

Oswaldo Feinstein*

EVALUACIÓN PRAGMÁTICA DE POLÍTICAS PÚBLICAS

Este artículo presenta un enfoque pragmático para la evaluación de políticas públicas. Comienza con una descripción de la evaluación en la que se discuten sus funciones y modalidades, continúa con una presentación sobre cómo la evaluación permite aprender de la experiencia, se complementa con la exposición de los criterios de evaluación, considera el problema de la atribución, presenta dos enfoques para la evaluación de políticas públicas utilizando los conceptos introducidos en las secciones precedentes, aborda las dificultades que se enfrentan en la evaluación de políticas públicas y concluye con un conjunto de soluciones posibles y diez conclusiones pragmáticas.

Palabras clave: *economía pública, desempeño del sector público, evaluación de políticas.*

Clasificación JEL: *H11, H83, P35.*

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 eva-

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

Versión de noviembre de 2006.

luaciones, sobre todo cuando hay prisa por aprobar políticas como señales o indicaciones de que se está haciendo algo en relación a 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.

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.

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 sólo 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 del mismo).

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 sólo 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 Ockam», 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 aún antes de que se complete la intervención y que no es afectado por problemas de atribución, algo que será considerado más adelante).

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.

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? Éste 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, asig-



narse 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.

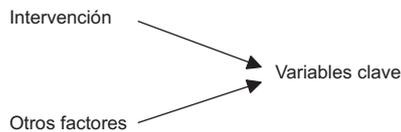
La relación entre relevancia, eficacia y eficiencia puede mostrarse mediante el diagrama siguiente: la relevancia examina los objetivos a la luz de la estrategia; la eficacia corresponde al grado de logro de los objetivos, y la eficiencia es la relación entre los niveles de logros alcanzados y los recursos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

ESQUEMA 2

MÚLTIPLES FACTORES CAUSALES



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 al 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 sólo 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).

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 (aquéllas que interesa influenciar, a veces llamadas variables de resultado).

Es útil plantear cuatro preguntas relacionadas:

- ¿Cómo identificar el efecto de la intervención X en las variables clave?
- ¿Cómo aislar o separar dicho efecto del correspondiente a otros factores?
- ¿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 sólo 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 aquéllos que han sido objeto de la intervención y otros cuyos integrantes difieren del primero sólo en que no han sido afectados por la intervención⁹. De poder establecerse un grupo de control, algo que discutiremos

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 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).

ESQUEMA 3

**GRUPOS CON Y SIN INTERVENCIÓN,
ANTES Y DESPUÉS DE LA INTERVENCIÓN**

	Con intervención	Sin intervención
Antes	CA	SA
Después	CD	SD

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 \$.

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 la misma. Se tienen así cuatro casos que corresponden a la situación antes y después de la intervención, y con/sin intervención (ver Esquema 3).

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).

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 período 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 fon-

¹⁰ Como ejemplos véanse los trabajos citados en PITMAN, G. K. *et al.* (2005) y WEISS, C. (1982).

dos suficientes y desarrollados por gente con la pericia necesaria para asegurar que dichos sistemas procedan a la recopilación de datos durante el período 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 los mismos pueden aportar una clarificación de la atribución¹³. Consecuentemente, 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 expondrá 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 que se encara la evaluación de fármacos,

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 sólo a las predicciones, cuando se juzgan las teorías económicas.

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 post-programas de personas que participan en programas de formación con las medidas preprogramas para dichas personas (procedimiento utilizado frecuentemente) sobrestiman sustancialmente las ganancias derivadas de la participación en la formación ya que buena parte de la mejora post-programa 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 período 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 progra-

mas 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

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.

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?	<i>Ex ante</i> , tiempo real, <i>ex post</i>
5. Cobertura	¿Cuánto evaluar?	Costes y beneficios Nivel óptimo
6. Métodos	¿Cómo evaluar?	«Tratamientos» Criterios: relevancia, eficacia, eficiencia
7. Uso	Factores que influyen en el uso	Oportunidad, credibilidad, disseminación

deberá realizar una reconstrucción (o interpretación) de los mismos 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 los mismos 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 dispo-

¹⁹ Estas cinco dimensiones pueden recordarse fácilmente utilizando un recurso mnemotécnico: las iniciales de las palabras subrayadas componen el nombre de ICARO, el personaje de la mitología griega que trató de volar hacia el sol.

nible 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).

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 los mismos), 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 éste 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 los mismos. Resulta útil distinguir entre la oferta actual o real de evaluadores, incluyendo instituciones, y la oferta potencial, incluyendo en ésta a uni-

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).

versidades 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 las mismas 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

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).

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 aquéllas que consisten o se componen de intervenciones muy específicas (programas o proyectos) que no varían durante el período 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 crea-

ción 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, *economies 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.

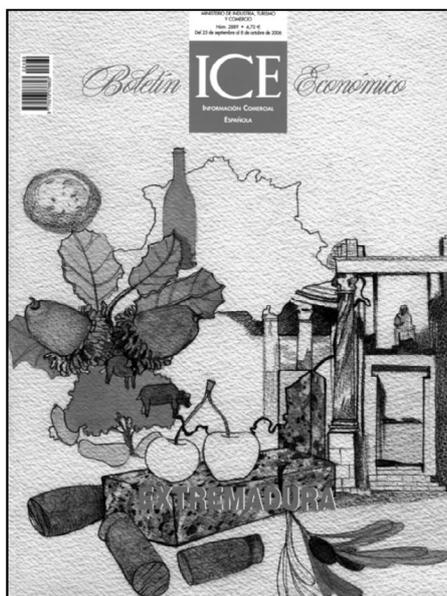
- Los evaluadores y los gestores de evaluaciones deben ocuparse no solamente de la producción sino también de la postproducción de evaluaciones, asumiendo una responsabilidad por facilitar su utilización. La división de trabajo que involucre a expertos en comunicación puede ser útil para que las evaluaciones 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.

Referencias bibliográficas

- [1] BANDURA, A. (1982): *Teoría del Aprendizaje Social*, Espasa Calpe, Madrid.
- [2] ELSTER, J. (2006): *Lógica y Sociedad*, Gedisa, Barcelona.
- [3] FEINSTEIN, O. (2002): «Use of Evaluations and the Evaluation of their Use», *Evaluation*, 8, 4, 433-439.
- [4] FEINSTEIN, O. (2006): «Evaluation of Development Interventions and Humanitarian Action», con TONY BECK, en SHAW, I. et al. (eds.): *Handbook of Evaluation*, Sage, Londres.
- [5] FEINSTEIN, O. y PICCIOTTO, R. (eds.) (2001): *Evaluation and Poverty Reduction*, Transaction, New Brunswick.
- [6] FRANCO, R. y LANZARO, J. (2006): *Política y Políticas públicas en los procesos de reforma en América Latina*, FLACSO/Miño y Dávila, Buenos Aires.
- [7] 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.
- [8] HECKMAN, J. (2001) en FEINSTEIN, O. y PICCIOTTO, R. (eds.) (2001): *Evaluation and Poverty Reduction*, Transaction, New Jersey, páginas 83-87.
- [9] LEEUW, F.; RIST, R. y SONNICHSEN, R. C. (eds.) (1993): *Can Governments Learn?*, Transaction, New Brunswick.
- [10] PITMAN, G. K.; FEINSTEIN, O. e INGRAM, G. K. (eds.) (2005): *Evaluating Development Effectiveness: Challenges and the Way Forward*, Transaction Publishers, New Brunswick.
- [11] RAMIRO SÁNCHEZ, A. (2002): *Demandas de calidad de la Administración Pública: Un derecho de la ciudadanía*, Dykinson, Madrid.
- [12] 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.
- [13] STERN, E. (2006): *Contextual Challenges for Evaluation Practice*, en SHAW, I. et al. (2006).
- [14] WEISS, C. (1982): *La investigación evaluativa*, Trillas, México.

ÚLTIMOS MONOGRÁFICOS PUBLICADOS



Información:
Ministerio de Industria, Turismo y Comercio
Paseo de la Castellana, 162-Vestíbulo
28071 Madrid
Teléf. 91 349 36 47