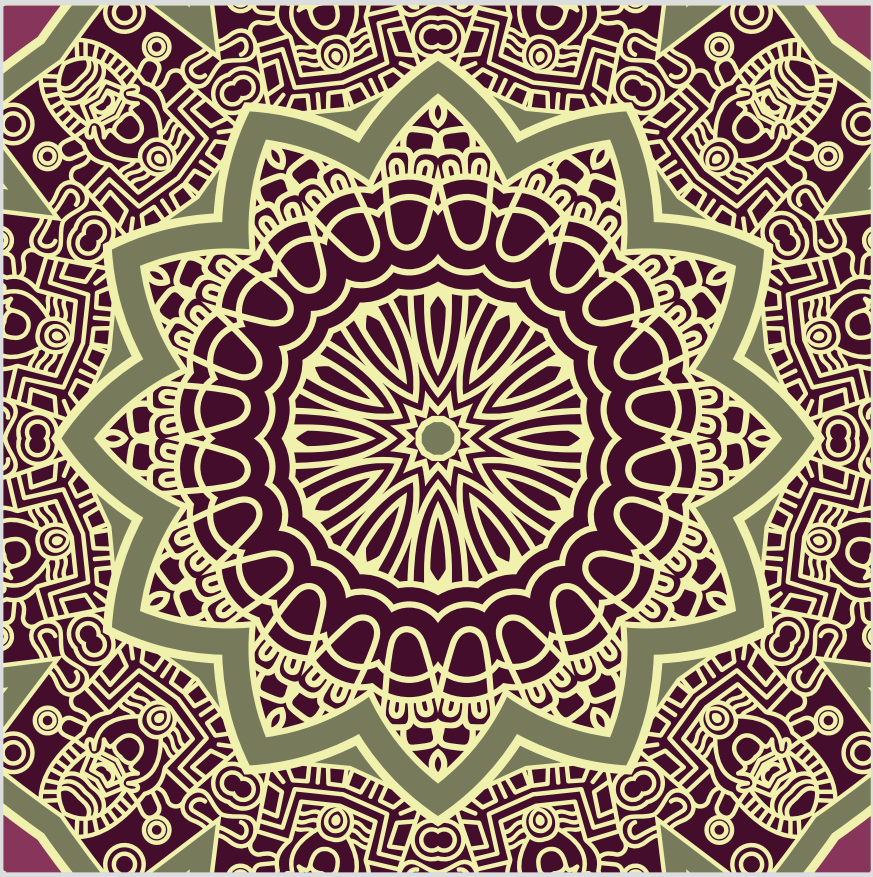
Claudia Maldonado Trujillo
Gabriela Pérez Yarahuán
(compiladoras)

Antología sobre evaluación

La construcción de una disciplina



GOBIERNO Y POLÍTICAS PÚBLICAS

CIDE

ANTOLOGÍA SOBRE EVALUACIÓN

La construcción de una disciplina

GOBIERNO Y POLÍTICAS PÚBLICAS

Claudia Maldonado Trujillo
Gabriela Pérez Yarahuán
(compiladoras)

Antología sobre evaluación

La construcción de una disciplina

Con artículos de:

Eleanor Chelimsky
Huey-Tsyh Chen
Howard Freeman
Jennifer Greene
Gary T. Henry
Mark W. Lipsey
Melvin M. Mark
Michael Q. Patton
Peter H. Rossi
Jean-Claude Thoenig
Carol H. Weiss

Prólogo de María Bustelo

CIDE

www.cide.edu

clear
Regional Centers for
Learning on Evaluation and Results

www.clear-la.cide.edu

Primera edición, 2015

Biblioteca del CIDE – Registro catalogado

Maldonado Trujillo, Claudia

Antología sobre evaluación. La construcción de una disciplina / Claudia Maldonado Trujillo y Gabriela Pérez Yahuarán (comps.); prólogo de María Bustelo; traducción de Erika Benton Rico y Román Villar Alonso — México, D. F.: Centro de Investigación y Docencia Económicas: Centro CLEAR para América Latina, 2015.

Primera edición

428 pp.: 23 cm. Colección: Gobierno y políticas públicas

I. Preparando el terreno / Carol H. Weiss – II. Evaluación con sentido: el enfoque basado en la teoría / Huey-Tsyh Chen y Peter H. Rossi – III. ¿Qué se puede construir con miles de ladrillos? Reflexiones sobre la acumulación de conocimientos en la evaluación de programas / Mark W. Lipsey – IV. La interfaz entre la evaluación y las políticas públicas / Carol H. Weiss – V. La evaluación como defensa / Jennifer Greene – VI. Los propósitos de la evaluación en una sociedad democrática / Eleanor Chelimsky – VII. Evaluación, gestión del conocimiento, mejores prácticas y lecciones aprendidas de gran calidad / Michael Q. Patton – VIII. La evaluación como conocimientos utilizables para las reformas de la gestión pública / Jean-Claude Thoenig – IX. Más allá de la utilización. La influencia de la evaluación sobre las actitudes y las acciones / Gary T. Henry y Melvin M. Mark – X. El contexto social de la evaluación / Peter H. Rossi, Howard Freeman y Mark W. Lipsey.

Incluye referencias bibliográficas.

ISBN 9786079367619

1. Evaluation
2. Public administration – Evaluation
3. Organizational effectiveness – Evaluation

I. Pérez Yahuarán, Gabriela
II. Benton Rico, Erika
III. Villar Alonso, Román
IV. Centro CLEAR para América Latina.

JF1355 M35 2015

Traducción: Erika Benton Rico y Román Villar Alonso

Revisión técnica de la traducción: Oliver Peña Habib y Alejandro Martínez Fierros

Dirección editorial: Natalia Cervantes

Imagen de portada: Elic. Usado bajo licencia de Shutterstock.

www.LibreriaCIDE.com

Centro CLEAR para América Latina

Impreso en México – *Printed in Mexico*

Índice

I. Preparando el terreno*

Carol H. Weiss

> Empieza por el principio —dijo el Rey con gravedad— y sigue hasta llegar al final; allí detente.
>
> Lewis Carroll

> El sistema federal emergente para la investigación en ciencias sociales refleja el reconocimiento de las necesidades de planificación y de que exista una estrecha relación entre los conocimientos y la formulación y ejecución de programas sociales.
>
> Gene Lyons

> En vez de símbolos triviales o nuevos programas federales a gran escala, [el Presidente] podría iniciar una serie de estudios audaces de investigación… Si bien no se resolverían inmediatamente los problemas de mayor magnitud, sí proliferarían las pruebas piloto y, al evaluarlas, aumentaría el aprendizaje.
>
> Richard Darman

¿En qué consiste una evaluación? Empecemos con un ejemplo sencillo, el de un programa para ayudar a las personas a dejar de fumar. Se llama a unas 100 personas cada mes para que participen en tres sesiones de actividades grupales. Una vez que

* Este artículo es una traducción al español del original en inglés: Carol H. Weiss, "Setting the scene", en su libro *Evaluation: Methods for studying programs and policies*, Prentice Hall, Jersey City, 1998, pp. 1-19 (Weiss, Carol H., *Evaluation*, 2.ª ed., © 1998. Permiso de reproducción otorgado por Pearson Education, Inc., Upper Saddle River, Nueva Jersey).

el programa haya operado alrededor de un año, el personal que lo ejecuta se pregunta cuál es el grado de avance. Sin embargo, no se ha entrevistado a los participantes una vez finalizadas las actividades, y no hay una manera de saber si se ha logrado el éxito esperado. ¿Las personas que atiende el programa realmente dejan de fumar y se mantienen alejadas del cigarro?

El personal decide efectuar una evaluación con el propósito de averiguar si se alcanzaron los objetivos del programa: ¿los participantes dejaron de fumar? En ese caso, la evaluación es relativamente sencilla, porque no existe un indicador claro de éxito. Que las personas dejen de fumar significa que el programa consiguió su objetivo; no se requiere indagar más, puesto que la investigación biomédica ha demostrado de forma contundente que dejar de fumar mejora la salud y la longevidad del individuo. El personal revisa las listas de participantes que asistieron a la sesión de la semana anterior, los llama por teléfono, les pregunta si dejaron de fumar, y 47% de ellos responde que sí.

Varios miembros del personal indican que una semana les parece muy poco tiempo y que no resulta tan difícil dejar de fumar durante siete días. Lo importante es que los participantes no vuelvan a tocar un cigarro. Por lo tanto, obtienen los nombres de los que asistieron a las sesiones de los primeros tres meses del programa, es decir, de hace casi un año, y los llaman por teléfono. En esa ocasión, la pregunta de si dejaron de fumar arroja 39% de respuestas afirmativas. Los entrevistadores se miran unos a otros desconcertados y se preguntan si dicho porcentaje es bueno. Habían esperado que 100% dejara de fumar, pero 39% equivale a un porcentaje de bateo de 0.39 (tremendo promedio para un jugador de beisbol) y probablemente representa a 39% más personas de las que habrían dejado de fumar sin el programa, ¿verdad?

Un miembro del equipo de evaluación recuerda a sus colegas que muchas personas dejan de fumar sin ayuda alguna y que quizá les habría ido bien de todas maneras sin inscribirse al programa. Por lo tanto, deciden llamar a los fumadores que conocen y saben que no asistieron a las sesiones para plantearles

la misma pregunta: "¿Dejó de fumar?". Los resultados así obtenidos demuestran que sólo 2% de esas personas lo logró. Este hallazgo lleva a una de las evaluadoras a decir: "¿Lo ven? Nuestro programa sí produjo cambios".

Sin embargo, alguien objeta: "La comparación no es justa. Las personas que asistieron a las sesiones se encontraban *motivadas* para dejar de fumar mientras que las demás no". Entonces, su problema radica en cómo encontrar personas con quienes comparar a los participantes, es decir, cómo encontrar personas motivadas a dejar de fumar que no asistieron al programa. Al ver la lista de asistentes de otro programa para fumadores, un miembro del equipo de evaluación sugiere: "Con ese programa se podría hacer una comparación justa". Con todo, los demás opinan que dicho ejercicio sólo compararía su programa con el de otros, y que esto no es realmente lo que buscan. Desean saber si el programa funciona, no si es peor o mejor que otro. En otras palabras, ambos programas podrían resultar deficientes o excelentes, pero la comparación no podría revelarlo.

Alguien capacitado en investigación sugiere que formen un grupo de control de asignación aleatoria, en el que incluirían a personas con la misma motivación para dejar de fumar. Por lo tanto, de las siguientes 200 solicitudes para entrar al programa (que representarán a personas con niveles similares de motivación), aceptarán al azar a 100, y a las 100 restantes les indicarán que ya no hay lugar (quien sugiere la idea explica que dicha medida es ética, porque el programa recibe muchas más solicitudes de las que puede aceptar y, en vez de inscribir a las personas interesadas conforme se vayan presentando, las aceptarán utilizando una técnica similar a la aplicada en la lotería y que se emplea también en la selección aleatoria). Una selección al azar no equivale a una selección accidental; implica apegarse a procedimientos muy estrictos basados en las leyes de probabilidad. Si todas las personas tienen la misma oportunidad de quedar en el grupo del programa, entonces habrá alta probabilidad de que tanto las personas aceptadas como las rechazadas (grupo de control) resulten muy parecidas en todas las dimensiones,

incluidas la edad, el tiempo que llevan fumando y el grado de apoyo familiar para dejar el hábito, entre muchas otras.

Por consiguiente, deciden ponerlo en práctica. De las siguientes 200 personas que presentan su solicitud, aceptan al azar a 100 e informan a las demás que deberán esperar a la sesión del próximo mes. Transcurrido dicho periodo, convocan a los dos grupos de personas (no pueden esperar más, porque ahora sienten la obligación de incorporar al programa al grupo de control). Las respuestas a las entrevistas telefónicas revelan que 46% del grupo del programa informa que dejó de fumar, mientras que sólo 15% del grupo de control lo logró.

El personal cree ahora haber demostrado el éxito inequívoco de su programa: las personas que no participaron fracasaron en el intento, a pesar de exhibir el mismo grado de motivación. Alguien subraya: "Esperen un momento. Como les dijimos que los podríamos integrar al programa al mes siguiente, ¿qué caso tendría que intentaran dejar de fumar por sí mismos? Esperan entrar al programa y, mientras tanto, disfrutan de su último mes de nicotina. Esto provocaría que el número de personas que dejó de fumar en ese grupo sea pequeño". Sus colegas lo miran molestos y le preguntan por qué no había pensado en eso antes. Otro de ellos añade: "En su defecto, si estaban realmente motivados, tal vez no quisieron esperar y fueron a buscar otro programa, situación que podría aumentar el número de quienes lograron dejar de fumar".

Un tercer miembro del personal indica que hay otro problema: "¿Qué significa que alguien diga que ha dejado de fumar? ¿Esa persona dejó de fumar un día antes de que la llamáramos o desde que salió del programa? Probablemente muchos dejaron y retomaron el cigarro una y otra vez". Por lo tanto, sugiere que tal vez deberían haber planteado muchas más preguntas sobre cuánto ha fumado cada persona desde que salió del programa. Alguien retoma la sugerencia y añade: "Estoy de acuerdo. Incluso si un participante logró disminuir el número de cigarros que fuma diariamente, sería un buen resultado, pero no preguntamos nada que nos ayudara a descubrirlo".

En este punto, el equipo estuvo a punto de descartar toda la evaluación. Incluso, se desanimaron aún más cuando la directora de la institución les preguntó: "¿Qué los lleva a pensar que las personas les dijeron la verdad? Ustedes se encontraban a cargo del programa y por ello algunos participantes que no dejaron de fumar pudieron haber mentido y decir que justamente lo dejaron para no herir sus sentimientos. Deberían verificarlo con parientes o amigos de los participantes o quizá someterlos a una prueba de saliva para detectar la presencia de nicotina". Con un gemido, alguien objetó que no contaban con el tiempo ni el dinero necesarios para realizar un estudio tan elaborado y preguntó si había algo que pudieran hacer en ese momento.

Espero que al terminar de leer este libro[1] puedan ayudar al equipo de evaluación a tomar decisiones razonables con respecto a qué acciones realizar en la siguiente etapa. Si el lector desea seguir trabajando con el ejemplo de los programas para dejar de fumar, puede consultar las siguientes referencias que lo remitirán a algunas de las miles de evaluaciones de programas de ese tipo.[2]

El ámbito de la evaluación

La gente hace evaluaciones todo el tiempo. Al escuchar una conversación, oímos cosas como: "Me encantó el programa de anoche", "su trabajo es muy deficiente", "el automóvil no vale lo que cuesta", "la comida en el Café de José es mucho mejor

[1] Weiss se refiere al libro *Evaluation* (*op. cit.*), del cual éste es su primer capítulo (N. del E.).

[2] Véanse C. Tracy Orleans *et al.*, "Self-help quit smoking interventions: Effects of self-help materials, social instructions, and telephone counseling", *Journal of Consulting and Clinical Psychology*, vol. 59, núm. 3, 1991, pp. 439-448; Deborah Ossip-Klein *et al.*, "Effects of a smokers' hot line: Results of a 10-county self-help trial", *Journal of Consulting and Clinical Psychology*, vol. 59, núm. 2, 1991, pp. 325-332; Chockalingam Viswesvaran y Frank L. Schmidt, "A meta-analytic comparison of the effectiveness of smoking cessation methods", *Journal of Applied Psychology*, vol. 77, núm. 4, 1992, pp. 554-561; Diane Zelman *et al.*, "Measures of affect and nicotine dependence predict differential response to smoking cessation treatments", *Journal of Counsulting and Clinical Psychology*, vol. 60, núm. 6, 1992, pp. 943-952.

ahora que antes". En términos más formales, uno escucha a un supervisor evaluando el desempeño de un empleado, a un maestro evaluar a un estudiante o, bien, leemos la evaluación de informes del consumidor relacionados con una línea de productos. Al igual que el *Bourgeois gentilhomme* de Molière, que se da cuenta que toda su vida ha hablado en prosa sin haber estado consciente de ello, nosotros nos percatamos de que hemos estado evaluando cosas diversas sin por ello utilizar la palabra *evaluación* para nombrar dicho proceso.

Evaluación es una palabra elástica que se extiende para abarcar juicios de todo tipo. Sin embargo, el común denominador de todos los usos de esa palabra es la noción de juzgar el mérito. Se examina y se compara un fenómeno (una persona, una cosa o una idea) con un estándar implícito o explícito. Dichos estándares de comparación pueden variar ampliamente. Por ejemplo, un criterio podría enfocarse a lo estético: ¿la entidad es hermosa y agradable? Otro podría medir la efectividad: ¿hace lo que se espera? Otro más podría basarse en la eficiencia: ¿los beneficios proporcionados justifican los costos? Existen estándares de comparación que se ocupan de cuestiones como la justicia, la equidad, la aceptabilidad en términos de normas comunitarias, el gozo, la satisfacción y las contribuciones a la armonía social, entre otras.

Los fenómenos que se pueden evaluar también son diversos, pero en este libro me centraré en la evaluación de un tipo especial de fenómeno: los programas y políticas diseñados para mejorar el bienestar de las personas. Existen programas y políticas de muchas clases: están los que se ocupan, por ejemplo, de la educación, del bienestar social, de la salud, de la vivienda, de la salud mental, de los servicios jurídicos; también existen los que se ocupan de asuntos correccionales, del desarrollo económico y de la construcción de carreteras, entre muchos otros. Pueden ser programas gubernamentales en el ámbito federal, estatal o local e, incluso, internacional; los pueden implementar organizaciones con o sin fines de lucro. Es característico que busquen cambiar el conocimiento, las actitudes, los valores y las conductas de las personas, así como hacer modificaciones a las organizaciones

donde trabajan, a las instituciones con las que deben negociar o a las comunidades donde viven. La cualidad en común de esos programas y políticas es la meta de mejorar y hacer más gratificante la vida de las personas a quienes deben beneficiar.

Más aún, me interesa en este capítulo explorar un método específico de evaluación, a saber, la investigación evaluativa. En lugar de depender del juicio de observadores expertos o de informes periódicos del personal, el evaluador en este escenario utiliza métodos de investigación de las ciencias sociales para lograr que el proceso de evaluación se vuelva más sistemático y preciso. En su carácter indagatorio, la evaluación establece cuestiones claras para investigar; recoge evidencia sistemáticamente de una variedad de personas involucradas en un programa dado; algunas veces traduce la evidencia en términos cuantitativos (por ejemplo, 23% de los participantes en el programa, calificaciones de 85 o más) y en ocasiones transforma los datos en narrativas convincentes. Posteriormente, llega a conclusiones acerca de la manera en que se ejecuta un programa, de sus consecuencias a corto plazo o de su efectividad para satisfacer las expectativas de quienes lo patrocinan, lo administran y lo dotan de personal o de participantes.

En el presente capítulo, ofreceré primero una definición de evaluación. Después, consideraré los diferentes tipos de fenómenos que normalmente se evalúan de manera formal. Existen varios géneros distintos de evaluación; aquí analizaré la diferencia entre la evaluación de la operación de los programas y la concerniente a los resultados del programa. Consecuentemente, exploraré las diferencias y las similitudes entre la evaluación y otros tipos de investigación en ciencias sociales. En otra sección examinaré la historia de la evaluación.

Definición de evaluación

Me gustaría plantear una definición provisional de evaluación que desarrollaremos a lo largo del libro: *la valoración sistemática*

de la *operación* y/o *de* los *impactos* de un programa o política al compararlos con un conjunto de *estándares implícitos* o *explícitos* para contribuir al *mejoramiento* del programa o política en cuestión.

Analicemos los cinco elementos clave de esa definición. El primero es la valoración sistemática; el énfasis en el sistema indica el carácter investigativo de los procedimientos de la evaluación. Sin importar si se realiza investigación cuantitativa o cualitativa, se debe llevar a cabo con formalidad y rigor, conforme a los cánones aceptados de investigación en ciencias sociales. El segundo y tercer elementos de la definición se centran en el objeto de investigación: la operación y los impactos del programa. Algunas evaluaciones se enfocan en estudiar el proceso, es decir, la manera en que se instrumenta un programa. Tales ejercicios podrían interesarse en determinar el grado en que el programa cumple con las prácticas prescritas (es decir, la fidelidad del programa hacia su diseño) o podrían abocarse sólo a plantear cómo van las cosas. Algunas otras evaluaciones se concentran en los impactos y en los efectos del programa para su población beneficiaria. Estos ejercicios buscan responder las siguientes preguntas: ¿los participantes obtienen los beneficios esperados? En su defecto y de forma más abierta, ¿qué sucede con los beneficiarios como resultado de la intervención del programa? Muchas evaluaciones examinan tanto el proceso del programa como los impactos para la población beneficiaria.

Los estándares de comparación constituyen el cuarto elemento de la definición. Una vez recopiladas las evidencias sobre el proceso y los resultados, la evaluación determina el mérito del programa al comparar las evidencias con una serie de expectativas. Sin importar si la evaluación se enfoca en el proceso o en los resultados del programa, siempre existirá un elemento de juicio. En ocasiones, el criterio aplicado para emitir juicios proviene de la declaración oficial de los objetivos del programa o de política al momento de su implementación. Si el programa se estableció para reducir el embarazo adolescente, un estándar explícito probablemente será la disminución en la frecuencia con que se embarazan las adolescentes que participan en el programa. Sin

embargo, los objetivos oficiales no constituyen la única fuente posible de criterios de evaluación, porque podrían haber cambiado en el transcurso de las operaciones. Quizá el objetivo actual consista, por ejemplo, en enseñar prácticas adecuadas de salud prenatal a las adolescentes embarazadas. Si el programa actualmente canaliza su energía hacia ese tema, entonces las prácticas sanitarias de las adolescentes participantes y el peso al nacer de sus bebés podrían considerarse criterios importantes. Otros estándares para emitir juicios pueden provenir de las expectativas que otros actores tienen respecto al contexto del programa o de la política. Los objetivos de los patrocinadores, de los gerentes de programa, de los profesionales y de los participantes también pueden convertirse en criterios de evaluación. En algunos casos, la evaluación ofrece evidencia sistemática sobre el programa en ámbitos específicos y deja en manos del lector la responsabilidad de llegar a una conclusión sobre el mérito del mismo.

El quinto elemento de la definición de evaluación es el propósito de llevarla a cabo; por ejemplo, contribuir al mejoramiento del programa y de la política. La evaluación es un quehacer práctico, diseñado para ayudar a mejorar el funcionamiento de los programas y para asignar recursos a los mejores. Los equipos de evaluación esperan que las autoridades empleen sus resultados para tomar las medidas adecuadas, y sienten satisfacción por la oportunidad de contribuir al mejoramiento social.

Las evaluaciones basadas en un proceso de investigación requieren más tiempo y resultan más costosas que las informales; estas últimas dependen de la intuición, de las opiniones o de una sensibilidad disciplinada. En cambio, las primeras exhiben una sistematicidad ausente en los ejercicios informales. El rigor se considera de particular importancia cuando *a)* los resultados por evaluar son complejos, difíciles de observar y constan de muchos elementos que reaccionan de formas diversas; *b)* las decisiones subsecuentes son importantes y costosas; y *c)* se requieren evidencias para convencer a otras personas de la validez de las conclusiones.

Gran variedad de programas sociales opera en Estados Unidos y en otros países, y continuamente nuevas necesidades

impulsan el desarrollo de programas adicionales. En la década pasada, surgieron programas para alimentar y dar refugio a personas sin hogar, para ofrecer cuidados a pacientes con SIDA, para educar a la juventud sobre los peligros de las drogas, para disminuir los costos de la atención médica y para ayudar a víctimas de desastres de todo el mundo a retomar el control de su vida. Algunos programas nuevos son extensiones lógicas de esfuerzos previos, mientras que otros representan desviaciones radicales del pasado para explorar audazmente territorios desconocidos.

Muchos desean —y necesitan— saber: ¿cómo se ejecuta el programa?, ¿qué acciones lleva a cabo en realidad?, ¿cuánto se apega a los lineamientos establecidos originalmente?, ¿qué tipo de resultados produce?, ¿en qué grado cumple con los propósitos para los cuales se estableció?, ¿vale el dinero que cuesta?, ¿se debería continuar, expandir, reducir, modificar o abandonar?, ¿la gente debería acudir en tropel al programa, inscribirse selectivamente o alejarse del todo?, y ¿funciona para todos o sólo para ciertos grupos?

Resulta difícil obtener respuestas mediante herramientas informales. Las personas mejor informadas (quienes ejecutan el programa) tienden al optimismo y, en todo caso, les interesa comunicar que tuvieron éxito. Muchos esfuerzos de los programas ofrecen servicios distintos y trabajan con grandes cantidades de participantes. Por ello, unos cuantos "testimonios de los consumidores" o un rápido viaje de inspección difícilmente podrían medir su efectividad. Las decisiones sobre operaciones futuras afectarán el destino de muchos e implicarán enormes sumas de dinero; quienes tienen derecho a opinar —legisladores, juntas directivas o futuros clientes— no son tan cercanos al programa y desean obtener información concluyente que les ayude en la toma de decisiones.

Cuando una empresa privada ejecuta un programa, generalmente el mercado emite el juicio. Si la empresa vende un *software* o un curso de capacitación, entonces la evaluación determinará si la gente adquiere o no el producto o servicio. Con el tiempo, los buenos programas prosperan y los malos quedan fuera o, al

menos, así sucede si los consumidores cuentan con información suficiente acerca de sus efectos. Sin embargo, cuando se trata de un programa gubernamental o de una organización no lucrativa, la satisfacción o insatisfacción del cliente por lo general produce poco impacto. En ocasiones, el programa es lo único interesante que sucede en la localidad —como es el caso de los esfuerzos de ayuda asistencial o de los tribunales penales y las correccionales—, mientras que, en otras, es el único programa para un grupo en particular —como son los cursos de capacitación para las fuerzas armadas— o el único programa gratuito —como las escuelas públicas—. Con frecuencia, esos programas continúan sin importar el nivel de demanda o de satisfacción del cliente.

Por otra parte, las compañías privadas también pueden ejecutar programas similares a los antes mencionados para su propio personal. Los programas internos, como las estancias infantiles para los hijos de los empleados o los procedimientos de seguridad para trabajadores, también cuentan con una población cautiva y tienden a mostrarse relativamente inmunes a la satisfacción o insatisfacción del cliente.

Sin embargo, resulta evidente que todos ellos constituyen cuantiosas inversiones. Gobiernos, negocios, organizaciones no lucrativas y fundaciones pagan sumas sustanciales para que estos programas sigan operando. A muchas personas interesa si los programas operan de la manera esperada y si obtienen los resultados buscados. A muchos preocupa si los resultados podrían mejorar o si sería posible gastar menos dinero o, mejor aún, si ambos objetivos se han alcanzado. Con todo, sin llevar a cabo algún tipo de revisión sistemática, resulta difícil averiguarlo. Por lo tanto, la evaluación es la mejor herramienta disponible para ofrecer la información necesaria.

¿Qué se evalúa? Una digresión sobre la terminología

Para garantizar que estemos sintonizados, debemos acordar el significado de un conjunto de términos que surgirá a lo largo

de los capítulos siguientes. Algunas de las definiciones son estrictamente convencionales y arbitrarias, pero si estamos de acuerdo con ellas, significarán lo mismo para todos nosotros en lo que resta del texto.

De aquí en adelante, a programas nacionales, como Head Start (programa preescolar para niños de cero a cinco años) o Superfund Environmental Cleanup (programa nacional de eliminación de desechos peligrosos), los llamaré solamente *programas*, y a sus operaciones locales las llamaré *proyectos*. Así, el Head Start, operado en el Centro Comunitario de Brigham, es un proyecto, y uno de sus elementos —lograr la participación de los padres de familia mediante reuniones semanales— constituye un componente. Las evaluaciones se pueden dirigir a cualquiera de esos ámbitos; podemos evaluar programas nacionales, proyectos locales o componentes de subproyectos con los mismos métodos básicos.

También podemos evaluar políticas. Para los fines de este texto, se define política como la declaración oficial de los objetivos asociados con una serie de actividades encaminadas a alcanzar los objetivos de una jurisdicción en particular. De esa manera, una política federal busca promover donativos para organizaciones sin fines de lucro al permitir que los contribuyentes los deduzcan en sus declaraciones de ingresos. Lo anterior no constituye un programa, sino una política, la cual se puede evaluar. Pregunta de evaluación: ¿la deducción de impuestos realmente promueve la donación filantrópica? Otro ejemplo: una política, también federal, busca garantizar la salud y la seguridad de quienes utilizan equipos médicos, como marcapasos y sillas de ruedas. El gobierno exige a los fabricantes que pongan a prueba la seguridad de los equipos antes de comercializarlos y una vez más después de venderlos al pedir a los usuarios que reporten defectos de fabricación. Tal política también se puede evaluar. Pregunta de evaluación: ¿los requisitos de las pruebas y los procedimientos para efectuar reportes mejoran la seguridad de los dispositivos? Muchas de las mismas estrategias se utilizan para evaluar políticas y programas.

De hecho, las técnicas de evaluación son muy adaptables; se pueden aplicar a programas y políticas sociales, así como a políticas ambientales, programas de tránsito masivo, proyectos forestales y elementos de compras militares. En muchos de esos campos, no se utiliza comúnmente la palabra *evaluación*, aunque el ejercicio se realiza con otros nombres; por ejemplo, auditorías sociales, revisiones sistémicas o medición del desempeño. Para simplificar la prosa, hablaré en general de la evaluación de programas, aunque utilizo esta expresión meramente para abreviar. Con el término "evaluación de programas" pretendo abarcar la evaluación de políticas, de proyectos y de componentes.

Otra convención que he adoptado aquí es la de llamar *evaluadora* a la persona encargada de la evaluación y usar el masculino para todos los otros actores del programa. De tal manera, los diseñadores de políticas, gerentes de programas, personal del programa y clientes del mismo aparecerán con pronombres masculinos y denotaré a los miembros del equipo de evaluación con un pronombre femenino. De ninguna manera sugiero que todas las evaluadoras son o deberían ser mujeres ni que todos los demás actores del ámbito programático son o deberían ser hombres. Sólo se trata de una convención para evitar usar "él" o "ella" cada vez que utilice un pronombre en singular.[3]

En una evaluación, se utilizan frecuentemente otros términos, muchos de los cuales les resultarán conocidos debido a la exposición que han tenido a trabajos de investigación; entre ellos se encuentran investigación cualitativa, investigación cuantitativa, experimento, grupos de control, asignación aleatoria, validez, confiabilidad, medición, indicadores, variables, muestreo e investigación empírica. Si no conocen su significado con certeza, quizá sería conveniente repasar las definiciones para manejar los términos con comodidad. En este texto,

[3] Lee Cronbach y sus asociados utilizaron dicha convención, pero al revés. Usaron un pronombre masculino para los evaluadores, y para el resto, un pronombre femenino. Yo prefiero la convención arriba descrita (Lee Cronbach *et al.*, *Toward reform of programme evaluation*, Jossey-Bass, San Francisco, 1980).

presento conceptos que se utilizan solamente en el ámbito de la evaluación: evaluación formativa y sumativa, y evaluación de resultados y procesos.

Evaluación de impacto y procesos

¿Qué clase de preguntas formula una evaluación sobre los programas (políticas, proyectos, componentes)? Desde el momento en que se convirtió en una actividad reconocida, en el periodo entre las décadas de 1930 y 1970, la principal pregunta se dirigía a los resultados. Las evaluaciones tendían a concentrarse en la siguiente pregunta: ¿el programa ha alcanzado las metas que debía lograr? O de manera más general: ¿cuáles son los resultados del programa? Dicho enfoque en los resultados sigue siendo la característica distintiva del propósito de la evaluación.

Los resultados se refieren a las consecuencias finales del programa para las personas a las que se buscaba servir. Utilizo el término "resultados" de manera intercambiable con el término "efectos". Algunos de los resultados producidos son las consecuencias previstas por los planificadores del programa —las cosas que deseaban que sucedieran—. Sin embargo, otros resultados constituyen efectos secundarios que nadie esperaba; frecuentemente, se trata de efectos que nadie deseaba. Así, por ejemplo, un programa dirigido a padres de familia en riesgo de abusar de sus hijos pequeños podría generar el efecto no buscado de etiquetarlos como abusadores potenciales, incluso con el conocimiento de los mismos padres. Una vez que acepten la etiqueta, podría producirse el efecto contrario y podrían volverse más abusivos. En su defecto, un programa que implementa grupos de asesoría para los ciudadanos respecto a las actividades de eliminación de residuos tóxicos podría contemplar la meta de acelerar el proceso mediante la presión ciudadana. Sin embargo, una vez que los ciudadanos se encuentren representados, podrían incorporar al análisis de la eliminación de residuos una serie de preocupaciones, como el empleo y el desempleo. Por consiguiente, podrían

redirigir las actividades de eliminación de residuos a las relacionadas con la retención de puestos de trabajo.

Cuando se habla del propósito de la evaluación, otro término que se utiliza también es el de *impacto*. La mayoría de las veces significa lo mismo que resultados finales. Un estudio de impacto analiza qué sucede a los participantes como resultado del programa y, en ocasiones, el impacto se interpreta como los resultados finales a largo plazo. De vez en cuando, se emplea la palabra impacto para referirse a los efectos del programa en la comunidad en general; por ejemplo, los impactos de un programa de capacitación laboral se podrían analizar en términos de la cantidad de participantes del programa que consiguen trabajo. En ese sentido más amplio, el impacto podría ser el efecto en la tasa de desempleo de toda el área geográfica en donde opera el programa, mientras que en otros, el impacto se interpone entre los procesos y los resultados finales de un programa.[4]

Otra acepción de impacto es el efecto o los efectos *netos* del programa, después de considerar qué habría sucedido de no haberse implementado. Desde el punto de vista operativo, significa examinar los impactos para los participantes del programa (por ejemplo, 9% tiene empleo después de la capacitación) y los impactos para una población *equivalente* que no haya participado en el programa (por ejemplo, 3% tiene empleo en el momento analizado).

Entonces, partiendo del supuesto de que la misma proporción de participantes (3%) tendría empleo si no hubiera asistido al programa, la evaluadora hace la resta de los impactos en la población no participante menos los impactos en los participantes y obtiene que el impacto del programa es un aumento de seis puntos porcentuales (9% menos 3%) en el número de personas con empleo (aquí el punto consiste en asegurarse de que los dos grupos sean equivalentes).[5]

[4] Lawrence W. Green y Marshall W. Kreuter, *Health promotion planning: An educational and environmental approach*, Mayfield Publishing, Mountain View, 1991.

[5] Este tema se analiza en el capítulo 9 [de la obra original].

Ese último significado de impacto resulta vital en la evaluación, pero no lo utilizo en este texto. El término que empleo es el de *resultados* y, en contexto, frecuentemente significa "resultados netos", es decir, la parte de los resultados que es atribuible al programa. Al ir avanzando en el tema, el concepto quedará más claro.

Además de los impactos, la evaluación puede enfocarse en los resultados a corto plazo y, si se cuenta con suficiente tiempo, la evaluación examina también los de largo plazo. El estudio *High Scope* del programa preescolar de Perry ha dado seguimiento durante 24 años a niños que recibieron educación preescolar a los 3 y 4 años de edad.[6] El seguimiento a largo plazo permite a las evaluadoras informar que, a los 27 años, los jóvenes que habían asistido al programa preescolar obtuvieron ingresos mensuales significativamente mayores que los de un grupo equivalente de personas que no participaron en el programa; además, tuvieron mayores probabilidades de ser propietarios de su casa, menos arrestos y mayor escolaridad. Los datos obtenidos han influido en gran medida para justificar el programa Head Start, mencionado arriba, que se basa en el programa preescolar de Perry, y para apoyar el aumento del financiamiento.

Aunque los resultados constituían el enfoque original de la evaluación, las preguntas actuales de las evaluadoras se dirigen tanto a éstos como al proceso del programa, es decir, examinan qué sucede. Las evaluadoras necesitan explorar qué *hace* realmente el programa. Anteriormente, se daba por hecho que el programa llevaba a cabo lo que sus operadores informaban, pero pronto fue evidente que a menudo este supuesto estaba mal fundamentado. Uno de los primeros evaluadores de la historia, Ralph Tyler, escribió en 1942 que su estudio de ocho años sobre treinta escuelas de educación media reveló que en el primer año "sólo dos o tres operaban el programa descrito en

[6] Lawrence Schweinhart *et al.*, *Significant benefits: The High/Scope Perry preschool study through age 27*, High/Scope Press, Ypsilanti, 1993.

la propuesta"[7] al evaluar un programa de salud rural en Egipto. Encontró que la mayoría de los centros de atención carecían de las cantidades esenciales de personal y que quienes laboraban en ellos trabajaban relativamente pocas horas. Para la evaluación, se debía saber en qué consistían realmente los programas antes de obtener conclusiones sobre si habían sido exitosos o no. Podía darse el caso de que nunca hubieran llegado a implementarse.

Hay otras razones para estudiar los procesos de los programas. A veces, la comunidad donde opera el programa plantea preguntas clave sobre sus procesos. Por ejemplo: ¿qué servicio reciben los participantes?, ¿el servicio se apega a las indicaciones del desarrollador del programa?, ¿con qué frecuencia se presentan los participantes (por ejemplo, de un programa de tratamiento médico)?, ¿qué problemas encuentra el personal?, y ¿los clientes se sienten contentos con el programa? Se necesitan llevar a cabo estudios dedicados a la evaluación sistemática de lo que sucede dentro del programa.

Otra razón para estudiar el proceso de un programa es ayudar a entender los datos de los resultados. La evaluadora podría encontrar que algunos participantes se desempeñaron particularmente bien, mientras que con otros sucedió todo lo contrario. Hay muchas razones posibles para dicho hallazgo; una de ellas es que se proporcionaron distintos tipos de servicio o de intensidad de servicio. En otras palabras, un grupo podría haber asistido con regularidad y haber recibido atención de parte de personal muy calificado o de la misma persona durante un largo periodo, mientras que otro grupo podría haber sido atendido por distintos miembros del personal con una capacitación deficiente, y quizá no asistieron con frecuencia. Para que la evaluadora pueda analizar qué condiciones produjeron

[7] Ralph Tyler, "General statement on program evaluation", en Milbrey Wally McLaughlin y Brenda D. Phillips (eds.), *Nineteenth yearbook of the National Society for the Study of Education*, Chicago Press, Chicago, 1991, t. 2, p. 7; Herbert Hyman y Charles Wright, "Evaluating social action programs", en William Hamilton Sewell *et al.* (eds.), *The uses of sociology*, Basic Books, Nueva York, 1967, p. 745.

los distintos impactos, requiere información sobre qué sucedió dentro del programa.

Por lo tanto, existen al menos tres situaciones donde se requieren datos de proceso. Una de ellas es cuando hay preguntas clave relativas al proceso; los patrocinadores de la evaluación desean saber qué sucede. Otra situación es cuando hay preguntas clave sobre los resultados; queremos asegurarnos *qué* los causó. Con frecuencia, se analiza un proyecto o algún grupo de ellos como muestra representativa de una clase de programas. Analicemos el ejemplo de centros de psicoterapia o de salud comunitaria de corto plazo; la evaluadora desea saber si en efecto existió un centro de salud comunitaria y en qué consistió el programa antes de concluir si este último fracasó o tuvo éxito. Una tercera situación es cuando la evaluadora desea asociar los impactos a elementos específicos del proceso programático; es decir, quiere descubrir qué características particulares del programa se asociaron a un grado mayor o menor de éxito.

Contribuciones de la evaluación

A lo largo de la historia, los simpatizantes de la investigación evaluativa la han considerado un medio para mejorar la racionalidad del diseño de políticas. Al contar con información objetiva sobre la implementación y los impactos de los programas, se pueden tomar decisiones atinadas sobre asignaciones presupuestarias y planificación de programas. Normalmente, se espera expandir los programas de buenos resultados y abandonar o modificar drásticamente los de resultados deficientes. A continuación, cito lo manifestado hace muchos años por un representante del Congreso. Me parece un buen resumen de la justificación de la evaluación de un programa:

> Resulta cada día más evidente que mucho de lo que invertimos en áreas como educación, salud, pobreza, empleo, vivienda,

> desarrollo urbano y transporte, entre otras, no genera dividendos adecuados en términos de resultados. Sin disminuir en ningún momento nuestro compromiso de trabajar para satisfacer tan urgentes necesidades humanas, un importante desafío que, sin embargo, el Congreso desatiende frecuentemente consiste en reevaluar la multitud de los programas sociales existentes, concentrar —y, de hecho, aumentar— los recursos en aquellos que *operan* donde se registran las mayores necesidades y disminuir o eliminar el resto. Ya no contamos con tiempo ni dinero para desperdiciar en cuestiones no esenciales que no produzcan, en los problemas, el impacto visible requerido.[8]

Para alcanzar los objetivos antes descritos, el Congreso estadounidense aprobó en 1993 la Ley de Desempeño y Resultados del Gobierno, que exige a los organismos federales que recolecten datos de desempeño de los programas y que elaboren un plan estratégico para las actividades de los programas; que establezcan metas de desempeño objetivas, cuantificables y medibles. También deben reunir datos que documenten el grado en que el programa alcanza las metas establecidas. Así también, los organismos se encuentran obligados a presentar, al presidente y al Congreso, un informe anual del desempeño de los programas.

Historia de la evaluación

Los esfuerzos antes mencionados son ejemplos recientes de intentos por institucionalizar la evaluación en los organismos gubernamentales. Son los últimos de una larga serie que pretende utilizar datos y evidencias para mejorar nuestra comprensión del comportamiento social y, con ello, el diseño de las políticas sociales. Si deseamos remontarnos a la prehistoria de la evaluación, quizá deberíamos comenzar en la década de 1660.

[8] F. Robert Dwyer, *Report to the people*, vol. 14, núm. 1, Nueva Jersey, 1970.

La evaluación se fundamenta en el estudio empírico de los problemas sociales, mismo que comenzó, con todo vigor, en esa década en Gran Bretaña.[9] Aunque los orígenes intelectuales de la alguna vez llamada "aritmética política" siguen siendo objeto de debate, resulta evidente que el siglo XVII fue testigo de los inicios de una búsqueda de leyes sociales comparables a las que se desarrollaban al mismo tiempo en las ciencias físicas. Quizá el primer estudio que se puede considerar como evaluación surgió casi dos siglos más tarde. El francés André Michel Guerry publicó en 1833 un estudio estadístico que buscaba demostrar que la educación no disminuía la criminalidad.[10] Otros expertos en estadística reunieron distintos datos para refutar sus hallazgos. A manera de contrapunto (que ha perdurado como característica permanente de la historia de la evaluación), tales expertos, además de citar distintos tipos de evidencia, criticaron los métodos de Guerry en el apasionado deseo que tenían de establecer que la educación sí disminuía la criminalidad.

Alrededor de los mismos años, otro francés, Jules Depuit, evaluó la utilidad de las obras públicas, como los caminos y los canales. Publicó un artículo en 1844 donde medía el valor de un proyecto de canales. Utilizó técnicas de *calcul économique*. Usó las cuotas máximas que pagaban los usuarios como evidencia del valor de los canales y planteó la caída en la demanda derivada del aumento de cuotas como una medida de los límites de su utilidad.[11]

A pesar de las incursiones tempranas como las antes descritas, la evaluación como se conoce actualmente constituye un desarrollo relativamente reciente en la historia del mundo, incluso dentro de la historia de los programas sociales. Las primeras políticas dirigidas a mejorar las condiciones sociales no

[9] Véase Michael J. Cullen, *The statistical movement in early Victorian Britain: The foundations of empirical social research*, Harper & Row, Nueva York, 1975.

[10] *Ibid*., p. 139.

[11] Véase Johnny Toulemonde y Lise Rochaix, "Rational decision-making through project appraisal: A presentation of French attempts", *International Review of Administrative Sciences*, 1994, pp. 37-53.

contemplaban la evaluación. Cuando los reformistas de finales del siglo XIX y principios del XX recurrieron a procedimientos de investigación de las ciencias sociales, su propósito era efectuar encuestas para documentar la magnitud de los problemas existentes e identificar a las personas necesitadas.[12] Dieron por hecho que los remedios que planteaban resolverían los problemas. No se evaluaron las reformas carcelarias promovidas por Dorothea Dix ni los servicios sociales proporcionados en el lugar conocido como Hull House de Jane Addams. Apenas se estudiaron los efectos de instalar luz eléctrica en las calles o de purificar el agua. Cuando el gobierno estadounidense, en la segunda década del siglo XX, aprobó la prohibición del trabajo infantil, a nadie se le ocurrió evaluar los impactos de tales leyes; simplemente se dio por sentado que se eliminaría el trabajo infantil e intrínsecamente los resultados de ello serían buenos. Cuando se instituyó el sistema estadounidense de beneficios por desempleo en 1935, no se diseñó ningún procedimiento de evaluación; evidentemente, se consideraban convenientes tales prestaciones encaminadas a ayudar al individuo durante el difícil periodo de búsqueda de empleo.

Los especialistas del campo de la educación y la salud fueron de los primeros profesionales en efectuar estudios sistemáticos de los impactos de su labor. En 1912, Richard Clarke Cabot examinó 3 000 informes de autopsia y los comparó con los diagnósticos efectuados para cada caso. El artículo que publicó en el *Journal of the American Medical Association* constituía esencialmente una evaluación de la calidad del diagnóstico clínico.[13] En 1914, el doctor Ernest Codman, cirujano del Hospital General de Massachusetts, subrayó que la forma de evaluar el desempeño de los cirujanos consistía en medir el estado de salud de los pacientes después de darlos de alta. El propio Codman recopiló numerosos datos, pero muchos

[12] Véase Martin Bulmer *et al.*, *The social survey in historical perspective*, Cambridge University Press, Nueva York, 1991.

[13] Véase Evelyn Flook y Paul J. Sanazaro (eds.), *Health services research and R&D in perspective*, Health Administration Press, s. l., 1973.

miembros de la comunidad médica ignoraron su trabajo.[14] En el ámbito de la educación, uno de los estudios más conocidos fue el ejercicio de ocho años, patrocinado por la Asociación de Educación Progresista en 1933 y dirigido por Ralph Tyler, que examinó los resultados de los programas implementados en treinta escuelas de educación media; quince eran progresistas y quince tradicionales. Por su parte, George Palmer, jefe de la división de investigación de la Asociación Estadounidense de Salud Infantil, realizó una evaluación temprana de la efectividad de los programas escolares de salud; la publicó en 1934. Al final de la década, el Commonwealth Fund patrocinó la evaluación más elaborada de Dorothy B. Nyswander; se publicó en 1942.[15]

En la década de 1940, las fundaciones privadas comenzaron a patrocinar evaluaciones de un grupo de programas sociales innovadores de sus donatarias. Uno de ellos, el famoso programa de trabajadores jóvenes, conocido como Cambridge-Somerville, buscaba prevenir la delincuencia en barrios suburbanos cercanos a Boston.[16] Los primeros resultados se consideraban prometedores, pero, con el seguimiento a largo plazo, se encontró que los jóvenes en situación de riesgo que habían recibido los servicios del programa progresaban aproximadamente igual que quienes no se habían beneficiado de los servicios.[17] Una interpretación de los hallazgos sugería que se habían vuelto tan dependientes de la ayuda que no habían desarrollado las habilidades necesarias para resolver sus propios problemas.

En la década de 1950, el gobierno federal empezó a patrocinar nuevos esfuerzos encaminados a desarrollar planes de estudio, tales como el Harvard Project Physics. Ésta fue una

[14] Véase Spencer Vibbert, *What works: How outcomes research will change medical practice*, Ground Rounds Press, Knoxville, 1993.

[15] Evelyn Flook y Paul J. Sanazaro, *op. cit.*

[16] Véase Edwin Powers y Helen Leland Witmer, *An experiment in the prevention of delinquency: The Cambridge-Sommerville youth study*, Columbia University Press, Nueva York, 1951.

[17] Véase William Maxwell McCord y Joan McCord, *Origins of crime: A new evaluation of the Cambridge-Sommerville youth study*, Columbia University Press, Nueva York, 1959.

respuesta al temor por el analfabetismo científico en Estados Unidos, que surgió después del lanzamiento del satélite Sputnik de los soviéticos. Por consiguiente, se financiaron evaluaciones para determinar el grado de éxito de los planes de estudio. Por su parte, a principios de la década de 1960, el Comité sobre Delincuencia Juvenil de la Presidencia otorgó fondos para una serie de proyectos que buscaba disminuir la criminalidad entre la juventud de todo el país; la administración federal exigió a cada proyecto que evaluara los resultados de sus actividades.

La Guerra contra la Pobreza de la misma década representa el inicio de los esfuerzos evaluativos a gran escala, financiados por el gobierno. El gobierno federal empezó a patrocinar programas de ayuda a los pobres y a exigir la evaluación sistemática de los resultados de los fondos utilizados. En la Ley de Educación Primaria y Secundaria de 1965, se incluyó el requisito de la evaluación, promovido por el senador Robert Kennedy. Él deseó asegurarse de que los nuevos fondos federales no se destinaran a apoyar prácticas escolares obsoletas, sino que se utilizaran para ayudar a niños desfavorecidos de manera innovadora. Asimismo, buscó que los padres con pocos recursos estuvieran bien informados sobre lo que sucedía en las escuelas, para que pudieran exigir a los educadores que atendieran a sus hijos de manera más efectiva. Para él, la evaluación era una herramienta que proveía a los padres la información necesaria.[18]

Se evaluaron otros programas de la Guerra contra la Pobreza, incluidos aquellos que proporcionaban servicios jurídicos, salud comunitaria, capacitación laboral, suplementos alimenticios para embarazadas y bebés, cupones de alimentos, cupones de vivienda, centros de servicios sociales múltiples, educación preescolar, innovaciones en la prevención de la delincuencia y las correccionales, servicios de salud mental y programas de acción comunitaria, que movilizaban a los residentes de barrios pobres para que identificaran sus prioridades y exigieran los

[18] Véase Milbrey Wallin McLaughlin, *Evaluation and reform: The elementary and secondary education Act of 1965, Title 1*, Ballinger, Cambridge, 1975.

servicios que requerían. Las evaluadoras desarrollaron nuevos métodos y herramientas para adaptarse a los diversos contenidos y escenarios de los programas. El desarrollo de la evaluación en ese periodo debe mucho a la pobreza, de la misma manera en que anteriormente debía al analfabetismo y a la criminalidad.

Ese mismo periodo fue testigo del surgimiento del análisis de costo-beneficio, que se implementó en la RAND Corporation, en el Departamento de Defensa y en otras instancias. Los analistas de políticas del Secretario de Defensa Robert MacNamara, apodados "los niños genio", examinaron las ventajas relativas de los sistemas de armas. Quisieron determinar, según la frase que empleaban, "la fuerza destructiva por dólar" de cada sistema. Como resultado, se lograron importantes avances en los métodos de análisis económico.

La evaluación se extendió a otras áreas. Con la nueva legislación, llegaron las evaluaciones a ámbitos como la protección ambiental, la conservación de los energéticos, el reclutamiento militar y el control de la inmigración. Al final de la década de 1970, la evaluación se había convertido en ejercicio común en todos los organismos federales. La mayoría de los departamentos contaban con su oficina de evaluación. Incluso algunos habían instalado oficinas para evaluar distintos niveles de la jerarquía: desde el secretario del departamento adjunto al área de programas más importante hasta el nivel operativo.

Se estableció una multitud de pequeños centros y compañías para la realización de evaluaciones con financiamiento federal. Los centros universitarios de investigación expandieron sus estatutos para incluir la evaluación; se instalaron centros especiales; surgieron nuevas organizaciones de investigación y consultoría con y sin fines de lucro. Las que ya estaban establecidas se extendieron hacia el ámbito de la evaluación. Muchos investigadores adaptaron sus habilidades para aprovechar los nuevos flujos de fondos.

Un punto sobresaliente de la historia de la evaluación fue la inauguración, durante la década de 1970, de una serie de experimentos sociales para ensayar políticas e ideas novedosas con el objetivo de desarrollar programas *antes* de su implementación.

El de mayor escala y el más difundido fue el experimento conocido como impuesto negativo al ingreso.[19] A éste siguieron ejercicios con beneficios para vivienda,[20] seguros de gastos médicos,[21] contratación por desempeño en el ámbito de la educación[22] y otros de menor escala. Como parte de tales experimentos, se implementaron programas piloto a escala suficiente como para promover condiciones reales de operación. Se esperaba que los resultados ayudaran a los diseñadores de políticas a decidir si procederían a implementar las políticas en el ámbito nacional. Sin embargo, cuando los resultados de los experimentos se encontraron disponibles, el clima político en general había cambiado; ejemplo de ello es el estudio acerca del impuesto negativo al ingreso. El entusiasmo inicial por el cambio había disminuido, y el movimiento reformista había perdido fuerza. Con el tiempo, se adoptaron pocos componentes de los programas experimentales, pero la información siguió disponible para diseñar políticas más adelante. Por ejemplo, en 1994, cuando la reforma de salud regresó a la agenda por un tiempo, se desempolvaron los resultados de los experimentos realizados con el seguro médico, se publicó un nuevo libro[23] y los hallazgos se sometieron a análisis.

[19] Véanse Glen C. Cain y Harold Watts, *Income maintenance and labor supply: Econometric studies*, Rand McNally College Publishing Co., Chicago, 1973; David Kershaw y Fair Jerilyn, *The New Jersey income-maintenance experiment: Operations, surveys and administration*, Academic Press, Nueva York, 1976-1977, ts. 1-2.

[20] Véanse David B. Carlson y John D. Heinberg, *How housing allowances work: Integrated findings from the experimental housing allowances program*, Urban Institute, Washington, D. C., 1978; Joseph Friedman y Daniel H. Weinberg (eds.), "The great housing experiment", *Urban Affairs Annual Review*, vol. 24, 1983; Stephen D. Kennedy, *The final report of the housing allowance demand experiment*, ABT Associates, Cambridge, 1980.

[21] Véanse Joseph Newhouse *et al.*, "Some interim results from a controlled trial of cost sharing in health insurance", *New England Journal of Medicine*, 1981; Charles E. Phelps y Joseph Newhouse, *Coinsurance and the demand for medical services*, Rand Corporation, Santa Mónica, 1973.

[22] Véase Alice M. Rivlin y P. Michael Timpane, *Planned variation on education: Should we give up or try harder?*, Brookings Institution, Washington, D. C., 1975.

[23] Véase Joseph Newhouse e Insurance Experiment Group, *Free for all? Lessons from the Rand health insurance experiment*, Harvard University Press, Cambridge (Mass.), 1993.

Cuando Reagan asumió la presidencia en 1981, la industria de la evaluación seguía en crecimiento, pero en ese momento el financiamiento para nuevas iniciativas sociales disminuyó drásticamente. Los programas nuevos e innovadores siempre se han considerado los mejores candidatos para la evaluación, pero cuando el número de programas nuevos se redujo a un puñado, también decrecieron las solicitudes de evaluación. Sin embargo, se continuaron dichos ejercicios en cantidades modestas, no tanto con fondos específicos para la evaluación, sino con recursos de las agencias encargadas de la operación. Se evaluaban programas en operación de atención médica a largo plazo, de padres adolescentes, de trabajadores desplazados, de exenciones estatales para reglamentos federales relacionados con el programa de Ayuda a Familias con Hijos Dependientes (AFDC, por sus siglas en inglés) y de apoyo al trabajo, entre otros.

A finales de la década de 1980 y principios de la siguiente, regresó parcialmente el financiamiento para evaluación y, consecuentemente, algunos organismos prosperaron. Por ejemplo, Ginsberg, McLaughlin, Pusko y Takai escribieron sobre la "revitalización" de la evaluación en el Departamento de Educación,[24] pero otras instancias permanecieron inactivas.[25] En general, la evaluación mantuvo su lugar dentro de la burocracia y se iniciaron nuevos esfuerzos importantes. Durante la presidencia de Clinton, se implementaron más programas sociales y se efectuaron más evaluaciones. No obstante, la revolución republicana, que comenzó con las elecciones de 1994, requería una reducción masiva del gobierno federal y el traslado de muchos programas sociales a los estados.

En el pasado, tanto conservadores como liberales consideraban la evaluación como una herramienta útil. Cuando gobiernan los conservadores, el énfasis de la evaluación tiende

[24] Véase Alan Ginsberg *et al.*, "Reinvigorating program evaluation at the U.S. Department of Education", *Education Researcher*, vol. 21, núm. 3, 1992, pp. 24-27.

[25] Véase Chris Wye y Richard Sonnichsen, "Another look at the future of program evaluation in the federal government: Five reviews", *Evaluation Practice*, vol. 13, núm. 3, 1992, pp. 185-195.

a dirigirse al recorte de gastos de los programas (¿el programa es coherente respecto a los criterios de eficiencia y reducción de costos?, ¿en qué medida?) y a la eliminación de servicios para beneficiarios inelegibles. En cambio, cuando los liberales están en el poder, los estándares tienden a enfocarse en la efectividad del servicio en términos del mejoramiento de las oportunidades de vida de los beneficiarios.

Con todos los avances y retrocesos en lo federal, una de las características notables de la historia reciente de la evaluación es el aumento de la actividad evaluativa estatal e incluso local. Por ejemplo, la Mancomunidad de Massachusetts evaluó su programa de educación y capacitación para los beneficiarios de la asistencia social;[26] el estado de Carolina del Sur evaluó un proyecto de integración de servicios humanos;[27] la ciudad de Chicago, junto con universidades locales y grupos de reforma ciudadana, evaluó la reforma escolar más importante de la ciudad.[28]

Recientemente, también se ha observado la tendencia a aumentar el uso de métodos cualitativos de evaluación. Hace poco tiempo, la evaluación cuantitativa representaba el único tipo de evaluación con legitimidad profesional —por lo menos, en el discurso—. Preferiblemente, se utilizaba un diseño experimental de asignación aleatoria.[29] Con todo, algunas evaluadoras

[26] Véanse los informes del Massachusetts Department of Public Welfare, *An analysis of the first 25 000 ET placements* (1986); *An evaluation of the Massachusetts employment and training choices program: Interim findings on participation and outcomes* (1989); *Follow-up survey of the first 25 000 ET placements* (1986), Executive Office of Human Services, Boston.

[27] Véase Comisión de Reorganización del Estado de Carolina del Sur, *An evaluation of the human service integration project, 1985-1988*, Columbia, 1989.

[28] Véase Anthony S. Bryk y Sharon G. Rollow, "The Chicago experiment: Enhanced democratic participation as lever for school improvement", *Issues in Restructuring Schools*, vol. 3, 1992, pp. 3-15; Anthony S. Bryk y Sharon G. Rollow, "The Chicago experiment: The potencial and reality of reform", *Equity and Choice*, vol. 9, núm. 3, 1993, pp. 3-15; Anthony S. Bryk *et al.*, "Measuring achievement gains in the Chicago public schools", *Education and Urban Society*, vol. 26, núm. 3, 1994, pp. 306-319.

[29] Véanse los capítulos 8 y 9 [de la obra original].

dependían más de las palabras que de las cifras, recopilaban sus datos mediante la observación y entrevistas informales en vez de utilizar encuestas estructuradas o registros cuantitativos. Sus análisis exploraban el significado del proceso y de los impactos por medio de análisis narrativos.[30] Durante las décadas de 1970 y 1980, irrumpieron en la literatura de la evaluación con numerosos libros y artículos publicados en revistas especializadas. Así, defendieron las ventajas de su enfoque[31] y provocaron un animado intercambio con simpatizantes de los métodos cuantitativos que rápidamente produjo gran revuelo.

Poco después de "las guerras de los paradigmas", como algunos llamaban a esas críticas e intercambios de duras palabras, llegaron los intentos por restablecer las relaciones amistosas. Numerosas personalidades clave de la evaluación llegaron a la conclusión de que ésta abarcaba muchos campos y daba cabida a gran variedad de enfoques. De hecho, los métodos cualitativos y cuantitativos podían complementarse bien entre sí y, en consecuencia, los estudios que empleaban ambos tipos de procedimientos empezaron a proliferar.

De esa trifulca surgió mayor legitimidad del trabajo cualitativo y mayor conciencia de las ventajas de sus técnicas. Cada día más evaluaciones incorporan un componente cualitativo, en especial en materia de educación. Los libros de texto actuales incluyen capítulos sobre los enfoques cualitativos para la evaluación.[32]

[30] Véase el capítulo 11 [de la obra original].

[31] Véanse Robert Bogdan y Sary Knopp Biklen, *Qualitative research for education: An introduction to theory methods*, Allyn & Bacon, Boston, 1982; David M. Fetterman, "A national ethnographic evaluation: An executive summary of the ethnographic component of the Career Intern Program Study", en David M. Fetterman (ed.), *Qualitative approaches to evaluation in education: The silent scientific revolution*, Praeger, Nueva York, 1988, pp. 262-273; Egon G. Guba, *The paradigm dialog*, SAGE, Newbury Park, 1990; Egon G. Guba e Yvvona S. Lincoln, *Fourth generation evaluation*, SAGE, Newbury Park, 1989; Michael Q. Patton, *Qualitative evaluation methods*, SAGE, Beverly Hills, 1980; Ruben E. Stake, *Evaluating the arts in education*, Merril, Columbus, 1975.

[32] Véase el capítulo 11 [de la obra original].

Otro notable avance ha sido la creación de asociaciones profesionales de evaluación, como la Asociación Estadounidense de Evaluación y sus equivalentes de Canadá, Europa, Gran Bretaña, Australia y Nueva Zelanda, entre otras. Dichas asociaciones ofrecen un espacio para que las evaluadoras compartan sus preocupaciones y sus experiencias laborales. En reuniones anuales y por medio de publicaciones, ofrecen oportunidades para que las evaluadoras divulguen sus hallazgos, se actualicen en las nuevas técnicas, consideren temas más amplios relacionados con el papel de la evaluación en la sociedad, diseminen estándares de conducta profesional y, en general, promuevan el avance de la disciplina.

Comparación entre la evaluación y otros tipos de investigación

En evaluación se aplican métodos de investigación de ciencias sociales, tanto cuantitativos como cualitativos. Los principios y la metodología que aplican a todos los demás tipos de investigación se emplean también en la evaluación. Todos nuestros conocimientos sobre diseño, medición y análisis intervienen en la planificación y en la conducción de un estudio de evaluación. El aspecto distintivo de la evaluación no es el método ni el tema de estudio, sino la intención, el propósito para el cual se lleva a cabo.

Diferencias

Utilidad. La evaluación se diseña para utilizarla. Si bien el énfasis de la investigación tradicional es la producción de conocimientos y deja su utilización en manos de los procesos naturales de divulgación y aplicación, la evaluación inicia con *el uso* en mente. En su más sencilla expresión, se lleva a cabo para un cliente que debe tomar decisiones y que recurre a ella

para obtener información que las sustente. Incluso en los casos en que el uso de la evaluación resulta menos directo e inmediato, su utilidad justifica llevarla a cabo.

Preguntas derivadas de un programa. Las preguntas consideradas en una evaluación surgen de las preocupaciones de la comunidad de políticas y programas, es decir, del conjunto de personas que participan o se ven afectadas por un programa. A diferencia del investigador tradicional que formula sus propias hipótesis, la evaluadora se ocupa de lo que concierne al programa. Evidentemente, tiene mucha injerencia en el diseño del estudio y lo aborda desde la perspectiva de sus propios conocimientos y disciplina. Puede elegir cómo plantear y promover la formulación de preguntas, y cómo ejercer control sobre el número de preguntas que puede atender la evaluación sin problemas. Generalmente, pueden incorporar preguntas sobre temas de gran interés para ella, aunque el núcleo del estudio representa cuestiones de interés administrativo y programático.

Calidad de la opinión. Una evaluación tiende a comparar "lo que es" y "lo que debería ser". Aunque, por lo general, el investigador trata de conservar su objetividad, normalmente se ocupa de los fenómenos que podrían demostrar el grado de idoneidad con que funciona el programa y con que se alcanzan los fines buscados. Siempre que se formulan las preguntas de estudio, aparece en algún punto la preocupación de cumplir con estándares explícitos o implícitos. Ese elemento de juicio, y no de criterios, se considera básico para la evaluación y la distingue de muchos otros tipos de investigación.

Escenario de acción. La evaluación se realiza en un escenario de acción, donde lo más importante que sucede es el programa. Un programa sirve al individuo; si existen conflictos entre los requisitos del programa y aquéllos de la evaluación, probablemente se dé prioridad al programa. Frecuentemente, los miembros del personal del programa controlan el acceso a los beneficiarios y también podrían controlar el acceso a registros y archivos. Son responsables de asignar los participantes a las actividades y los lugares donde se lleva a cabo el programa.

Es común que los requisitos de investigación (en cuanto a los datos obtenidos "antes" y para grupos de control) se enfrenten a los procedimientos establecidos de los programas. Esto genera tensión con respecto a qué debe prevalecer, los requerimientos o los procedimientos.

Conflictos de responsabilidades. Frecuentemente, las evaluadoras y los operadores tienen fricciones interpersonales. Los papeles que estos últimos desempeñan y las normas de sus profesiones de servicio tienden a insensibilizarlos a las demandas y a las promesas de la investigación. Desde su punto de vista, el imperativo es el beneficio. Consideran que, probablemente, la contribución de la evaluación no será suficientemente amplia como para justificar las interrupciones y los retrasos. A menudo, los operadores creen firmemente en el valor del programa que ofrecen y consideran que hay poca necesidad de evaluarlo. Más aún, la cualidad de una evaluación de emitir juicios significa que se sopesará el mérito de sus actividades. Desde su perspectiva y en cierto sentido, están bajo escrutinio; si la evaluación produce resultados negativos, si se encuentra que el programa no cumple con los fines deseados, entonces éste, y posiblemente sus empleos, correrán peligro. Por ende, la posibilidad de fricción resulta obvia.

Publicación. La investigación básica se publica y es fundamental divulgarla a la comunidad de investigadores y profesionales. Respecto a la evaluación, quizá nunca se publica la mayoría de los informes. Frecuentemente, los administradores y el personal de programas creen que la información se generó para responder sus preguntas; no sienten la imperiosa necesidad de exhibir sus fallas en público. A veces, las evaluadoras se encuentran tan presionadas por el tiempo o por la impaciencia de comenzar con su siguiente contrato que presentan el informe requerido al organismo correspondiente y proceden a trabajar en un nuevo estudio.

Afortunadamente, las últimas décadas fueron testigos de la apertura de nuevos canales de divulgación. Casi una docena de publicaciones periódicas incluyen actualmente artículos sobre

estudios de evaluación y sus métodos, filosofía y aplicaciones. Dichas publicaciones incluyen *Evaluation Review*, *Educational Evaluation and Policy Analysis*, *Evaluation and the Health Professions*, *New Directions for Evaluation*, *Evaluation Practice* (que ahora se llama *American Journal of Evaluation*), *Evaluation and Program Planning*, *Studies in Educational Evaluation* y *Evaluation: The International Journal of Theory, Research, and Practice*. Estas revistas especializadas y las de campos importantes, como el abuso de sustancias y la criminología, proveen espacios donde las evaluadoras pueden compartir sus resultados y analizar los avances innovadores de su quehacer.

Las publicaciones también otorgan visibilidad a los resultados de los estudios entre personas interesadas en programas y políticas. Para avanzar en la comprensión de cómo conducir mejor los programas, y dónde y cuándo implementar las mejoras, se debe contar con una base acumulativa de información. Sólo mediante las publicaciones se podrá construir sobre los resultados.[33] Incluso cuando los resultados de la evaluación revelan que cierto programa tuvo poco efecto, situación que en general lleva a autores y editores a mostrarse poco dispuestos a publicarlo, es importante que otros se enteren de los hallazgos. Así se evitará que se repitan los programas inefectivos una y otra vez. Cuando un programa produce resultados mixtos, es decir, unos buenos y otros no tan buenos, las personas encargadas de la ejecución de un programa se beneficiarían de saber qué componentes se asociaron con mayor éxito.

Por supuesto, no todo estudio de evaluación merece publicarse. Los que están realizados deficientemente son más engañosos que útiles, y si la evaluadora analizó los temas de forma tan concreta y específica que los resultados no se pueden generalizar más allá del proyecto inmediato, habrá poco que informar a los demás. Para evitar tales limitaciones, las evaluadoras necesitan tener en mente las necesidades de públicos

[33] En esta era de avances tecnológicos, se desarrollan servicios computarizados para complementar lo publicado en revistas especializadas.

amplios al momento de desarrollar sus planes de trabajo. Así, los informes publicados pueden añadirse al inventario de conocimientos de programas.

Lealtad. La evaluadora tiene una alianza doble o quizá triple. Tiene obligaciones con la organización que financia el estudio; entre ellas, entregar un informe de alta calidad e incorporar el mayor grado posible de utilidad traducible en acciones. Busca que su informe sea útil a diseñadores de políticas, directores, profesionales y participantes de programa. Más que hacia una organización específica, es responsable de contribuir al mejoramiento de los programas en su área de especialidad (educación en ciencias, reglamentación de armas de fuego, etc.). Sin importar si la organización apoya las conclusiones del estudio, la evaluadora a menudo se siente obligada a trabajar en la aplicación de las conclusiones para el avance del programa y de las políticas en ese campo. En ambos escenarios, tiene compromisos con la práctica, con el desarrollo de conocimientos y con su profesión. Como científica social, busca expandir las fronteras de los conocimientos relacionados con la forma en que una intervención afecta las instituciones y la vida humana.

Si bien algunas de las diferencias que hay entre la investigación en evaluación y la investigación social más académica han producido la percepción de que el destino de la evaluadora es demasiado severo, también existen compensaciones. Entre ellas, la más gratificante es la oportunidad de participar activamente en la convergencia de conocimientos científicos y de acción social. Además, tiene la oportunidad de contribuir al mejoramiento de programas sociales. Es esa ventaja la que ha atraído a tantos investigadores capaces al campo de la evaluación, a pesar de las limitaciones de su práctica.

Similitudes

Por otra parte, hay importantes similitudes entre la evaluación y otros tipos de investigación. Al igual que otros tipos de

estudio, la evaluación busca *a)* describir, *b)* entender las relaciones entre las variables y *c)* delinear la secuencia causal entre una variable y otra. Como la evaluación estudia un programa que interviene en la vida de las personas con el propósito de producir cambios, a veces puede efectuar inferencias directas sobre los vínculos causales que van del programa hacia el efecto.

Al igual que otros investigadores, las evaluadoras emplean toda la gama de métodos de investigación existentes para recopilar información: entrevistas, cuestionarios, pruebas de conocimientos y habilidades, inventarios de actitudes, observación, análisis de contenido de documentos, registros e inspección de evidencia física, entre otros. Las evaluadoras ingeniosas pueden encontrar maneras adecuadas de explorar gran variedad de procesos y efectos. Por lo tanto, el esquema de recolección de datos que se usará dependerá de la información requerida para responder preguntas específicas planteadas por la evaluación.

El diseño clásico utilizado en una evaluación es el experimento aleatorio, que implica la medición de variables relevantes para un mínimo de dos grupos equivalentes: uno que se haya expuesto al programa y otro sin exposición. Con todo, se emplean muchos otros diseños experimentales en la investigación en evaluación: estudios de caso, encuestas efectuadas después de la finalización de un programa, series de tiempo, estudios de correlación, etcétera.

No existe una fórmula definitiva para ofrecer a las evaluadoras el "mejor" método o el más adecuado para efectuar sus estudios. Los programas y los organismos con que trata una evaluadora resultan tan diversos y multifacéticos que los aspectos específicos de cada caso en particular ejercen influencia significativa. Muchas cosas dependen de los usos que el estudio tendrá: las decisiones inminentes y las necesidades de información de los tomadores de decisiones o las incertidumbres del campo y la necesidad de entender mejor cómo funcionan los programas. Asimismo, y desgraciadamente, otro tanto depende de las limitantes del escenario del programa: de los límites impuestos al estudio por las realidades de tiempo, por

la ubicación y por las personas. El dinero también es un problema. Los libros de texto rara vez mencionan la sórdida cuestión del financiamiento, pero los fondos limitados imponen restricciones inevitables sobre cuánto se puede estudiar a lo largo de qué periodo. Si la evaluadora está en la nómina de la instancia que ejecuta el programa, podría haber límites a su libertad de explorar aspectos o impactos negativos. Por lo tanto, los métodos de evaluación frecuentemente representan un conflicto entre lo ideal y lo viable.

En ocasiones, y en particular dentro de círculos académicos, la evaluación se considera investigación de menor importancia que la "tradicional" o "pura". A veces, las evaluadoras se perciben como profesionales aburridas de la comunidad de investigación. Se piensa que trabajan hasta el agotamiento en cuestiones poco interesantes y comprometen su integridad en el mundo corrupto. Sin embargo, cualquier evaluadora activa les indicará fervientemente que la evaluación requiere de habilidades más elevadas que las de un estudio diseñado y ejecutado bajo el control del investigador. Se requiere habilidad —y valor— para lograr que la investigación resulte de utilidad sin perder rigor; deben lidiar con las complejidades de personas reales en programas reales, ejecutados por organizaciones también reales.

La evaluadora debe contar con conocimientos vastos sobre la formulación de la pregunta de investigación, el diseño experimental, el muestreo, la recopilación de datos, el análisis y la interpretación. Debe encontrarse al tanto de lo publicado en los textos de metodología de investigación. Después, es preciso que aprenda a aplicar esas enseñanzas en un escenario que frecuentemente dirige su hostilidad hacia áreas importantes de sus conocimientos. Si persiste en adherirse estrictamente a las normas prescritas, corre varios riesgos: que su labor resulte irrelevante para las necesidades del organismo que la contrata, que cause molestias en el personal del programa con quien trabaja y que ignoren sus resultados, si acaso llegara a concluir el estudio. Por lo tanto, a veces tiene que encontrar formas alternas de conducir el estudio. Al mismo tiempo, debe estar lista

para defender hasta la muerte los elementos que no puedan menoscabarse; debe hacer todo sin disminuir la calidad de la investigación. Finalmente, requiere habilidades bien afinadas para divulgar los resultados del estudio. El fin es promover y apoyar la aplicación de los hallazgos en el mejoramiento de la política y el programa.

Resumen

En este texto, se define la evaluación como la valoración sistemática de la operación y/o de los impactos de un programa o política mediante su comparación con estándares explícitos e implícitos para ayudar a mejorar dicho programa o política. Una evaluación se efectúa cuando el programa es complejo, difícil de observar y consta de elementos que interactúan de numerosas maneras; cuando se deben tomar decisiones importantes y costosas, y cuando se requieren evidencias para convencer a otros de los méritos o las fallas del programa. Asimismo, la evaluación puede ser una herramienta de rendición de cuentas; los patrocinadores y operadores de un programa pueden emplear la evidencia obtenida de la evaluación para elaborar informes sobre sus procedimientos e impactos, que divulgarán a un público amplio.

La evaluación se puede dirigir al proceso del programa (es decir, la forma en que se implementa) y puede examinar sus impactos (es decir, las consecuencias del programa para sus participantes).

La evaluación forma parte de una tendencia de largo plazo que busca la racionalidad científica en la toma de decisiones, y se puede rastrear hasta el siglo XVII. A mediados del XIX, algunos estudios efectuados en Francia calcularon el efecto de la educación en las tasas de criminalidad, y el valor de canales y caminos. A principios del siglo XX, un cirujano estadounidense evaluó el desempeño de sus colegas: determinó el estado de salud de los pacientes después de la cirugía. Un estudio trascendental en la

historia de la evaluación fue el análisis de las consecuencias del proyecto de Cambridge-Somerville de trabajadores jóvenes, dirigido a adolescentes en riesgo de convertirse en delincuentes. En las décadas de 1970 y 1980, la Guerra contra la Pobreza dio gran impulso a la evaluación; el gobierno empezó a financiar cientos de evaluaciones sobre educación, justicia penal, servicios sociales, asistencia jurídica, organización comunitaria, atención a la salud, desarrollo internacional, servicios de salud mental y nutrición, entre otros. En la década de 1970, el gobierno apoyó varios experimentos sociales de gran escala para ensayar la viabilidad de las ideas de políticas antes de su adopción en el ámbito nacional. Otro momento notable fue la realización del experimento del impuesto negativo al ingreso.

La evaluación se distingue de otros tipos de investigación por distintas razones: los diseñadores de políticas y los profesionales proveen sus preguntas medulares, el uso de los resultados se dirige al mejoramiento de programas, se ubica en un turbulento escenario de acción y elabora informes para públicos que no están dedicados a la investigación. La evaluadora tiene obligaciones con la organización que financia el estudio, con el amplio campo de programas (educación y rehabilitación física, por mencionar algunos) y con las ciencias sociales. Entre las similitudes que exhibe con otros tipos de estudio se encuentran las técnicas de investigación que utiliza y el afán de entender y encontrar explicaciones.

Bibliografía

Bogdan, Robert y Sary Knopp Biklen, *Qualitative research for education: An introduction to theory methods*, Allyn & Bacon, Boston, 1982.

Bryk, Anthony S. y Sharon G. Rollow, "The Chicago experiment: Enhanced democratic participation as lever for school improvement", *Issues in Restructuring Schools*, vol. 3, 1992, pp. 3-15.

——, "The Chicago experiment: The potencial and reality of reform", *Equity and Choice*, vol. 9, núm. 3, 1993, pp. 3-15.
Bryk, Anthony S. *et al.*, "Measuring achievement gains in the Chicago public schools", *Education and Urban Society*, vol. 26, núm. 3, 1994, pp. 306-319.
Bulmer, Martin *et al.*, *The social survey in historical perspective*, Cambridge University Press, Nueva York, 1991.
Cain, Glen C. y Harold Watts, *Income maintenance and labor supply: Econometric studies*, Rand McNally College Publishing Co., Chicago, 1973.
Carlson, David y John D. Heinberg, *How housing allowances work: Integrated findings from the experimental housing allowances program*, Urban Insitute, Washington, D. C., 1978.
Comisión de Reorganización del Estado de Carolina Del Sur, *An evaluation of the human service integration project, 1985-1988*, Columbia, 1989.
Cronbach, Lee *et al.*, *Toward reform of programme evaluation*, Jossey-Bass, San Francisco, 1980.
Cullen, Michael J., *The statistical movement in early Victorian Britain: The foundations of empirical social research*, Harper & Row, Nueva York, 1975.
Massachusetts Department of Public Welfare, *An analysis of the first 25 000 ET placements*, Executive Office of Human Services, Boston, 1986.
——, *An evaluation of the Massachusetts employment and training choices program: Interim findings on participation and outcomes*, Executive Office of Human Services, Boston, 1989.
——, *Follow-up survey of the first 25 000 ET placements*, Executive Office of Human Services, Boston, 1986.
Darman, Richard, "Riverboat gambling with government", *New York Times Magazine*, 1 de diciembre de 1996, pp. 116-117.
Dwyer, Robert, *Report to the people*, vol. 14, núm. 1, Nueva Jersey, 1970.
Fetterman, David M., "A national ethnographic evaluation: An executive summary of the ethnographic component of

the Career Intern Program Study", en David M. Fetterman (ed.), *Qualitative approaches to evaluation in education: The silent scientific revolution*, Praeger, Nueva York, 1988, pp. 262-273.

Flook, Evelyn y Paul J. Sanazaro (eds.), *Health services research and R&D in perspective*, Health Administration Press, s. l., 1973.

Friedman, Joseph y Daniel H. Weinberg (eds.), "The great housing experiment", *Urban Affairs Annual Review*, vol. 24, 1983.

Ginsberg, Alan *et al.*, "Reinvigorating program evaluation at the U.S. Department of Education", *Education Researcher*, vol. 21, núm. 3, 1992, pp. 24-27.

Green, Lawrence W. y Marschall W. Kreuter, *Health promotion planning: An educational and environmental approach*, Mayfield Publishing, Mountain View, 1991.

Guba, Egon G., *The paradigm dialog*, SAGE, Newbury Park, 1990.

Guba, Egon G. e Yvvona Lincoln, *Fourth generation evaluation*, SAGE, Newbury Park, 1989.

Hyman, Herbert y Charles Wright, "Evaluating social action programs", en William Hamilton Sewell *et al.* (eds.), *The uses of sociology*, Basic Books, Nueva York, 1967.

Kennedy, Stephen D., *The final report of the housing allowance demand experiment*, ABT Associates, Cambridge, 1980.

Kershaw, David y Jerilyn Fair, *The New Jersey income-maintenance experiment: Operations, surveys and administration*, Academic Press, Nueva York, 1976-1977, 2 ts.

McCord, William Maxwell y Joan McCord, *Origins of crime: A new evaluation of the Cambridge-Sommerville youth study*, Columbia University Press, Nueva York, 1959.

McLaughlin, Milbrey Wallin, *Evaluation and reform: The elementary and secondary education Act of 1965, Title 1*, Ballinger, Cambridge, 1975.

Newhouse, Joseph *et al.,* "Some interim results from a controlled trial of cost sharing in health insurance", *New England Journal of Medicine*, 1981.

Newhouse, Joseph e Insurance Experiment Group, *Free for all? Lessons from the Rand health insurance experiment*, Harvard University Press, Cambridge (Mass.), 1993.

Orleans, Tracy C. *et al.*, "Self-help quit smoking interventions: Effects of self-help materials, social instructions, and telephone counseling", *Journal of Consulting and Clinical Psychology*, vol. 59, núm. 3, 1991, pp. 439-448.

Ossip-Klein, Deborah *et al.*, "Effects of a smokers' hot line: Results of a 10-county self-help trial", *Journal of Consulting and Clinical Psychology*, vol. 59, núm. 2, 1991, pp. 325-332.

Patton, Michael Q., *Qualitative evaluation methods*, SAGE, Beverly Hills, 1980.

Phelps, Charles E. y Joseph Newhouse, *Coinsurance and the demand for medical services*, Rand Corporation, Santa Mónica, 1973.

Powers, Edwin y Helen Leland Witmer, *An experiment in the prevention of delinquency: The Cambridge-Sommerville youth study*, Columbia University Press, Nueva York, 1951.

Rivlin, Alice y P. Michael Timpane, *Planned variation on education: Should we give up or try harder?*, Brookings Institution, Washington, D. C., 1975.

Schweinhart, Lawrence *et al.*, *Significant benefits: The High/Scope Perry preschool study through age 27*, High/Scope Press, Ypsilanti, 1993.

Stake, Ruben, *Evaluating the arts in education*, Merril, Columbus, 1975.

Toulemonde, Johnny y Lise Rochaix, "Rational decision-making through project appraisal: A presentation of French attempts", *International Review of Administrative Sciences*, 1994, pp. 37-53.

Tyler, Ralph, "General statement on program evaluation", en Mildrey Wallin McLaughlin y Brenda D. Phillips (eds.), *Nineteenth yearbook of the National Society for the study of Education*, Chicago Press, Chicago, 1991, t. 2.

Vibbert, Spencer, *What works: How outcomes research will change medical practice*, Ground Rounds Press, Knoxville, 1993.

Viswesvaran, Chockalingam y Franck L. Schmidt, "A meta-analytic comparison of the effectiveness of smoking cessation methods", *Journal of Applied Psychology*, vol. 77, núm. 4, 1992, pp. 554-561.

Wye, Chris y Richard Sonnichsen, "Another look at the future of program evaluation in the federal government: Five reviews", *Evaluation Practice*, vol. 13, núm. 3, 1992, pp. 185-195.

Zelman, Diane *et al*., "Measures of affect and nicotine dependence predict differential response to smoking cessation treatments", *Journal of Consulting and Clinical Psychology*, vol. 60, núm. 6, 1992, pp. 943-952.

En las últimas décadas, se han fortalecido los espacios profesionales para la evaluación de políticas y programas públicos, y se ha abierto el espacio político para canalizar la información que ésta produce. Este intercambio se puede nutrir significativamente con el acceso a investigaciones, tanto teóricas como aplicadas, que hasta ahora han sido referentes en Estados Unidos y Europa, pero que no se han difundido en español.

Antología sobre evaluación contiene diez traducciones de artículos fundamentales para el desarrollo de la evaluación como disciplina. Su objetivo principal es contribuir a la socialización y difusión de un lenguaje compartido entre la comunidad de profesionales de la evaluación y los interesados en ella: estudiantes de ciencias sociales, en licenciaturas y posgrados, así como evaluadores en proceso de formación.

El lector podrá hallar aquí artículos de Eleanor Chelimsky, Huey-Tsyh Chen, Howard E. Freeman, Jennifer Greene, Gary T. Henry, Mark W. Lipsey, Melvin M. Mark, Michael Q. Patton, Peter H. Rossi, Jean-Claude Thoenig y Carol H. Weiss.

ISBN 978-607-8367-61-9
9 786079 367619

GOBIERNO Y POLÍTICAS PÚBLICAS es una colección que busca estimular el debate intelectual sobre administración pública, con textos que den cuenta de la agenda renovada en las democracias latinoamericanas, que actualicen diagnósticos, enfoques y métodos de análisis, y que contribuyan a la construcción de una comunidad académica hemisférica. Ofrece textos de académicos hispanohablantes y algunas traducciones seleccionadas, que sirvan para el estudio, la enseñanza y la discusión práctica sobre asuntos públicos.

CIDE

clear
Regional Centers for
Learning on Evaluation and Results