

Emma García García*

EVALUACIÓN DE LAS POLÍTICAS PÚBLICAS PARA EL DESARROLLO: EFECTIVIDAD, EFICIENCIA Y RENTABILIDAD

Este artículo describe un marco general de evaluación aplicable a programas y políticas públicas, en especial aquellas destinadas a promover el desarrollo y la investigación en el campo de la educación. En primer lugar, se ofrece un análisis de qué significa y cómo se evalúa la efectividad de los programas. Se repasan aspectos tales como la estimación de impacto, qué otras dimensiones del impacto deben ser evaluadas, y otras cuestiones relacionadas con algunos de los retos que la inferencia causal plantea. Se explican también las ventajas de incorporar análisis relacionados con los costes y el marco de evaluación y en particular, los objetivos de los análisis coste-efectividad y coste-beneficio, así como las características principales de sus metodologías y limitaciones. El documento muestra cómo los resultados de estos estudios se pueden combinar con el fin de proporcionar orientación valiosa al diseño y gestión de las políticas públicas; las ventajas derivadas de la cooperación entre quienes se encargan de la formulación de las políticas, quienes las ponen en práctica y quienes se dedican a la investigación, a fin de maximizar el impacto y alcance de nuestras políticas públicas en general.

Palabras clave: análisis de costes, evaluación, inferencia causal, investigación en educación, políticas públicas.
Clasificación JEL: D6, H4, I2, I3.

1. Introducción

Todos los agentes implicados en la toma de decisiones en materia de políticas públicas, estarían de acuerdo

sobre la creciente importancia atribuida a políticas y programas que están respaldados por resultados basados en la evidencia empírica¹. Este acuerdo existiría entre los profesionales cuyo trabajo está conectado a políticas

* Economic Policy Institute y Georgetown University, McCourt School of Public Policy.

La autora agradece las orientaciones y comentarios de Vicente Montes Gan. También mencionar que los materiales utilizados para la elaboración del artículo incluyen material docente desarrollado conjuntamente con Ilya Cornelisz. Agradezco igualmente sus consejos, que he tenido en cuenta para la versión final de este trabajo.

¹ En inglés, este término se denomina *evidence-based policy*. Un enfoque y gestión de políticas basadas en la evidencia es aquel que se refiere a un proceso político que ayuda a los planificadores a tomar decisiones bien documentadas sobre políticas, programas y proyectos, al proveerse de la mejor y más precisa evidencia, y hacer de esta el punto central del desarrollo e implementación de las políticas o del proceso político (DAVIES, 1999; UNICEF, 2008).

microeconómicas o de nivel individual (por ejemplo, programas de formación para jóvenes con el objetivo de mejorar su empleabilidad, actividades extraescolares para el desarrollo de las habilidades cognitivas y no-cognitivas de los niños, o disminuciones en los impuestos a las rentas de capital para incentivar el ahorro a largo plazo), así como entre las personas cuya actividad tiene como finalidad el influir en decisiones globales o a nivel macroeconómico (por ejemplo, políticas a nivel nacional relacionadas con la reducción de la pobreza, el desempleo o la ayuda internacional, etcétera). Esta tendencia y estrategia persiguen priorizar programas que han sido rigurosamente evaluados, así como aumentar las evaluaciones disponibles analizando programas que aún no lo han sido. Como consecuencia de este enfoque, nuestro conocimiento sobre los efectos causales de los programas está aumentando, o dicho de otro modo, contamos con mejores mecanismos para identificar los programas que funcionan, es decir que tienen un impacto, y para cuantificar tales efectos. Además, otra contribución importante de estas evaluaciones consiste en que pueden contribuir a informar de los mecanismos a través de los cuales un programa funciona. Finalmente, una evaluación global de un programa permite evaluar el impacto del programa por unidad de coste, el coste por participante, el retorno de la inversión y los beneficios netos del programa (Evans y Ghosh, 2008).

La combinación de todos los aspectos contenidos en la evaluación de los programas mencionados conduce a la medición del impacto de un programa, a si los recursos económicos se han utilizado de manera eficiente y a si la intervención tendrá tasas positivas de retorno. Este proceso de pensamiento tan lógico se asienta en el trabajo desarrollado por la comunidad científica durante varias décadas (eminentemente en el campo de las ciencias sociales), que ha contribuido al desarrollo de métodos analíticos relativamente sofisticados capaces de producir tales respuestas. Estos métodos son conocidos como métodos de inferencia causal, análisis coste-efectividad y análisis coste-beneficio (cuatro referencias seminales al respecto son Angrist y Pischke, 2008; Levin y

McEwan, 2001; Murnane y Willett, 2011 y Shadish, Cook y Campbell, 2002).

Es importante destacar que el desarrollo y la utilización de estos métodos se entienden desde su asociación y mejora de teorías, así como desde el incremento en la cantidad y calidad de datos disponibles en el campo de las ciencias sociales. A partir de las estrechas relaciones entre la investigación teórica, aplicada y la disponibilidad de datos, estamos presenciando una renovación profunda en nuestra comprensión de la orientación y el diseño de políticas, su evaluación y la rendición de cuentas por parte de los profesionales implicados en estas tareas. Así, cuando los investigadores, autoridades y profesionales utilizan información fundamentada en la evidencia empírica, su trabajo y sus conclusiones no solo son más lógicos y razonados, sino que redundan en mejores tomas de decisiones alrededor de las políticas públicas, y consiguen aumentar la transparencia y la rendición de cuentas sobre las consecuencias de sus decisiones.

Este documento proporciona una revisión del marco general de evaluación de programas y políticas públicas. Si bien el objetivo es demostrar la viabilidad e incluso necesidad de evaluar las políticas públicas en materia de desarrollo, tanto la conceptualización como la mayoría de los ejemplos utilizados en este artículo provienen de la disciplina de políticas públicas en general, y más concretamente, de la investigación en educación y desarrollo humano. Cabe señalar que el marco que se describe en este artículo puede adaptarse a otros ámbitos y a otros niveles, incluida la evaluación de las políticas micro y macroeconómicas en ayuda al desarrollo. En el siguiente apartado, el artículo ofrece un análisis de las cuestiones relacionadas con el impacto o efectividad de programas², incluyendo los distintos aspectos que deben ser evaluados,

² Para un análisis de los orígenes de la evaluación, consulte *Información Comercial Española. Revista de Economía*, número 836, 2007: «Evaluación de políticas públicas». A pesar de que la mayoría de los artículos incluidos en ese número son relevantes para el presente artículo, este se centra en revisar el marco de evaluación necesaria para proporcionar una evaluación completa de los impactos y consecuencias de los programas, siendo los contenidos complementarios a la obra publicada en 2007.

los métodos, los requerimientos de datos y algunas de las limitaciones de este tipo de análisis. En el apartado 3 se revisan los métodos o análisis de costes, y se explica por qué los estudios relacionados con costes se están convirtiendo en una herramienta cada vez más exigida en la formulación de políticas. En este apartado se explican los objetivos de los estudios coste-efectividad y coste-beneficio, así como sus características metodológicas y sus limitaciones, y cómo sus resultados pueden ser de valor para la orientación de las políticas públicas. Por último, en el apartado 4 se exponen las conclusiones.

2. El estudio de los efectos o impactos de un programa

En este apartado se describe el marco de evaluación utilizado para estimar el impacto o efecto de políticas, programas o intervenciones, conocido como evaluación de programas. En términos generales, el concepto de efecto de un programa refleja el cambio producido por la intervención en los participantes o beneficiarios del programa. Técnicamente hablando, el impacto de un programa es la diferencia entre lo ocurrido como consecuencia del programa y lo que habría sucedido en ausencia del mismo (Angrist y Pischke, 2008; Morgan y Winship, 2014 y Murnane y Willett, 2011).

Una investigación detallada de los efectos de un programa o evaluación está compuesta por varios componentes y etapas. En primer lugar, es esencial que la conceptualización y la investigación comiencen con una declaración clara de la cuestión o circunstancia que motivan el proyecto (basada en la evaluación de una determinada necesidad) y, de forma muy importante, de la teoría que enmarca el funcionamiento del programa (adaptado de Murnane y Willett, 2011, 27). Tal y como Murnane y Willett explican, la teoría informa de «las cuestiones que deben abordarse, la población de interés, la naturaleza de las intervenciones, los resultados que esperamos que una determinada intervención pueda influir o afectar, así como los mecanismos mediante los cuales los impactos se producirán», lo

que, en otras palabras, puede denominarse la teoría del cambio completa (Murnane y Willett, 2011, 333). Finalmente, vinculados a la ejecución de los programas, la evaluación del proceso y su control son componentes importantes, y pueden también informar tanto de los efectos como de los mecanismos que los originan.

El impacto o efecto causal de un programa es el elemento central de la evaluación de programas. Por ejemplo, si un programa ofrece transferencias en efectivo a las familias, a fin de que los adolescentes se inscriban en las escuelas de educación secundaria, el impacto del programa se reflejará en el cambio, en la proporción de los adolescentes matriculados, o quizás en el incremento en el número de graduados al final de la etapa secundaria (o de hecho, con cualquier otra relacionada con el objetivo del programa). Si un programa ofrece apoyo a los adolescentes en su rendimiento en matemáticas, en función de si el programa brinda capacitación, apoyo personal o ambos, podríamos determinar el impacto atendiendo a los cambios en el rendimiento en un determinado examen de matemáticas o, también, en modificaciones en la puntuación en un índice que mida la confianza de los estudiantes en sus habilidades en el campo de las matemáticas. En términos generales, la unidad de medida en la que el impacto del programa se expresa está vinculada al propósito u objetivo específico del programa. A la hora de convertir esta idea en algo más específico y propio de cada programa, la pregunta de investigación, la teoría cambio y el objetivo del programa son quienes determinan quién es la variable dependiente y por tanto, la definición y unidad de medida de su impacto.

El principal reto de la evaluación de impactos radica en la identificación del «factual» del programa, que representa la condición de tratado, y del «contrafactual», o representante de la ausencia de tratamiento. Como puede verse en esta introducción y simple definición, la estimación del efecto de un programa implica, por construcción, que dicha condición de control existe (dado que «el efecto de una causa es siempre

relativo a otra causa», Holland, 1986, 946). Sin embargo, la dificultad en la identificación de la condición de contrafactual proviene del hecho de que no podemos observar los resultados de haber sido sometido al tratamiento y, a la vez, no haber sido sometido al tratamiento, para la misma unidad o persona (esto es a veces conocido como el problema fundamental de la inferencia causal, Holland, 1986, 947). Por lo tanto, es necesario construir un contrafactual válido, o, en otras palabras, una estimación válida de los posibles resultados en ausencia del programa.

Distintas estrategias de identificación han sido desarrolladas con el fin de proveer al factual de un programa con su correspondiente contrafactual, y, por consiguiente, de estimar los efectos causales de los mismos. Los experimentos, que se definen como investigaciones empíricas en las que los niveles de una posible causa son manipulados por un agente externo que funciona de forma independiente a los participantes en la investigación, después de la cual las consecuencias en un resultado determinado son medidas (Murnane y Willett, 2011, 30), constituyen la manera más persuasiva y convincente de abordar cuestiones causales. Siguiendo la presentación de Murnane y Willett, se definen dos tipos de experimentos: experimentos aleatorios y cuasi-experimentos. Los primeros se conocen a veces como el *gold-standard* (o el estándar más elevado) de la inferencia causal, y se definen como aquellos experimentos en los que las unidades se asignan a las condiciones experimentales de tratamiento y control mediante un proceso aleatorio. En los cuasi-experimentos, las unidades no están asignadas al azar a las condiciones de tratamiento o control, pero existe una estrategia de identificación de una fuente exógena de asignación de unidades a los tratamientos. Por ejemplo, pueden basarse en «circunstancias que ocurren naturalmente» (como un cambio en el reglamento de los nacidos después de un momento específico en el tiempo) o en «instituciones que, quizá sin proponérselo, dividen a la población o muestra en los grupos de tratamiento y de control de una manera

similar a una asignación aleatoria» (Angrist, 2003). A pesar de que la estimación de interés seguiría siendo el efecto individual del tratamiento, dado que este no puede ser estimado, como se explicó anteriormente, la evaluación del impacto causal persigue obtener estimaciones confiables del promedio del efecto del programa o del tratamiento, que sería la diferencia entre los resultados promedio de los grupos de tratamiento y de control³.

En el Cuadro 1 se resumen los métodos más importantes utilizados para estimar el efecto causal de un programa. Además de los métodos experimentales aleatorios, los enfoques estadísticos utilizados para realizar inferencias causales con estudios observacionales y datos no-experimentales, son el método de emparejamiento por propensión a participar (o más conocido por el término en inglés, *propensity score matching*), variables instrumentales, regresión discontinua, diferencia en las diferencias, modelos con controles y modelos de efectos fijos. El Cuadro 1 incluye una breve descripción de cada método, una explicación de quién se encuentra en el grupo de control y una mención de los supuestos necesarios para la identificación. La última columna ofrece algunas referencias de análisis empírico utilizando cada una de las técnicas y/o que proporcionan una discusión metodológica de los mismos.

Una vez se ha estimado el impacto del programa, es posible pensar en si dicha intervención podría haber tenido alguna repercusión o podría haber impactado otras dimensiones relacionadas directa o indirectamente con el objetivo del programa. Estas consideraciones son importantes porque pueden ser muy relevantes a la hora de informar sobre la toma de decisiones

³ Los métodos que se describen a continuación pueden producir diferentes estimandos del impacto. Consúltense las referencias en el Cuadro 1 y ANGRIST y PISCHKE (2008) y MURNANE y WILLET (2011), para una explicación de los diferentes efectos del tratamiento (efecto promedio, tratamiento en los tratados, intención de tratar, o efecto de tratamiento promedio local, o en inglés, *average treatment effects, treatment on the treated estimates, intent to treat or local average treatment effect estimates*).

CUADRO 1

DESCRIPCIÓN GENERAL DE LOS MÉTODOS UTILIZADOS PARA ESTIMAR EFECTOS CAUSALES DEL PROGRAMA

	Método	Descripción	Los individuos en el grupo de comparación	Hipótesis necesarias	Ejemplos
Métodos experimentales	Experimento aleatorio	Método experimental para medir la relación causal entre dos variables	Los participantes son asignados aleatoriamente a los grupos de tratamiento y control	Contexto aleatorio, no existe ninguna fuente de selección bajo la asignación a tratamiento y control. Esto garantiza que los grupos de tratamiento y de control son estadísticamente idénticos, en tanto características observables y no observables	(BANERJEE, COLE, DUFLO y LINDEN, 2007; DUFLO, GLENNERSTER y KREMER, 2007 y SCRIVENER <i>et al.</i> , 2015)
	Métodos de puntuación de propensión	Los individuos en el grupo control se comparan a individuos similares o emparejados creados en el grupo experimental.	Los no-participantes que tienen una mezcla de características que permiten predecir que hubieran tenido la misma probabilidad de ser tratados como los participantes.	Los factores que fueron excluidos en el proceso de emparejamiento correspondiente (ya que son no observables y/o no han sido medidos) no sesgan las estimaciones, debido a que no están correlacionadas con los resultados o no difieren entre participantes y no participantes.	(BERGER, HILL y WALDFOGEL, 2005; DEHEJIA y WAHBA, 1999 y HILL, 2008)
Métodos cuasi-experimentales	Variables instrumentales	Participación en un programa puede ser predicha por un factor accidental (casi aleatorio), o variable instrumental, que no está correlacionada con el resultado, más allá de que tal factor predice la participación (y la participación afecta el resultado)	Las personas que, por la intervención de este factor casi aleatorio, fueron asignadas a no participar y (posiblemente como resultado de la asignación) no participaron.	Si no fuera por la capacidad de predecir la participación por parte de la variable instrumental, este «instrumento» no tendría ningún efecto o no estaría correlacionado con los resultados.	(ANGRIST, IMBENS y RUBIN, 1996; ANGRIST y KRUEGER, 1991; CARNEIRO, HECKMAN y VYTLACIL, 2003)
	Regresión discontinua	Los individuos se ordenan según criterios específicos y medibles, y existe un punto de corte que determina si una persona es elegible para participar en un programa. Los participantes son comparados con los no participantes y el criterio de elegibilidad es controlado.	Las personas que están cerca del punto de corte, pero caen en el lado correspondiente a la ausencia de tratamiento o programa.	Después de controlar por los criterios (y otras medidas) de elección o asignación a tratamiento, las diferencias que aún subsisten entre las personas exactamente por debajo del punto de corte y por encima del mismo no son estadísticamente significativas y no desvirtúan los resultados. Un requisito necesario y suficiente es que los criterios de selección se cumplan estrictamente.	(BLOOM, 2009; G.W. IMBENS y LEMIEUX, 2008 y JACOB y LEFGREN, 2004)
	Diferencia en las diferencias	Miden el grado de mejora (o cambio) en el tiempo de los participantes en el programa relativo a la mejora (o cambio) de los no participantes.	Las personas que no participan en el programa (por cualquier razón), pero para los cuales los datos fueron recolectados antes y después del programa.	Si el programa no existiera, los dos grupos que hubieran tenido trayectorias idénticas durante el periodo de estudio.	(CARD y KRUEGER, 2000)
	Análisis de regresión, regresión multivariante con controles	Las personas que reciben el tratamiento se comparan con quienes no reciben el tratamiento, y se controlan todos los factores que podrían explicar las diferencias en los resultados.	En este caso los datos incluyen los indicadores de resultados y las variables de control necesarias que expliquen los resultados.	Las variables no incluidas entre las variables explicativas (porque no eran observables y/o no han sido medidos) no sesgan los resultados debido a que están correlacionadas con los resultados o no difieren entre participantes y no participantes.	(BERGER, PAXSON y WALDFOGEL, 2009 y BLAU, 1999)
	Efectos Fijos (datos de panel)	Las personas se encuentran en los grupos de tratamiento y control en diferentes momentos en el tiempo (a veces dentro de un grupo).	Cada unidad sirve efectivamente como su propio control (cambios en el tiempo de una política). No se puede usar para comparar los niveles de los resultados de tratamiento y de control, solo los cambios en las tendencias.	Una vez se incluyen los efectos fijos, las variables observables y no observables que son constantes en el tiempo se eliminan. Las tendencias en el contrafactual en los grupos de tratamiento y control son iguales.	(CLOTFELTER, LADD y VIGDOR, 2007a, 2007b y GARCÍA, 2013)

NOTA: Algunas de las referencias proporcionan ejemplos prácticos de la utilización de ciertos métodos y referencias que aportan explicaciones metodológicas de los diferentes métodos.

FUENTE: Reproducción parcial, adaptada y ampliada, de los materiales elaborados por el Abdul Latif Jameel Poverty Action Lab, para evaluar programas sociales, JPAL101x.

en política pública. Siguiendo el marco establecido por Levin (2002) para evaluar los efectos educativos de programas de vales escolares, las cuatro posibles dimensiones del impacto del programa que deben tenerse en cuenta son: la eficiencia productiva, la equidad, la libertad de elección y la cohesión social⁴. La evaluación de todos estos resultados o impactos (que pueden ser intencionados y en ocasiones también no intencionados) es importante para comprender aspectos tales como quiénes se han beneficiado y quiénes no se han beneficiado del programa, y qué otras variables, además del propio impacto, pudieron ser afectadas por la intervención. Además de estas, siguiendo las recomendaciones de Rossi, Freeman y Lipsey (2004), el conjunto completo de las cuestiones que deben ser examinadas incluye: la estimación del impacto del programa, un examen de las causas y los mecanismos por los cuales dicho programa funcionó o no funcionó, tal y como se explicará en el apartado 3, su eficiencia, escalabilidad y posible aplicación en un contexto diferente, en comparación con otros programas diseñados para cumplir objetivos similares.

Por último, es necesario hacer referencia a dos importantes advertencias relacionadas con estos métodos estadísticos avanzados y su uso en la información de cuestiones de política pública⁵. En primer lugar, la utilización de estos métodos está necesariamente asociada con importantes demandas y requisitos en relación con los datos y las estadísticas. Van Klaveren y De Wolf (2015) explican estas condiciones como cuestiones relacionadas con los análisis y los datos, y las denominan

condiciones de segundo orden⁶. Aspectos tales como la atrición o desaparición de la muestra, la falta de valores o *missing*, los cruces, los errores estándar no ajustados por pertenencia a un grupo, los efectos de aprendizaje, la falta de equivalencia de los grupos de tratamiento y control, los problemas de la validez del instrumento utilizado para identificar el efecto de la intervención y los problemas relacionados con la potencia del análisis, así como cualquier otra fuente de datos contradictorios o erróneos, serían obstáculos signitativos en la evaluación de programas.

La segunda consideración tiene que ver con la validez de las estimaciones —interna y/o externa⁷—, que puede ser puesta en tela de juicio dada la rigurosidad de los supuestos de los métodos utilizados para la identificación. Por ejemplo, aunque las evaluaciones experimentales constituyen la referencia absoluta en la evaluación de programas, en ocasiones hay limitaciones logísticas y éticas, o incluso limitaciones de tiempo cuando los resultados del experimento aleatorio deben recogerse en un horizonte temporal lejano, así como razones políticas y financieras que pueden hacer inviable la asignación aleatoria de un programa, o constituir formas intrusivas de evaluación (Cook, 2001 y Rubin, 1974). En relación con los requisitos de identificación, es importante destacar que, si bien la validez interna de experimentos aleatorios es indiscutible en circunstancias normales, en algunas ocasiones existe una clara preocupación sobre su validez externa, que se corresponde con la situación en la cual las estimaciones se corresponden con resultados de equilibrio general o de equilibrio parcial (Angrist y Pischke, 2009 y Murnane y Willett, 2011). Por el contrario, en

⁴ El marco expuesto por LEVIN (2002) ha sido utilizado recientemente por GARCÍA y CORNELISZ (2014) para clasificar las asociaciones público-privadas en el ámbito de la educación y la evaluación de sus resultados educativos. Su trabajo comenta los equilibrios y sacrificios existentes entre los distintos resultados, y también propone ampliar el marco de evaluación para poder considerar el cálculo de la relación coste-efectividad en los distintos tipos de colaboraciones público-privadas (GARCÍA y CORNELISZ, 2014 y LEVIN, 2002).

⁵ Las siguientes advertencias se aplican tanto a estudios cuasi-experimentales como a aquellos que utilizan datos experimentales.

⁶ En contraste con cuestiones relativas a la metodología, que serían condiciones de primer orden. Esta clasificación se basa en IMBENS (2009).

⁷ La validez interna se refiere a la seguridad de que la diferencia en el valor del resultado entre los grupos de tratamiento y control es en verdad una medida creíble del impacto causal del tratamiento. La validez externa se refiere a cuán generalizables son los resultados de un programa, implementado en un contexto, a toda la población o en otro contexto. Véase MURNANE y WILLETT (2011).

estudios cuasi-experimentales, la validez interna solo es posible si los (estrictos) supuestos de identificación pueden ser verificados. Como es de esperar, si ese requisito no se cumple, pueden plantearse serias dudas sobre los resultados de la evaluación de los efectos de un programa.

3. Análisis de costes

Como se ha comentado anteriormente, los diferentes métodos de evaluación relacionados con los costes de un programa se entienden cada vez más como componentes intrínsecos, en vez de complementarios, de la evaluación de programas. A pesar de que es fundamental conocer el impacto de una intervención, desde el punto de vista de las políticas es igualmente imprescindible disponer de una estimación adecuada de los costes, dado que esta información va a ser necesaria a la hora de determinar si un programa es factible y cuántos beneficiarios van a poder servirse del mismo. A continuación se explican los objetivos de los estudios coste-efectividad y coste-beneficio, así como los marcos analíticos y las herramientas necesarias para su utilización. Por último, se argumenta cómo los resultados de estos análisis pueden ser de gran importancia en el asesoramiento de las políticas públicas.

Los beneficios de la educación, o la idea de rendimiento de las inversiones en educación, son conocidos desde hace tiempo en el campo de la economía. En la década de los sesenta, Gary Becker introdujo los conceptos de los beneficios de la educación en su libro «*Human Capital*» (Becker, 1964). En cuanto al origen de los estudios de coste-efectividad, las primeras aplicaciones de este análisis en el campo de la educación se desarrollaron a principios de los años setenta, con el trabajo pionero de Henry M. Levin. Levin recibió el encargo de analizar los costes económicos para la sociedad del abandono escolar (Levin, 1972), a instancias del Congreso de Estados Unidos. Después de este primer análisis, su trabajo se centró en el desarrollo de métodos más consistentes basados en criterios

económicos ya aceptados en otros campos (Levin, 1975, 1983), cuya utilización está en clara expansión en la investigación actual (algunas muestras se mencionan a continuación).

Los estudios de coste-efectividad evalúan la productividad relativa de distintas intervenciones que persiguen el mismo objetivo, y por tanto determinan la eficiencia en el uso de los recursos disponibles por parte de dichas intervenciones. Por su parte, los estudios coste-beneficio estiman el valor monetario de los beneficios y los costes durante la vida del proyecto de las distintas intervenciones. Los resultados de estos últimos análisis proporcionan la información necesaria para evaluar el rendimiento de las inversiones en programas, así como el valor económico de una posible ampliación de la escala de las intervenciones. Como los nombres de los métodos indican, ambas técnicas comparten un elemento común, que es la estimación de los costes de la intervención. Por esta razón, antes de explicar los detalles metodológicos de los métodos coste-efectividad y coste-beneficio, se describen en primer lugar las características de los análisis de costes.

Es importante empezar esta descripción explicando qué análisis de costes de políticas públicas utilizan la definición económica de costes⁸. Este enfoque implica que debe estimarse el valor de todos los recursos (o ingredientes) utilizados por un programa que contribuyen a su efecto, tomando como referencia su valoración si se hubieran asignado a sus más valiosos usos alternativos (Belfield y Levin, 2010 y Levin y McEwan, 2001). Este concepto es conocido en economía como coste de oportunidad de los recursos, y se ejemplifica típicamente con el caso de los costes de los voluntarios en un programa: aunque el coste de los voluntarios no sería parte de un presupuesto destinado a la ejecución del programa, desde el punto de vista económico, el

⁸ Estimaciones alternativas de costes pueden provenir de análisis contables o presupuestarios. Para conocer las limitaciones de esos enfoques, véase LEVIN y MCEWAN (2002).

tiempo dedicado por los voluntarios es esencial para el programa y su efecto, y, por tanto, es un componente del coste total del programa. La metodología para calcular los costes totales de una intervención se conoce como «método de los ingredientes» (Levin, 1975, 1983 y Levin y McEwan, 2002).

El método de los ingredientes combina el análisis cuantitativo y cualitativo para la estimación de los costes totales de un programa, y su aplicación se organiza alrededor de cinco etapas distintas. En primer lugar, los recursos que se utilizan para obtener el efecto o impacto del programa son identificados⁹. Esta información suele estar disponible en diversos documentos del programa, tales como los informes de ejecución, y también puede ser obtenida a partir de observaciones directas de la implementación del programa, o a través de la realización de entrevistas sistematizadas con personas que participan en el programa¹⁰. En segundo lugar, se identifican los precios de los ingredientes, principalmente los precios de los ingredientes se miden utilizando precios de mercado (y, siempre que sea posible, valores nacionales). Cuando estos no están disponibles, los precios presupuestados y los precios sombra (que indican lo que la sociedad está dispuesta a pagar por un recurso específico) pueden ser utilizados, con la idea de que se aproximen o indirectamente capturen lo que un ingrediente se valoraría en un mercado competitivo. El tercer paso implica atribuir los gastos a la agencia que está soportando los distintos costes. Esto incluye identificar los costes soportados por los particulares, el Estado o contribuyente, o por la sociedad. Esta distribución de los ingredientes es una característica altamente

valorada para la formulación de políticas públicas, dado que permite identificar la parte del coste que el sector público está destinado a soportar¹¹. Las últimas dos etapas consisten en calcular los costes totales y el coste medio por participante, así como determinar cómo se financian los costes totales (realizando cualquier ajuste que sea necesario en este aspecto).

El análisis coste-efectividad combina los resultados del análisis de costes con las estimaciones de impacto del programa obtenidas por los procedimientos explicados en el apartado 2¹². El resultado identifica, entre una serie de programas alternativos o políticas con objetivos similares, la intervención con menor coste en relación con su efecto (o, alternativamente, cuál logra la máxima efectividad dado un nivel de costes). Un requisito imprescindible en los análisis coste-efectividad es que las múltiples intervenciones deben tener un objetivo común, y sus efectos deben expresarse en unidades comparables¹³. Programas con costes más bajos con respecto a los efectos son elegidos, dado que son los que ofrecen maneras más eficientes en la utilización de los recursos existentes con el fin de obtener un impacto.

Ejemplos recientes de análisis coste-efectividad en educación incluyen a Hollands *et al.* (2013), con una aplicación de este método a cinco programas que persiguen mejorar la alfabetización de niños durante sus primeros años escolares; Levin, García, y Morgan (2012), que utilizan el análisis coste-efectividad para evaluar una intervención destinada a mejorar la progresión y opciones de graduación en educación superior, denominada ASAP (Accelerated Study

⁹ Normalmente los ingredientes se clasifican en cuatro categorías principales, que incluyen personal, instalaciones, equipos y materiales, y otros insumos del programa.

¹⁰ El método de los ingredientes es facilitado cuando el análisis de los costes se desarrolla al mismo tiempo que la evaluación de impacto, y resulta más complejo cuando el análisis de los costes se realiza retrospectivamente. Para ver un ejemplo de protocolos para entrevistas directas, consulte HOLLANDS *et al.* (2013). Para ver un ejemplo de análisis de los costes retrospectivo (con la finalidad de llevar a cabo un análisis de costo-efectividad de intervenciones destinadas a prevenir la deserción escolar en la escuela secundaria), consulte HOLLANDS *et al.* (2014).

¹¹ Es importante señalar que las estimaciones de costes están sujetas a una serie de ajustes, los cuales incluyen ajustes por inflación, cálculo del valor presente y la incertidumbre. Para obtener más información, consúltese LEVIN y MCEWAN (2002).

¹² Para el cálculo de la ratio de efectividad de costes, se utilizan los métodos descritos en el apartado anterior, la estimación de efecto o impacto. El coste por participante es utilizado en el denominador de la relación.

¹³ Un método también relacionado con el análisis de costes que podría ser utilizado cuando los programas a comparar tienen distintos objetivos, y cuando hay resultados no-monetarios y monetarios es el análisis coste-utilidad. Véase LEVIN y MCEWAN (2002) para más detalles.

in Associate Programs) por sus siglas en inglés e implementada en la Universidad de la Ciudad de Nueva York y Dhaliwal, Duflo, Glennerster y Tulloch (2013), quienes ofrecen una aplicación del método coste-efectividad a programas que buscan mejorar la eficiencia de los programas de ayuda internacional que apoyan el desarrollo escolar y personal a niños en países en desarrollo.

Los análisis de coste-efectividad comparten las limitaciones y los desafíos que hemos descrito anteriormente en cuanto a la estimación de impacto o efecto, relacionados principalmente con su precisión y su validez. Además, los resultados del análisis de coste-efectividad son sensibles a posibles sesgos en las estimaciones de los costes. Finalmente, el análisis de coste-efectividad plantea una limitación de tipo conceptual, pues no puede informar sobre la magnitud del proyecto, ni puede justificar si una acción política debiera implementarse en modo alguno. Es importante enfatizar que este método solo nos permite ordenar las políticas según su efecto por unidad de coste¹⁴.

El análisis coste-beneficio compara los beneficios monetarios de uno o varios programas con sus costes durante la vida útil (o un período de tiempo determinado) del proyecto. Una de las ventajas más significativas del análisis coste-beneficio es que su resultado proporciona la información necesaria para evaluar el rendimiento de las inversiones, e informa también del valor económico y la viabilidad de expansiones de las intervenciones. Para desarrollar un análisis coste-beneficio, el primer objetivo es establecer un vínculo causal entre el programa y los resultados posteriores (Nores, Belfield, Barnett y Schweinhart, 2005, 246). Posteriormente, los resultados del programa o las áreas de beneficios identificadas se monetizan utilizando valores de mercado o precios sombra (Boardman, Greenberg, Vining y Weimer, 2011 y Levin y McEwan, 2001). Posteriormente, se calcula el

valor presente (descontado) del flujo de beneficios y costes. La ratio de esos valores proporciona la relación coste-beneficio, que se utiliza para seleccionar los proyectos con una mayor ratio beneficio-costes, o con mayor beneficio neto.

Las evaluaciones coste-beneficio de programas de educación más conocidas se encuentran en la evaluación de intervenciones de educación para la primera infancia, algunos de cuyos efectos se han estimado durante largos períodos de tiempo (en los ejemplos que se mencionan a continuación, desde la década de los sesenta y setenta). Programas en esta categoría son el programa Perry Preschool (Barnett, 1985; Belfield, Nores, Barnett y Schweinhart, 2006; Heckman, Moon, Pinto, Savelyev y Yavitz, 2010; Heckman, Moon, Pinto, Savelyev y Yavitz, 2010 y Nores *et al.*, 2005), el Abecedarian Project (Barnett y Masse, 2007) o el programa Chicago Child Parent Centers (Reynolds *et al.*, 2002). En el ámbito de la enseñanza superior, la evaluación coste-efectividad del programa ASAP, descrita anteriormente fue complementada con un análisis coste-beneficio que informaba de los beneficios netos para el contribuyente, tanto de la implementación de ASAP como de una posible ampliación de la intervención (Levin y García, 2013). Estas distintas evaluaciones ejemplifican y subrayan los aspectos metodológicos implicados en la estimación de los costes y los beneficios que, se miden a lo largo del tiempo, y los que se estiman a partir de los datos actuales (observacionales)¹⁵.

Aunque los requisitos en materia de datos de los análisis de coste-beneficio son superiores a las necesidades de realizar análisis de coste-efectividad, este método ofrece algunas ventajas importantes en cuanto a formulación de políticas. A diferencia del análisis de coste-efectividad, un análisis coste-beneficio de un único programa es informativo y útil. Por otra parte, si se utiliza para comparar la rentabilidad de varios

¹⁴ Para obtener más información y directrices sobre cómo llevar a cabo el análisis de coste-efectividad, consúltense LEVIN y BELFIELD (2015).

¹⁵ Un libro de texto sobre análisis de coste-beneficio es BOARDMAN *et al.*, (2011). Los diferentes estudios mencionados también resaltan la importancia de realizar análisis de sensibilidad y ofrecer algunas pautas sobre cómo llevarlo a cabo.

proyectos, o para seleccionar cuál es la inversión más rentable, proyectos en distintas áreas de política pública pueden ser comparadas entre ellas —educación, salud, medio ambiente, transporte, etcétera—.

4. Conclusiones

Más allá de las múltiples cualidades y bondades de la evaluación global de un programa —su impacto, eficiencia y rentabilidad—, hay aspectos importantes que el marco y los métodos explicados en este documento no contribuirían a informar. En numerosas áreas de conocimiento, la interpretación de los resultados de la evaluación de programas podría aún llevar a ciertos (al menos aparentes) desacuerdos. Estos pueden ocurrir entre los investigadores, los responsables de las políticas y también entre individuos particulares, por razones técnicas, ideológicas, entre otros motivos¹⁶. A pesar de que la evaluación de programas no sea capaz de conciliar interpretaciones contrarias, incluso en esas situaciones, su aplicación puede contribuir a crear un terreno común no solo para saber si las políticas tienen o no un impacto, si no, quizás más importante, para aumentar nuestra comprensión de cómo y por qué funcionan las distintas intervenciones.

El hecho de que administraciones e instituciones¹⁷ estén en una marcada senda de búsqueda de iniciativas y políticas basadas en evidencia, demuestra que estamos abocados a adentrarnos en una etapa de gestión y

decisión de las políticas públicas, en la que la aplicación de procedimientos sistematizados para su evaluación global constituirá la norma y no la excepción. Además de que este proceder está en total consonancia con las tendencias actuales en múltiples regiones y disciplinas, incluida la ayuda microeconómica al desarrollo; también lo está con la esperable evolución de las ciencias sociales, tanto en aspectos metodológicos como logísticos. Si bien todos estos aspectos son importantes, es clave entender que la evaluación de las políticas públicas responde a la más que razonable necesidad de asegurar que funcionen, y que persiguen trabajar de un modo más efectivo y eficiente a medida que nuestro conocimiento también avanza (indudablemente, el último aspecto sería especialmente relevante en tiempos de recortes presupuestarios, pero la búsqueda de la mayor efectividad y eficiencia deberían ser el *motu* de los gestores, practicantes e investigadores en todo momento).

Antes de concluir, es importante destacar que el marco expuesto en este artículo y las evidencias aportadas por los estudios empíricos, mencionados a lo largo del mismo apuntan además a otras relevantes razones por las cuales debería ampliarse la evaluación en el campo de las ciencias sociales y las políticas públicas. Cuando la política pública utiliza un enfoque basado en la evidencia a la hora de la selección de los distintos proyectos que persiguen objetivos similares, los gestores políticos están a la vez fomentando una mayor transparencia en la rendición de cuentas de sus actividades. De hecho, una mayor conciencia de las consecuencias de la adopción de decisiones en las políticas públicas, tanto en el plano económico como en términos de las consecuencias de sus impactos, —asesoradas mediante la aplicación del marco de evaluación propuesto—, contribuiría a aumentar la transparencia de los procesos y a ayudar a mantener a los distintos profesionales involucrados en estas tareas, responsables de su desempeño y actuación profesional. Además, cuando las políticas públicas demandan evidencia de los programas que han obtenido un éxito rigurosamente probado, es decir, determinado por una evaluación diseñada científicamente, la

¹⁶ Un ejemplo de lo anterior, que es particularmente relevante para los actuales volúmenes de la revista ICE, es la discusión mantenida por JEFFREY SACKS y WILLIAM EASTERLY sobre si la ayuda al desarrollo funciona o no funciona (SACKS, 2006 y EASTERLY, 2001). Además, para un análisis de por qué no siempre resultados basados en evidencias son el único insumo utilizado en el diseño de la política pública, véase FEINSTEIN (2007). Sobre si todos los programas deben ser evaluados, véase DAHLER-LARSEN, 2007.

¹⁷ Estos incluyen la Unión Europea, EE UU, y la mayoría de los organismos multilaterales. Algunas instituciones han venido desarrollando documentos de orientación y uso con el fin de crear estrategias y metodologías sistemáticas para realizar sus evaluaciones. Algunos ejemplos son FLORIO *et al.*, (2008); GERTLER, MARTINEZ, PREMAND, RAWLINGS y VERMEERSCH (2011); HASKINS y MARGOLIS (2014) y LOI y RODRIGUES (2012).

cooperación entre investigación y políticas se intensifica y se vuelve, a la larga, más fructífera.

La descripción de los componentes de la evaluación de programas proporcionada por este documento tenía como objetivo el presentar una introducción a los métodos que se utilizan, las preguntas que tienen por objeto responderse, así como mencionar algunos de los problemas a los que los evaluadores se enfrentan a nivel práctico. A pesar de dichos obstáculos, la investigación y la evaluación son cruciales para entender si las políticas y los programas en funcionamiento están, en efecto, logrando sus objetivos, así como para conformar los esfuerzos destinados a atender las necesidades de los beneficiarios de las intervenciones y para mejorar las estructuras de rendición de cuentas. En definitiva, una mayor cooperación y entendimiento entre la formulación, la implementación y la investigación en la evaluación de programas, pueden contribuir a maximizar el alcance, los resultados y el propio funcionamiento de las políticas públicas.

Referencias bibliográficas

- [1] ANGRIST, J.D. (2003). *Randomized Trials and Quasi-experiments in Education Research*. NBER Reporter Online. Verano 2003, pp. 11-14.
- [2] ANGRIST, J. D y KRUEGER, A. B. (1991). «Does Compulsary School Attendance Affect Schooling and Earnings?». *The Quarterly Journal of Economics*, vol. 106, n° 4, pp. 973-1014.
- [3] ANGRIST, J. D. y PISCHKE, J. S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: University Princeton.
- [4] ANGRIST, J. D; GUIDO, W. y RUBIN, D. B. (1996). «Identification of Causal Effects Using Instrumental Variables». *Journal of the American statistical Association*, vol. 91, n° 434, pp. 444-455.
- [5] BANERJEE, A. V.; COLE, S.; DUFLO, E. y LINDEN, L. (2007). «Remedying Education: Evidence from Two Randomized Experiments in India». *The Quarterly Journal of Economics*, vol. 122, n° 3, pp. 1235-1264.
- [6] BARNETT, W. S. (1985). «Benefit-cost analysis of the Perry Preschool Program and its Policy Implications». *Educational Evaluation and Policy Analysis*, vol. 7, n° 4, pp. 333-342.
- [7] BARNETT, W. S. y MASSE, L. N. (2007). «Comparative Benefit-cost Analysis of the Abecedarian Program and its Policy Implications». *Economics of Education Review*, vol. 26, n° 1, pp. 113-125.
- [8] BECKER, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. (3rd edition ed.) New York: National Bureau of Economic Research.
- [9] BELFIELD, C. R. y LEVIN, H. M. (2010). «Cost-benefit Analysis and Cost-effectiveness Analysis. In D. B. P.». *Encyclopedia of Economics of Education*, MCEWAN (Ed.), pp. 199-203.
- [10] BELFIELD, C. R.; NORES, M.; BARNETT, W. S. y SCHWEINHART, L. (2006). «The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup». *The Journal of Human Resources*, vol. 41, n° 1, pp. 162-190.
- [11] BERGER, L. M.; HILL, J. y WALDFOGEL, J. (2005). «Maternity Leave, Early Maternal Employment and Child Health and Development in the United States». *The Economic Journal*, vol. 115, n° 501, pp. F29-F47.
- [12] BERGER, L. M.; PAXSON, C. y WALDFOGEL, J. (2009). «Income and Child Development». *Children and Youth Services Review*, vol. 31, n° 9, pp. 978-989.
- [13] BLAU, D. M. (1999). «The Effect of Income on Child Development». *Review of Economics and Statistics*, vol. 81, n° 2, pp. 261-276.
- [14] BLOOM, H. S. (2009). «Modern Regression Discontinuity Analysis». *MDRC Working Papers on Research Methodology*: Newbury Park, CA. Sage.
- [15] BOARDMAN, A. E.; GREENBERG, D. H.; VINING, A.R. y WEIMER, D. L. (2011). *Cost-benefit analysis: concepts and practice*. 4th Ed.: Prentice Hall.
- [16] CARD, D. y KRUEGER, A. B. (2000). «Minimum Wages and Employment: a Case Study of the Fast-food Industry in New Jersey and Pennsylvania: Reply». *American Economic Review*, vol. 90, n° 5, pp. 1397-1420.
- [17] CARNEIRO, P.; HECKMAN, J. J. y VYTLACIL, E. (2003). *Understanding what Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education*. University of Chicago: The American Bar Foundation and Stanford University, July, 19.
- [18] CLOTFELTER, C. T.; LADD, H. F. y VIGDOR, J. L. (2007a). *How and Why do Teacher Credentials Matter for Student Achievement?*. National Bureau of Economic Research.
- [19] CLOTFELTER, C. T.; LADD, H. F. y VIGDOR, J. L. (2007b). «Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects». *Economics of Education Review*, vol. 26, n° 6, pp. 673-682.
- [20] COOK, T. D. (2001). «Sciencephobia». *Education Next*, vol. 1, n° 3, pp. 62-68.
- [21] DAHLER-LARSEN, P. (2007). «¿Debemos evaluarlo todo?: O de la estimación de la evaluabilidad a la cultura de

la evaluación». *Información Comercial Española. Revista de economía*, nº 836, pp. 93-106.

[22] DAVIES, P. (1999). «What is Evidence-based Education?». *British Journal of Educational Studies*, vol. 47, nº 2, pp. 108-121.

[23] DEHEJIA, R. H. y WAHBA, S. (1999). «Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs». *Journal of the American statistical Association*, vol. 94, nº 448, pp. 1053-1062.

[24] DHALIWAL, I.; DUFLO, E.; GLENNERSTER, R. y TULLOCH, C. (2013). «Comparative Cost-effectiveness Analysis to Inform Policy in Developing Countries: a General Framework with Applications for Education». *Education Policy in Developing Countries*, en P. GLEWWE (Ed.): University of Chicago Press, pp. 285-338.

[25] DUFLO, E.; GLENNERSTER, R. y KREMER, M. (2007). «Using Randomization in Development Economics Research: A Toolkit». *Handbook of Development Economics*, nº 4, pp. 3895-3962.

[26] EASTERLY, W. y EASTERLY, W. R. (2001). *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*: MIT Press.

[27] INFORMACIÓN COMERCIAL ESPAÑOLA (2007). *Evaluación de Políticas Públicas. Información Comercial Española. Revista de Economía*, nº 836. Monográfico de mayo-junio 2007.

[28] EVANS, D. y GHOSH, A. (2008). *Prioritizing Educational Investments in Children in the Developing World*.

[29] FEINSTEIN, O. N. (2007). «Evaluación pragmática de políticas públicas». *Información Comercial Española. Revista de economía*, nº 836, pp. 19-32.

[30] FLORIO, M.; FINZI, U.; GENCO, M.; LEVARLET, F.; MAFFII, S.; TRACOGNA, A. y VIGNETTI, S. (2008). «Guide to Cost-benefit Analysis of Investment Projects». *Evaluation Unit, DG Regional Policy, European Commission*.

[31] GARCÍA, E. (2013). *What we Learn in School: Cognitive and Non-cognitive Skills in the Educational Production Function*. Columbia University.

[32] GARCÍA, E. y CORNELISZ, I. (2014). «Public-Private Partnerships in Education. In D. J.». *Encyclopedia of Education Economics and Finance*. Brewer, en LAWRENCE O. PICUS (Ed.), SAGE Publications.

[33] GERTLER, P. J.; MARTINEZ, S.; PREMAND, P.; RAWLINGS, L. B. y VERMEERSCH, C. M. J. (2011). *Impact evaluation in practice*. World Bank Publications.

[34] HASKINS, R. y MARGOLIS, G. (2014). *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*. Brookings Institution Press.

[35] HECKMAN, J. J.; MOON, S. H.; PINTO, R.; SAVELYEV, P. y YAVITZ, A. (2010). «A New Cost-benefit and Rate of Return

Analysis for the Perry Preschool Program» *National Bureau of Economic Research*, A summary.

[36] HECKMAN, J. J.; MOON, S. H.; PINTO, R.; SAVELYEV, P. A. y YAVITZ, A. (2010). «The Rate of Return to the HighScope Perry Preschool Program». *Journal of Public Economics*, vol. 94, nº 1-2, pp. 114-128.

[37] HILL, J. (2008). «Discussion of Research Using Propensity-score Matching: Comments on 'A Critical Appraisal of Propensity-score Matching in the Medical Literature Between 1996 and 2003' by Peter Austin, Statistics in Medicine». *Statistics in Medicine*, vol. 27, nº 12, pp. 2055-2061.

[38] HOLLAND, P. W. (1986). «Statistics and Causal Inference». *Journal of the American Statistical Association*, vol. 81, nº 396, pp. 945-960.

[39] HOLLANDS, F. M.; BOWDEN, A. B.; BELFIELD, C. R.; LEVIN, H. M.; CHENG, H.; SHAND, R. y HANISCH-CERDA, B. (2014). «Cost-Effectiveness Analysis in Practice Interventions to Improve High School Completion». *Educational Evaluation and Policy Analysis*, vol. 36, nº 3, pp. 307-326.

[40] HOLLANDS, F. M.; PAN, Y.; SHAND, R.; CHENG, H.; LEVIN, H. M.; BELFIELD, C. R. y HANISCH-CERDA, B. (2013). *Improving Early Literacy: Cost-effectiveness Analysis of Effective Reading Programs: Center for Benefit Cost Studies of Education*. Teachers College: Columbia University.

[41] IMBENS, G. W. (2010). «Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)». *Journal of Economic Literature*, nº 48, pp. 399-423.

[42] IMBENS, G. W. y LEMIEUX, T. (2008). «Regression Discontinuity Designs: A Guide to Practice». *Journal of Econometrics*, vol. 142, nº 2, pp. 615-635.

[43] JACOB, B. A. y LEFGREN, L. (2004). «Remedial Education and Student Achievement: A Regression-discontinuity Analysis». *Review of economics and statistics*, vol. 86, nº 1, pp. 226-244.

[44] LEVIN, H. M. (1972). *The Costs to the Nation of Inadequate Education* (Report to the Select Committee on Equal Educational Opportunity, US Senate, 92d Cong.). Washington, DC: US Government Printing Office.

[45] LEVIN, H. M. (1975). «Cost-effectiveness Analysis in Evaluation Research». *Handbook of Evaluation Research*, nº 2, pp. 89-122.

[46] LEVIN, H. M. (1983). *Cost-effectiveness: A Primer*. Sage Publications Beverly Hills, CA.

[47] LEVIN, H. M. (2002). «A Comprehensive Framework for Evaluating Educational Vouchers». *Educational Evaluation and Policy Analysis*, vol. 24, nº 3, p. 159.

[48] LEVIN, H. M. y BELFIELD, C. R. (2015). «Guiding the Development and Use of Cost-effectiveness Analysis in Education». *Journal of Research on Educational Effectiveness*, nº 0, pp. 1-19.

- [49] LEVIN, H. M. y GARCÍA, E. (2013). *Benefit-cost Analysis of Accelerated Study in Associate Programs (ASAP) of the City University of New York (CUNY)*. New York: Center for Benefit-Cost Studies in Education, Teachers College, Columbia University, 2013.
- [50] LEVIN, H. M. y MCEWAN, P. J. (2001). *Cost-effectiveness Analysis: Methods and Applications*. SAGE Publications, Incorporated.
- [51] LEVIN, H. M.; GARCÍA, E. y MORGAN, J. (2012). *Cost-Effectiveness of Accelerated Study in Associate Programs (ASAP) of the City University of New York (CUNY)*. New York: City University of New York (CUNY).
- [52] LOI, M. y RODRIGUES, M. (2012). *A Note on the Impact Evaluation of Public Policies: the Counterfactual Analysis*. Luxembourg: Publications Office of the European Union, European Commission.
- [53] MORGAN, S. L. y WINSHIP, C. (2014). *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press.
- [54] MURNANE, R. J. y WILLETT, J. B. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Cambridge: Oxford University Press.
- [55] NORES, M.; BELFIELD, C. R.; BARNETT, W. S. y SCHWEINHART, L. (2005). «Updating the Economic Impacts of the High/Scope Perry Preschool Program». *Educational Evaluation and Policy Analysis*, vol. 27, nº 3, pp. 245-261.
- [56] REYNOLDS, A.; TEMPLE, J.; ROBERTSON, D. y MANN, E. (2002). «Age 21 Cost-Benefit Analysis of the Title I Chicago Child-Parent Centers». *Educational Evaluation and Policy Analysis*, vol. 24, nº 4, pp. 267-303.
- [57] ROSSI, P. H.; LIPSEY, M. W. y FREEMAN, H. E. (2004). *Evaluation: A Systematic Approach* 7th Edition ed.: Sage Publications.
- [58] RUBIN, D. B. (1974). «Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies». *Journal of Educational Psychology*, vol. 66, nº 5, pp. 688.
- [59] SACHS, J. (2006). *The End of Poverty: Economic Possibilities for Our Time*. Penguin Books: Reprint Edition.
- [60] SCRIVENER, S.; WEISS, M. J.; RATLEDGE, A.; RUDD, T.; SOMMO, C. y FRESQUES, H. (2015). «Doubling Graduation Rates: Three-Year Effects of CUNY's Accelerated Study in Associate Programs (ASAP) for Developmental Education Students». *MDRC, Building Knowledge to Improve Social Policy*.
- [61] SHADISH, W. R.; COOK, T. D. Y CAMPBELL, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Cengage learning: 2nd Edition.
- [62] FONDO DE NACIONES UNIDAS PARA LA INFANCIA (UNICEF) (2008). *Bridging the Gap: The Role of Monitoring and Evaluation in Evidence-based Policy Making*. Geneva: UNICEF Regional Office for Central and Eastern Europe/Commonwealth of Independent States.
- [63] VAN KLAVEREN, C. y DE WOLF, I. (2015). *Systematic Reviews In Education Research: When Do Effect Studies Provide Evidence?*. De WITTE, K. Ed. Contemporary Education Issues from an Economic Perspective, Leuven: Leuven University Press, pp. 1-26.