

Evaluación de impacto: más allá de la experimentación

Impact evaluation: beyond experimentation

*Ester García Sánchez**

*Myriam Cardozo Brum***

Resumen

Tanto en España como en América Latina, la evaluación de impacto de programas públicos ha iniciado en la primera década de este siglo con un notorio énfasis en la aplicación de métodos cuantitativos y diseños experimentales contrafactuales. En el trabajo se presentan los casos de Colombia, Chile, México y España. Con la intención de ampliar las opciones metodológicas consideradas en estos países, se revisa el concepto de impacto, los enfoques teóricos en que se sustenta este tipo de evaluación y algunas experiencias en las que se han empleado con éxito métodos alternativos.

Palabras clave: evaluación de impacto, contrafactual, experimentación, América Latina, España.

Abstract

In Latin American countries and in Spain, impact evaluation has begun in the first decade of this century with a clear emphasis on quantitative methods and counterfactual quasi-experimental design. The cases of Chile, Colombia, Mexico and Spain are discussed in this paper. In order to expand the methodological options considered in these countries, we review the existing definitions of impact, the theoretical approaches on which impact evaluation is based and several successful experiences in applying alternative methods.

Key words: impact evaluation, counterfactual, quasi-experimental design, Latin America, Spain.

Artículo recibido el 16-06-2016

Apertura del proceso de dictaminación: 10-10-16

Artículo aceptado el 24-03-17

* Universidad Carlos III de Madrid, España [estgarsa@polsoc.uc3m.es].

** Profesora-investigadora, Departamento de Política y Cultura, UAM-Xochimilco, México [mcardozo@correo.xoc.uam.mx].

La evaluación de impacto de políticas y programas públicos es uno de los temas de mayor interés en los últimos años a un lado y otro del Atlántico debido a la importancia de la información que genera para la toma de decisiones, al mismo tiempo que por la complejidad metodológica que implica su realización.

El predominio del positivismo ha provocado que, al menos en Iberoamérica,¹ la evaluación de impacto se vincule con el empleo de diseños experimentales (o cuasiexperimentales) e hipótesis contrafactuales, como si se tratara de las dos caras de una misma moneda. Sin embargo, un examen pormenorizado de la bibliografía² generada en otros países pone de relieve que el panorama de la evaluación de impacto es más rico y diverso de lo que, en un principio, pudiera parecer a tenor del discurso dominante, asumido o impulsado por muchos organismos, instituciones y foros científicos y académicos.

Ya en una primera aproximación a los trabajos teóricos y empíricos realizados en este campo, se percibe que varios omiten la conceptualización de impacto (también de su evaluación) o parten de una definición que se asume como indiscutible, como si existiera al respecto un consenso universal. Clarificar el concepto es imprescindible para construir el trabajo científico sobre bases sólidas³ y permitir el debate entre diferentes enfoques teórico-metodológicos. En este caso, la dificultad se agrava por los problemas de traducción.

En consecuencia, el presente artículo tiene como objetivo presentar la situación de la evaluación de impacto reinante en las instituciones especializadas de España y de tres de los países latinoamericanos con mayores avances en la materia (Colombia, Chile y México), que se caracteriza por la presencia casi exclusiva de enfoques y métodos experimentales, para luego ofrecer una revisión documental amplia del marco teórico-conceptual-metodológico de la evaluación de impacto, recuperando especialmente los diversos enfoques y métodos no experimentales disponibles, con sus ventajas y desventajas, y

¹ Sería imposible hacer una revisión de experiencias mundiales; elegimos trabajar en nuestro contexto más inmediato, donde el predominio planteado está presente.

² Por razones de espacio no se indicarán las direcciones URL públicamente conocidas.

³ Giovanni Sartori, “¿Hacia dónde va la ciencia política?”, *Revista Española de Ciencia Política*, núm. 12, 2005, pp. 9-13.

señalando algunas experiencias en las que éstos han sido empleados con éxito.

Se trata fundamentalmente de analizar y compartir un abanico de posibilidades, cuya difusión es particularmente limitada en los tres países latinoamericanos seleccionados, para enriquecer sus opciones teórico-metodológicas. Por razones de espacio no es posible exponerlas detalladamente, pero se ofrecen referencias para quienes se interesen en profundizar en cada una de ellas.

LA EVALUACIÓN DE IMPACTO EN IBEROAMÉRICA

Si bien existen evaluaciones de impacto de programas⁴ en diversos países de Iberoamérica (Argentina, Brasil, Ecuador, Costa Rica, Uruguay, Nicaragua, Bolivia, Perú, entre otros), como ya se mencionó, esta revisión sólo incluye el caso de España y de los tres países latinoamericanos que presentan el mayor número de evaluaciones de este tipo en los últimos años: Colombia, Chile y México. Para efectuarla, se procedió a una intensa búsqueda de información en libros, revistas, internet, así como consultas a especialistas de los países incluidos.

Como antecedente de este trabajo, en 2005 la Comisión Económica para América Latina (Cepal)⁵ ya había analizado evaluaciones de impacto en la región, encontrando el mismo énfasis experimental de los casos que a continuación se refieren: dos utilizaron un modelo experimental, quince cuasiexperimental, uno combinando los dos anteriores y sólo uno no experimental (en Perú).

España

En los últimos años y desde distintos frentes, en España se insiste en la necesidad de que el diseño de las políticas y los programas públicos se base en evidencias sólidas y de que la evaluación se convierta en uno de los principales instrumentos para lograrlo, poniendo en práctica las recomendaciones de evaluadores y otros expertos, en lugar de dejarse llevar por la inercia o por la

⁴ No se identificaron evaluaciones de impacto de políticas públicas, sólo se incluyeron las de programas públicos. No se consideran tales el análisis del impacto de las crisis económicas y sociales, la sistematización de experiencias y resultados realizada por los responsables de los propios programas o las evaluaciones con criterios distintos al impacto.

⁵ Hugo Navarro, *Manual de evaluación de impacto de proyectos y programas de lucha contra la pobreza*, Chile, Cepal, 2005.

ideología en programas cuya efectividad no está comprobada o que resultan demasiado costosos. Casi siempre, el reclamo ha acompañado la apuesta por los diseños experimentales (y la evaluación de impacto entendida como equivalente).⁶

España cuenta con tres organismos públicos de evaluación, todos de carácter generalista: la Agencia Estatal de Evaluación y Calidad (Aeval), creada en 2006, la Comisión Foral de Evaluación de las Políticas Públicas y la Calidad de los Servicios de la Comunidad Autónoma de Navarra y el Instituto Catalán de Evaluación de Políticas Públicas (Ivàlua), constituidos en 2005 y 2006 respectivamente.⁷ Junto a éstos, existen unidades u organismos de evaluación sectoriales (por ejemplo, centrados en educación, cooperación al desarrollo, investigación, etcétera) tanto a nivel estatal como en algunas comunidades autónomas, lo que genera un elevado –y no siempre positivo– grado de fragmentación.⁸

⁶ Ya sea como postura oficial, discurso o práctica habitual de organizaciones (el caso de Ivàlua), think tanks (por ejemplo, Fedea), institutos de investigación (IVIE, que cuenta con el apoyo de la Fundación BBVA) o blogs (Nada es Gratis o Politikon, entre otros) o como propuesta a título individual de personas con proyección pública y/o reputación académica vinculadas con éstos. Véanse, a modo de ejemplo, David Casado, “¿Por qué no evaluamos las políticas públicas como los fármacos? Una apuesta por la experimentación social”, *Avaluació per al Bon Govern*, núm. 3, 2012; José García Montalvo (ed.), *El análisis experimental de la ayuda al desarrollo. La evaluación de lo que funciona y de lo que no funciona*, Madrid, Fundación BBVA, 2008; Manuel Bagés, “Evaluación experimental de políticas: el (buen) ejemplo de las ONG”, *Nada es gratis*, 31 de julio de 2012 [<http://nadaesgratis.es/bagues/evaluacion-experimental-de-politicas-el-buen-ejemplo-de-las-ongs>], fecha de consulta: 2 de abril de 2017; Kiko Llaneras, “Una política basada en la evidencia”, *Politikon*, 26 de mayo de 2013 [<http://politikon.es/2013/05/26/por-una-politica-basada-en-la-evidencia/>], fecha de consulta: 3 de abril de 2017; Kiko Llaneras y Octavio Medina, “Esther Duflo y el auge de lo aleatorio”, *Politikon*, 14 de mayo de 2015 [<http://politikon.es/2015/05/14/esther-duflo-y-el-auge-de-lo-aleatorio/>], fecha de consulta: 3 de abril de 2017. El artículo de Alejandra Agudo, “Menos cooperación y bajo la lupa”, publicado en *El País* el 21 de diciembre de 2012, ofrece algunas claves sobre cómo se ha concebido la evaluación en los anteriores gobiernos del Partido Popular.

⁷ Si bien el último no comenzó a funcionar hasta 2008.

⁸ Ester García Sánchez, “Sistemas de evaluación en España: análisis y propuestas”, *Studia Politicae*, núm. 34, 2015, pp. 117-133. Hasta la fecha, la descripción más completa de la actividad de evaluación en el marco de la AGE es la que ofrece el *Mapa de la Evaluación de Políticas y Programas Públicos en la Administración General del Estado (2007-2011)*, investigación realizada por Francisco J. Vanaclocha, Ester García Sánchez, Gema Sánchez y Rubén Sánchez en 2012 por encargo de Aeval, en la que se identificaron cuatro evaluaciones centradas sólo en impactos, de un total de 71. Fueron realizadas a instancias del Centro para el Desarrollo Tecnológico Industrial (CDTI), entidad dependiente del Ministerio de Economía y Competitividad.

De los tres organismos, Ivàlua desarrolla una actividad más destacada en materia de evaluación de impacto; de hecho, este tipo de evaluación, en su orientación cuantitativista y experimental, constituye una de sus principales señas de identidad. Las evaluaciones de impacto realizadas (15) suponen casi 33% del total (46) en el periodo 2008-2015.⁹ La relevancia que se otorga a ese tipo y método de evaluación se advierte también en otros detalles. Así, su definición de evaluación de impacto gira en torno a la hipótesis contrafactual, propia de los diseños experimentales. El discurso que impregna los informes, los boletines de noticias y los eventos científicos, de formación y de difusión de Ivàlua –basta una rápida consulta a su página electrónica– subraya la necesidad de fomentar ese tipo de evaluación y de diseños. Es lógico, pues, que seis de las siete evaluaciones de impacto de las que tenemos información hayan seguido una perspectiva experimentalista y que, en el caso que utilizó una alternativa diferente (teoría del programa), el propio informe señale que esta perspectiva no permite evaluar los impactos de forma robusta.

CUADRO 1
Evaluaciones de impacto realizadas por Ivàlua 2008-2015

Sector	Número de evaluaciones	Método
Educación	2	Informe no disponible
	2	Matching
	1	Cuasi-experimental (dobles diferencias y matching)
		Entrevistas semiestructuradas para percepción de impactos
	1	Diseño experimental (en una parte de la evaluación)
Política social	3	Teoría del programa Cuasi-experimental (dobles diferencias) Asimilado a cuasi-experimental
	2	Informe no disponible
Turismo	1	
Economía / trabajo	2	Informe no disponible
NNTT	1	

Fuente: [<http://www.ivalua.cat/generic/static.aspx?ID=782&>], fecha de consulta: 12 de junio de 2016.

⁹ Este cómputo incluye sólo las evaluaciones de programas públicos y cuyos informes se mencionan en las memorias institucionales o se encuentran disponibles en la web. Por ello, en algunos casos no ha sido posible averiguar el método específico de evaluación empleado.

Por el contrario, la experiencia de la Aeval en evaluación de impacto es ciertamente limitada. Ninguna de las 40 evaluaciones realizadas hasta el momento se ha centrado en exclusiva en impactos, si bien en seis de ellas sí se contempló este criterio junto con otros aspectos o fases de las intervenciones públicas. A diferencia de Ivàlua, la perspectiva metodológica que se adopta es abierta, en la medida en que se entiende que los métodos han de depender de la naturaleza del objeto evaluado y de los objetivos que persiga la evaluación, razón por la que en los seis trabajos mencionados han primado los métodos mixtos:

CUADRO 2
Evaluaciones realizadas por Aeval que incluyen impactos 2007-2015

Sector	Número de evaluaciones	Método
Función pública / administraciones públicas	2	Mixtos pero incluyen análisis tentativo cuasi experimental con Propensity Score Matching Mixtos
Trabajo e inmigración	1	Mixto (técnicas estadísticas y econométricas para impactos)
Medio ambiente	1	Mixtos
Economía	1	Mixtos
Política social	1	Mixtos

Fuente: Mapa de la Evaluación de Políticas y Programas Públicos en la AGE (2007-2011), 2012 [<http://www.aeval.es/>], fecha de consulta: 12 de junio de 2016.

En el caso de Navarra, encontramos una situación similar a la de Aeval. En el conjunto de las administraciones públicas de esta Comunidad Autónoma se han realizado, entre 2007 y 2015, un total de 40 evaluaciones de las cuales sólo dos incluyen el criterio de impacto, todo ello a tenor de los datos recogidos en el registro de evaluaciones. En una de esas dos evaluaciones –la única en la que se dispone del informe– se emplean métodos mixtos para conocer las percepciones sobre los impactos, pero se recuerda que con ello sólo se logra obtener una medida indirecta de los mismos.

A excepción de Ivàlua, cuyo énfasis en la experimentación es notorio, las evaluaciones de impacto son aún escasas en los organismos de evaluación en España, no así las publicaciones sobre el tema, que son cada vez más

frecuentes,¹⁰ en especial aquellas que apuestan por los métodos experimentales, y mucho más que en América Latina.

La oferta formativa es también limitada en España. Además de los módulos sobre evaluación de impacto que se incluyen en los másteres y/o cursos de experto de evaluación existentes (ofertados por la Universidad Complutense y la Universidad de Sevilla), se pueden mencionar el curso monográfico organizado por el Centro de Estudios de Cooperación al Desarrollo de la Universidad San Pablo-CEU (Cecod), la FIIAPP, el Banco Interamericano de Desarrollo (BID) y el Fondo Multilateral de Inversiones (Fomin), cuya segunda edición tuvo lugar en 2016, y el curso sobre métodos cuantitativos de evaluación de impacto del ICEX-CECO.

Colombia

El Sistema Nacional de Evaluación de Resultados de la Gestión Pública (Sinergia) fue creado en 1994 y encargado al Departamento Nacional de Planeación (DNP), que depende directamente de la Presidencia. Sus evaluaciones cubren los denominados eslabones de la cadena de valor, al final de la cual se ubican las relativas a impacto. No fue posible encontrar una definición específica de este tipo ni una exposición de los métodos disponibles.

En 2004, el DNP decidió iniciar con la evaluación de los programas que conforman la Red de Apoyo Social: Empleos en Acción, Familias en Acción y Jóvenes en Acción. De manera acumulada se informa que se habrían realizado 34 evaluaciones de impacto hasta 2012 (30% del total); sin embargo, muchas son sólo levantamiento de líneas de base, por lo que nuestra lista se reduce a

¹⁰ La lista es bastante amplia. Valgan, a modo de ejemplo, Javier Mato, *La formación para el empleo: una evaluación cuasi-experimental*, Madrid, Civitas, 2002; los trabajos de José Ma. Larrú, “Las evaluaciones de impacto aleatorias y los microcréditos”, *Revista de Economía Mundial*, núm. 19, 2008, pp. 33-62, y “Evaluaciones de impacto en la cooperación al desarrollo”, *Revista de Evaluación de Programas y Políticas Públicas*, núm. 3, 2014, pp. 117-153, además del referido en la nota a pie núm. 73; Mauro Mediavilla y Jorge Calero, “Metodología para la evaluación de impacto de las políticas públicas educativas a través del diseño cuasi-experimental”, VV.AA., *Lecturas sobre economía de la educación: homenaje a María Jesús San Segundo*, 2011, pp. 189-218; Brindusa Anghel, Antonio Cabrales y Jesús M. Carro, “Evaluating a Bilingual Education Program in Spain: The Impact Beyond Foreign Language Learning”, *Economic Inquiry*, 54(2), 2016, pp. 1202-1223; César Alonso-Borrego, Antonio Romero-Medina, Rocío Sánchez-Mangas, “The Impact of Public Research Contracts on Scientific Productivity”, UC3M Working papers del Departamento de Economía, núm. 5, 2014.

24 (más de dos en promedio anual) y sólo tres en los cuatro años siguientes. De todas éstas 15 fueron realizadas por una misma consultora, con especialización en técnicas microeconómicas y recolección y tratamiento de datos,¹¹ mientras en las restantes nueve participaron organizaciones diferentes. Su desglose por sectores es el siguiente:

CUADRO 3
Evaluaciones de impacto colombianas 2004-2016

Sector	Número de evaluaciones	Método
Presidencia de la República	11	Cuasiexperimentales (uno se combina con métodos cualitativos y otro enfatiza la teoría del cambio)
Salud y protección social	4	Cuasiexperimentales (uno se combina con métodos cualitativos)
Vivienda y desarrollo territorial	3	1 cuasiexperimental y 2 encuestas y entrevistas
Comercio, industria y turismo	2	1 cuasiexperimental (combinado con métodos cualitativos) y 1 encuestas
Agropecuario	1	Cuasiexperimental (combinado con métodos cualitativos)
Trabajo	1	Encuestas, entrevistas y modelo de regresión
Educación	1	Cuasiexperimental (combinado con métodos cualitativos)
Transporte	1	Entrevistas y modelos econométricos

Fuente: reelaborado con datos de [<http://sinergiapp.dnp.gov.co/#Evaluaciones/EvalFin>] y [<https://sinergiapp.dnp.gov.co/PortalDNP/default.aspx>], fecha de consulta: 25 de mayo de 2016.

¹¹ André-Noel Roth, "La evaluación de políticas públicas en Colombia: una mirada crítica a partir de las prácticas evaluativas oficiales de los programas de la Red de Apoyo Social", *Revista del CLAD Reforma y Democracia*, núm. 45, 2009, pp. 161-186.

Como puede verse, salvo en cinco casos, las evaluaciones utilizaron metodología cuasiexperimental, sin participación de los beneficiarios ni de los funcionarios ejecutores. A diferencia de los otros países latinoamericanos revisados, en nuestra búsqueda identificamos algunos libros y artículos de investigadores o consultores que analizan la situación colombiana en la materia.¹² La oferta académica colombiana no incluye posgrados en evaluación, menos aún especializados en impacto.

Chile

El Sistema de Control de Gestión se consolidó en 2006 con la creación del Sistema de Seguimiento de la Programación Gubernamental (SPG), coordinado por el Ministerio de Hacienda a través de la Dirección de Presupuestos (Dipres).

A partir de 2001 se aplican evaluaciones de impacto a los programas que ya han sido evaluados, en los que no se pudieron medir los resultados finales. Las mismas tienen por objeto cubrir ese faltante: los efectos netos de un programa en los beneficiarios. Utilizan enfoques cuasiexperimentales ampliamente expuestos en diversos documentos institucionales¹³ y deben ser aprobadas por la Dipres en materia metodológica y operativa. Requieren trabajo de campo y presupuestos elevados; su duración varía entre 12 y 18 meses y son realizadas por consultores externos, universidades o consultoras privadas, seleccionados en concursos públicos.

Entre 2001 y 2011 se informa la realización de 96 evaluaciones de impacto (cerca de nueve en promedio anual); sin embargo, en los años siguientes sólo se realizó una (2014). Aunque varían en tamaño y complejidad, la metodología utilizada es prácticamente la misma para todas (cuasiexperimental). En algunos casos se han presentado dificultades para aislar los impactos, debido principalmente a insuficiencias de información sobre líneas de base (datos previos al inicio del programa). En el Cuadro 4 se muestra su distribución.

¹² Son ejemplos de éstos el trabajo de Víctor Manuel Quintero, *Características de las demandas de servicios de seguimiento y evaluación – Colombia*, Informe final de la Consultoría BID-RELAC, 2010 o el artículo de Roth recién citado.

¹³ Dipres, *Evaluación ex-post: conceptos y metodologías*, División de Control de Gestión, artículo 22557, 2015.

CUADRO 4
Evaluaciones de impacto chilenas 2001-2011

Sector	Número de evaluaciones	Método
Educación	38	Cuasiexperimental
Planificación	12	
Economía, fomento y turismo	12	
Agricultura	11	
Salud	6	
Trabajo y previsión social	4	Con recolección de información
Vivienda	3	
Interior y seguridad pública	3	
Hacienda	2	
Relaciones exteriores	2	Primaria en campo
Minería	1	
Desarrollo social	1	
Secretaría General de Gobierno	1	

Fuente: reelaborado con datos de [<http://www.dipres.gob.cl/574/w3-propertyvalue-23076.html>], fecha de consulta: 31 de mayo de 2016.

A diferencia del caso colombiano, en Chile las evaluaciones de impacto son más numerosas y muestran mayor diversificación por sectores. Su opción metodológica es claramente cuasiexperimental y su sistema de evaluación ha sido considerado robusto y muy avanzado por el Banco Mundial;¹⁴ sin embargo, este organismo recomendó que se complemente, con base en la *theory-based evaluation*, para poner especial atención en las relaciones causales. Sus publicaciones sobre el tema también son muy pocas,¹⁵ en cambio tiene diversas maestrías y diplomados, dos de ellos especializados en evaluación de impacto general y ambiental (Universidad de Chile y de Lagos, respectivamente).

¹⁴ Banco Mundial, *Chile: estudio de evaluación de impacto del Programa de Evaluación de Programas*, Informe núm. 34589-CL, 2005.

¹⁵ Por ejemplo, la misma Dipres publicó, en 2007, *Anexo. Metodología evaluación de impacto*, que expone el tema a nivel conceptual y metodológico, con instrumental cuantitativo y cualitativo, al que se agrega el manual de Hugo Navarro, ya referido.

México

El Sistema de Monitoreo y Evaluación de programas sociales federales en México¹⁶ prevé distintos tipos de evaluación (incluido impacto) que se precisan en los Lineamientos¹⁷ publicados en 2007 por su órgano rector, el Consejo Nacional de Evaluación de la Política Social (Coneval), conjuntamente con la Secretaría de Hacienda y Crédito Público (SHCP) y la Secretaría de la Función Pública (SFP).

Los Lineamientos establecen que la evaluación de impacto mide cambios en los indicadores de resultados atribuibles a la gestión del programa, con base en metodologías que serán revisadas por la Secretaría de Desarrollo Social (Sedesol), la SFP y el Coneval. Con objeto de especificarlas fue publicado un guión¹⁸ que inicia directamente con las técnicas cuantitativas, como si fueran únicas, y explica las características de las herramientas estadísticas disponibles para la aplicación del modelo experimental o cuasiexperimental, con diseño del contrafactual, que pueden incluir trabajo de campo o sólo utilizar información ya generada.

Con base en la información que cada programa interesado en realizarla presenta al Coneval, éste emite una opinión sobre su factibilidad. Existen programas que han intentado realizar evaluaciones de impacto pero no han podido cumplir los requisitos del diseño experimental o cuasiexperimental exigido (Programa de Conservación para el Desarrollo Sostenible, Sistema Nacional de Investigadores, etcétera).¹⁹ En estos casos el Coneval no las ha autorizado, ni ha reconocido las realizadas con métodos alternativos.

¹⁶ Nos referiremos exclusivamente al sistema de evaluación de programas ligados a la lucha contra la pobreza, si bien existen otros como el educativo, y los creados en el ámbito local, entre los que sobresale el de la Ciudad de México, conocido como Evalúa DF. Véase una exposición más amplia en Myriam Cardozo, “El estado del arte de la evaluación de políticas y programas públicos en México”, *Studia Politicae*, núm. 34, 2015, pp.135-162.

¹⁷ Coneval/SHCP/SFP, “Lineamientos generales para la evaluación de los programas federales de la administración pública federal” [http://www.coneval.org.mx/rw/resource/coneval/eval_mon/361.pdf], fecha de consulta: 23 de abril de 2016.

¹⁸ Coneval, “Guion de análisis de factibilidad para llevar a cabo una Evaluación de Impacto” [http://www.coneval.org.mx/rw/resource/coneval/EVALUACIONES/EVALUACIONES_PROGRAMAS_POLITICAS_DS/Evaluaciones%20de%20Impacto/guion_analisis_de_factibilidad.pdf], fecha de consulta: 23 de abril de 2016.

¹⁹ El órgano de evaluación de la Ciudad de México (Evalúa DF) intentó una evaluación alternativa de impacto en el sector salud cuyos Términos de Referencia pueden consultarse en su página institucional.

Según la página electrónica institucional, a la fecha se han realizado 18 evaluaciones de impacto de programas federales entre 2007 y 2012 (con un promedio anual de tres) y ninguna posterior. Se distribuyen por sectores como aparece en el Cuadro 5, donde Desarrollo Social atiende 46.2% de la población en pobreza.²⁰

CUADRO 5
Evaluaciones de impacto mexicanas 2007-2016

Sector	Número de evaluaciones	Método
Desarrollo social	8	8 cuasiexperimentales: 3 de éstos combinados con estudios cualitativos, 1 de ellos con dos grupos experimentales con distinta exposición al programa
Economía	3	2 cuasiexperimental, 1 encuestas con seguimiento de indicadores
Trabajo y previsión social	2	Cuasiexperimentales
Educación pública	2	Cuasiexperimentales
Agrario	1	No fue posible hacerlo cuasiexperimental y se analizaron las encuestas que pudieron aplicar
Ciencia y tecnología	1	Cuasiexperimental
Salud	1	Diseño experimental que se contaminó y se aplicaron análisis complementarios

Fuente: elaboración con datos de Coneval, Evaluaciones de impacto [<http://www.coneval.org.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 22 de abril de 2016.

En respuesta a los Lineamientos, la metodología más utilizada ha sido la cuasiexperimental,²¹ en algunos pocos casos combinada con métodos cualitativos –uno de ellos el Programa de Desarrollo Humano Oportunidades (PDHO) 2009. Los únicos dos trabajos no experimentales se debieron a la falta de la información suficiente para proceder experimentalmente. Han participado diversas instituciones académicas entre las que destacan el Instituto Nacional de Salud Pública (35% de los casos).

²⁰ Coneval, “Medición de la pobreza” [http://www.coneval.org.mx/Medicion/MP/Paginas/Pobreza_2014.aspx], 20 de abril de 2016.

²¹ *Idem.*

Si bien México cuenta con más publicaciones sobre evaluación que los dos países previos, no se localizó ninguna especialmente referida a evaluación de impacto, sólo algún apartado en trabajos más amplios.²² Sus pocos programas formativos incluyen algún componente de impacto, con enfoques variados.

En síntesis, el estudio de los casos de Chile, Colombia y México muestra un fuerte énfasis en trabajos experimentales y cuasiexperimentales, que coincide con el identificado por la Cepal en 2005 para un conjunto mayor de países latinoamericanos. Esa homogeneidad metodológica se explica en buena parte por la influencia en ese sentido del Banco Mundial –a veces en colaboración con el Centro Latinoamericano de Administración para el Desarrollo (CLAD)– y de las experiencias de nueva gerencia pública de Gran Bretaña, Australia y Nueva Zelanda, a partir de sus asesorías e intercambio de profesionales. En un número importante de los casos latinoamericanos revisados por la Cepal, que han solicitado la asistencia técnica del mencionado Banco, éste aparece también como evaluador.²³

Sorprende que el promedio anual de evaluaciones de impacto ha sufrido un fuerte retroceso a partir de 2012 en los tres países latinoamericanos revisados: Colombia pasa de un promedio superior a 2, a otro inferior a 1; Chile de un promedio de casi 9, a sólo 0.2, y México de 3 a 0. No se dispone de información que explique esta coincidencia pero, aunque los motivos pueden ser diversos, es probable que esté influyendo la poca factibilidad de los diseños experimentales predominantes.

En el caso español, este énfasis metodológico se presenta fundamentalmente en uno de organismos autonómicos de evaluación y en el discurso predominante en ciertos frentes a los que aludimos en páginas previas. Es probable que en ese énfasis hayan influido la crisis económica, las disciplinas de procedencia de quienes la reivindican, pero sin duda también el deseo de acortar distancias con quienes cuentan con una mayor tradición en evaluación de impacto.²⁴

²² Banco Mundial, *La pobreza en México: una evaluación de las condiciones, las tendencias y la estrategia del gobierno*, Washington DC., 2004; Fernando Cortés, Agustín Escobar y Mercedes González de la Rocha, *Método científico y política social*, México, Colmex, 2008; Claudia Maldonado y Gabriela Pérez (comps.), *Antología sobre evaluación*, México, CIDE-CLEAR, 2015; Myriam Cardozo (comp.), *Institucionalización del proceso de evaluación. Calidad y utilización de sus resultados*, UAM/Cámara de Diputados, 2015.

²³ Hugo Navarro, *Manual de evaluación de impacto...*, *op. cit.*

²⁴ Se debe mencionar la creación, en 2007, del Fondo Español para Evaluaciones de Impacto (Spanish Trust Fund for Impact Evaluation, SIEF) en la Red de Desarrollo Humano del Banco Mundial, gracias a los fondos aportados por el gobierno español. La financiación se mantuvo hasta 2012, año en que fue asumida por la cooperación británica con el nombre de Strategic Impact Evaluation Fund (SIEF).

Sin embargo, la opción experimental preferida en los contextos revisados enfrenta fuertes dificultades metodológicas, económicas y éticas, razón por la que resulta muy relevante conocer otras alternativas útiles para los países analizados. Una rápida mirada a la producción mundial en el tema muestra una importante diversidad conceptual, teórica y metodológica, incluyendo estudios no experimentales poco conocidos o valorados en Iberoamérica, que puede complementar o sustituir a los trabajos experimentales, en función de las necesidades de cada caso. El resto del artículo se dedica a su exposición.

CONCEPTO DE IMPACTO

Se ha señalado que la evaluación de impacto en los países de Iberoamérica analizados tiende a partir de una definición de este término, no siempre explicitada, que se presupone universalmente aceptada. Relaciona los impactos con el cambio producido en una población, comparado con lo que habría sucedido si no se hubiera intervenido en ella (lo que se conoce como *contrafactual*) e incorpora al concepto el método que considera idóneo para su cálculo numérico (experimental) en función de los efectos netos que genera un programa en ciertos indicadores predefinidos.

Sin embargo, el examen detenido de la bibliografía sobre evaluación de impacto muestra una mayor diversidad de definiciones, entre las que además existen diferencias importantes, ligadas al enfoque teórico y metodológico en que se encuadran. Dicho examen nos permite identificar los tres elementos principales que permiten su caracterización, aunque no todas les conceden la misma importancia: *i*) el marco temporal; *ii*) su diferencia respecto del término *resultados*; y *iii*) el ámbito en el que se producen los impactos.

Las definiciones que contemplan la variable “tiempo” suelen construirse en torno a la idea de que los diferentes elementos de una intervención pública están vinculados entre sí en una suerte de progresión lógica y temporal (recursos, actividades, productos, resultados e impactos), ya sea de forma lineal, ya de forma más compleja. La referencia más conocida en este caso es la proporcionada por el Development Assistance Committee (DAC) de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), que define *impacto* como los “efectos de largo plazo positivos y negativos, primarios y secundarios, producidos directa o indirectamente por una intervención para el desarrollo, intencionalmente o no”. Los *resultados*, en cambio, son los efectos directos “de corto y medio plazo, logrados o probables, de los productos de

una intervención”.²⁵ En esta línea se sitúan también Rogers²⁶ y Peersman²⁷ al asumir la definición del DAC, puntualizando que los impactos se producen después y como consecuencia de los resultados.

Otras definiciones consideran, además de la idea de temporalidad y de su diferenciación respecto de *resultados*, el ámbito en el que se produce el impacto. Por ejemplo, para la Kellogg Foundation, el término *resultado* denota los cambios específicos en la población destinataria que se derivan de las actividades del programa y que se dan tanto en el mediano (entre 1 y 3 años) como en el largo plazo (entre 4 y 6 años); el de *impacto* se reserva para los cambios sociales futuros (a 7-10 años vista), previstos o imprevistos, que el programa genera.²⁸ Cabe también incluir aquí las contribuciones de Montague²⁹ y de Montague, Porteous y Sridharan,³⁰ vinculadas con el enfoque de las esferas de influencia y al mapeo de resultados (*outcome mapping*). En éste, los impactos se conciben como los efectos que se dan en la última de las tres esferas de influencia de un programa, la de control.

En el trabajo de Leeuw y Vaessen, uno de los más interesantes del panorama actual, la denominación de *impacto* cubre todo el abanico de efectos de la *cadena de resultados*, tanto en la población destinataria como en los contextos institucionales y sociales más amplios –esto es, incorporan los tres elementos mencionados pero no asocian *impacto* a un ámbito concreto ni sólo al largo plazo. Esto debido a que, a veces, el impacto surge inmediatamente y también debe ser cubierto por la evaluación.³¹

²⁵ DAC-OCDE, *Glossary of Key Terms in Evaluation and in Results Based Management*, París, 2010.

²⁶ Patricia Rogers, “Introduction to Impact Evaluation”, *Impact Evaluation Notes*, 1, Interaction/Rockefeller Foundation, 2012, p. 2.

²⁷ Greet Peersman, “Impact evaluation”, *BetterEvaluation* [http://www.betterevaluation.org/themes/impact_evaluation], fecha de consulta: 15 de mayo de 2016.

²⁸ Development Guild/DDI y Kellogg Foundation, *Evaluating Outcomes and Impacts: A Scan of 55 Leaderships Development Programs*, Brookline, Massachusetts, 2002; y Kellogg Foundation, *Logic Model Development Guide*, Battle Creek, Michigan, 2004.

²⁹ Steve Montague, “Focusing on inputs, outputs, and outcomes: are international approaches to performance management really so different?”, *The Canadian Journal of Program Evaluation*, 15(1), 2000, pp. 139-148.

³⁰ Steve Montague, Nancy Porteous y Anjeev Sridharan, “The Need to Build Reach into Results Logic and Performance Frameworks” [<http://www.pmn.net/wp-content/uploads/Build-Reach-into-Results-Logic.pdf>], fecha de consulta: 31 de mayo 2016.

³¹ Frans Leeuw y Jos Vaessen, *Impact Evaluations and Development Nonie Guidance on Impact Evaluation*, Washington DC., NONIE/IEG, 2009.

Por su parte, la propuesta de Hearn y Bufardi³² añade nuevos elementos a los tres ya indicados: uso o aplicación que se dará al concepto (proyectiva, prospectiva o retrospectiva), alcance (si incluye o no los impactos específicos, pre-establecidos y conocibles y los indefinidos, inesperados e imprevistos), sujetos y niveles a los que afecta (población objetivo, diferentes contextos micro y macro), vinculación entre actividades e impactos (directa e inevitable; directa pero incierta; moderada; e indirecta e incierta), temporalidad (inmediatos, estacionales y de corto, medio y largo plazo) y grado de permanencia (irreversibles, reversibles, duraderos pero reversibles).

Como se observa, existe una complejidad conceptual mayor que la rescatada en los países analizados, situación que se repetirá en torno a los enfoques y métodos que se revisan a continuación.

ENFOQUES DE LA EVALUACIÓN DE IMPACTO

En los países incluidos en el estudio prima el empirismo que promueve la evaluación basada en evidencias (*evidence-based evaluation*) y pone énfasis en el rigor científico de la medición cuantitativa y experimental. Nuevamente, no es el único enfoque posible, algunos de los enfoques alternativos más desarrollados se presentan a continuación.

a) Aunque con frecuencia se oye hablar del enfoque de la “teoría del programa” (*theory-based* y *theory-driven evaluations*), en puridad esta denominación cubre distintas aproximaciones,³³ entre las que destacan las pioneras de Chen,³⁴ Chen y Rossi³⁵ y la *teoría* del cambio de Weiss, la *evaluación realista* de Pawson y Tilley³⁶ y, más recientemente, los trabajos de White³⁷ y Rogers.³⁸

³² Simon Hearn y Anne L. Bufardi, *What is Impact?*, A Methods Lab publication, Londres, ODI, 2016.

³³ Para una clasificación tentativa de los enfoques, véase Frans L. Leeuw y Stewart I. Donaldson, “Theory in Evaluation: Reducing confusion and encouraging debate”, *Evaluation*, 21(4), 2015, pp. 467-480.

³⁴ Huey T. Chen, “Theory-driven evaluation: Needs, difficulties, and options”, *American Journal of Evaluation*, 15(1), 1994, pp. 79-82.

³⁵ Huey T. Chen y Peter E. Rossi, “Issues in the Theory-driven Perspective”, *Evaluation and Program Planning*, (12)4, 1989, pp. 299-306.

³⁶ Ray Pawson y Nick Tilley, *Realistic Evaluation*, Londres, Sage, 1997.

³⁷ Howard E. White, “Theory-Based Impact Evaluation: Principles and Practice”, *3ie Working Paper* 3, 2009.

³⁸ Patricia Rogers, “Theory of Change”, *Methodological Briefs. Impact Evaluation* 2, Florencia, UNICEF-Oficina de Investigación, 2014.

Todas estas aproximaciones buscan identificar y/o reconstruir la lógica, la cadena de resultados o las hipótesis que subyacen a la intervención, comprobar su solidez y abrir la *caja negra* que relaciona el programa y sus efectos, sin olvidar el papel que desempeña el contexto de la intervención. Se trata de una perspectiva totalmente abierta, que no exige ni excluye opciones metodológicas (Leeuw y Vaessen)³⁹ y que ofrece alternativas consistentes a los métodos experimentales y cuasi-experimentales.

Destacados autores –como White⁴⁰ o Leeuw y Vaessen⁴¹ y gobiernos, organismos y redes de investigación han incorporado este enfoque a su acervo teórico y/o a su práctica evaluativa con un reconocimiento equivalente al que se ha dado a enfoques y métodos más difundidos. Es el caso del Centre of Excellence for Evaluation del gobierno canadiense, la Comisión Europea en su Guía EVALSED de 2013, la International Initiative for Impact Evaluation (3ie) o la Oficina de Investigación de Unicef.

b) Relacionadas con la anterior, algunas otras propuestas descansan, en mayor o menor medida, en la búsqueda y el examen sistemático de la racionalidad del programa, si bien no se limitan a éstos. Entre las más conocidas se encuentran el enfoque de la complejidad, el análisis de contribución y los enfoques participativos –que se revisarán a continuación–, si bien cabe citar también el General Elimination Methodology (GEM) de Scriven (que se apoya en el principio de falsación de Popper)⁴² y el método de la averiguación (*elicitation method*).

El enfoque de la complejidad (*complex evaluations*)⁴³ –que tiene en el GEM su primer antecedente– concibe a los determinantes del impacto percibido como un sistema de causas interrelacionadas, capaz de provocar efectos no

³⁹ Frans Leeuw y Jos Vaessen, *Impact Evaluations and Development...*, *op. cit.*

⁴⁰ Howard E. White, “Theory-Based Impact Evaluation...”, *op. cit.*

⁴¹ Frans Leeuw y Jos Vaessen, *Impact Evaluations and Development...*, *op. cit.*

⁴² Michael Scriven, “Summative Evaluation of RCT Methodology: An Alternative Approach to Causal Research”, *Journal of Multidisciplinary Evaluation*, 5(9), 2008, pp.11-24.

⁴³ Rick Davies, “Scale, complexity and the representation of theories of change: Part II”, *Evaluation* 11(2), 2005, pp. 133-149; Ray Pawson, “Simple principles for the evaluation of complex programmes”, Amanda Killoran, Catherine Swann y Michael Kelly (eds.), *Public health evidence: tackling health inequalities*, Oxford, Oxford University Press, 2006; Neil Spicer y Penny Smith, “Evaluating complex, area-based initiatives in a complex of change”, *Evaluation*, 14(1), 2008, pp. 75-90; Patricia Rogers, “Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions”, *Evaluation*, 14(1), 2008, pp. 29-48; Kim Forss, Mita Marra y Robert Schwartz (eds.), *Evaluating the complex: Attribution, contribution and beyond*, Nueva York, Transaction Publishers, 2011; Barbara Befani, Ben Ramalingam y Elliot Stern, “Introduction. Towards Systemic Approaches to Evaluation and Impact”, *IDS Bulletin*, 46-1, 2015, pp. 1-6.

lineales (por ejemplo, sinérgicos). Dada esta complejidad, en la polémica *atribución-contribución* se inclina por la última –que se revisa en el siguiente apartado por estar muy ligada a la discusión metodológica. Los trabajos de Mayne⁴⁴ sobre análisis de contribución en contextos complejos están también entre las aportaciones más relevantes.

Los enfoques participativos constituyen otra de las alternativas más sugerentes y prometedoras en el campo de la evaluación de impacto. En sentido estricto, responden más a un principio inspirador para la realización de evaluaciones que a enfoques propiamente dichos. Su objetivo es implicar a los distintos *stakeholders* en el proceso evaluativo para que éstos contribuyan a enriquecer el trabajo con la pluralidad de sus puntos de vista –algo que el experimentalismo descarta e incluso rechaza.⁴⁵ La participación puede buscarse por motivos éticos, prácticos o, incluso, por ambos.

No se plantea *a priori* ninguna exigencia respecto de cómo entender la participación, quiénes serán los que participen o de qué modo lo harán. Estas cuestiones quedan abiertas y se deciden en cada caso particular. Los enfoques participativos no se vinculan tampoco con métodos o diseños concretos, ya sean cuantitativos o cualitativos, ni con una fase determinada de la evaluación, pero resultan mucho más viables y habituales en trabajos no experimentales, de menor complejidad técnica. Se han de mencionar aquí los estudios de Guijt,⁴⁶ Chambers⁴⁷ o Rogers,⁴⁸ en los que encontramos reflexiones teóricas, orientaciones prácticas para incluir y gestionar la participación en

⁴⁴ John Mayne, “Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly”, *Canadian Journal of Program Evaluation*, 16(1), 2001, pp. 1-24; “Contribution analysis: an approach to exploring cause and effect”, *ILAC Brief* 16, Roma, Institutional Learning and Change Initiative, 2008; y “Contribution Analysis: Addressing Cause and Effect”, Kim Forss, Mita Marra y Robert Schwartz (eds.), *Evaluating the complex: Attribution...*, *op. cit.*, pp. 53-96.

⁴⁵ Sandra Mathison, “Serving the Public Interest through Educational Evaluation. Salvaging Democracy by Rejecting Neoliberalism”, Katherine E. Ryan y J. Bradley Cousins (eds.), *The SAGE International Handbook of Educational Evaluation*, Thousand Oaks, CA, Sage, 2009, pp. 527-539.

⁴⁶ Irene Guijt, *Participatory Approaches, Methodological Briefs. Impact Evaluation 5*, Florencia, UNICEF-Oficina de Investigación, 2014.

⁴⁷ Robert Chambers, “Making the Poor Count: Using Participatory Methods for Impact Evaluation”, Robert Chambers, Dean Karlan, Martin Ravallion y Patricia Rogers, *Designing impact evaluations: different perspectives, 3ie Working Paper* 4, 2009, pp. 4-7.

⁴⁸ Patricia Rogers, “Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation”, Robert Chambers, Dean Karlan, Martin Ravallion y Patricia Rogers, *Designing impact evaluations...*, *op. cit.*, pp. 24-33.

las evaluaciones, así como descripciones de casos en que esta perspectiva ha sido empleada con éxito.

POSTURAS EPISTEMOLÓGICAS Y MÉTODOS PARA LA EVALUACIÓN DE IMPACTO

El debate sobre los métodos de evaluación de impacto es un reflejo de toda la discusión epistemológica generada en torno a las posibilidades y formas de acceder al conocimiento. En los últimos años ha girado en torno a dos polémicas centrales: atribución frente a contribución y experimentación frente a no experimentación.⁴⁹

En líneas generales, los defensores de la idea de “atribución” se han posicionado del lado del experimentalismo, que hemos visto predominar en los cuatro países estudiados, en tanto que los partidarios de la “contribución” se han opuesto al mismo; al tiempo, encontramos enfoques intermedios y también posiciones de quienes no participan de este debate:

a) Para los primeros, el objetivo de la evaluación de impacto es averiguar si existen relaciones de causalidad entre el programa y los impactos observados⁵⁰ y si éstos son atribuibles exclusivamente a la intervención y no a otros factores causales (efectos netos). El impacto es entendido como el cambio producido en una población, comparado con lo que habría sucedido si no se hubiera intervenido en ella (lo que se conoce como *contrafactual*). Se presupone asimismo que es posible medir sus efectos netos, para lo que es necesario recurrir a los diseños experimentales.

Es esta perspectiva sobre la evaluación de impacto –que podríamos denominar *técnica*, siguiendo a Bamberger⁵¹ la que ha recibido mayor

⁴⁹ Howard White, “A Contribution to Current Debates in Impact Evaluation”, *Evaluation*, 16(2), 2010, pp. 153-164; Elliot Stern, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies y Barbara Befani, *Broadening the range of designs and methods for impact evaluations*, DFID Working Paper núm. 38, 2012; Barbara Befani, Ben Ramalingam y Elliot Stern, “Introduction. Towards Systemic Approaches to Evaluation and Impact”, *op. cit.*; e IEG-World Bank, *World Bank Group Impact Evaluations. Relevance and Effectiveness*, Washington DC., 2012.

⁵⁰ Sobre el concepto de causalidad, Barbara Befani “Models of causality and causal inference”, Elliot Stern, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies y Barbara Befani, *Broadening the range of designs...*, *op. cit.*, pp. 92-126.

⁵¹ Michael Bamberger, “Institutionalizing impact evaluation. A key element in strengthening country-led monitoring and evaluation systems”, Marco Segone (ed.), *From policies to results Developing capacities for country monitoring and evaluation systems*, Nueva York, Unicef, 2010, pp. 189-217.

promoción en los últimos años, de la mano de organismos internacionales, fundaciones, *think tanks* y prestigiosos centros de investigación que la han presentado, no como una alternativa metodológica sino como la única opción válida para obtener conclusiones con validez científica.⁵² Un claro referente es el Abdul Latif Jameel Poverty Action Lab (J-PAL), una red internacional formada por cerca de 140 investigadores en evaluación aleatoria de impacto y cuyo centro operativo está asociado con el Massachusetts Institute of Technology (MIT).⁵³ Destacan también los trabajos realizados desde el Independent Evaluation Group (IEG) del Banco Mundial,⁵⁴ la Government Accounting Office (GAO)⁵⁵ –la agencia federal encargada de la evaluación y la auditoría fiscal y financiera en Estados Unidos–, la What Works Network (asociada con el gobierno británico), las contribuciones de Ravallion,⁵⁶ Baker,⁵⁷ Khandker, Koowal y Samad,⁵⁸ Gertler *et al.*⁵⁹ y las de la 3ie⁶⁰ –que si bien en sus inicios no apostaba por una opción metodológica concreta, más tarde se decantó por los diseños experimentales o cuasi-experimentales, de los que nos ocuparemos en breve.

⁵² Incluso a través de los medios de comunicación. Por ejemplo, James Ball, “Test policies in ‘randomised controlled trials’, nudge unit urges”, *The Guardian*, 20 de junio de 2012 o Clara Marín, “Un gran impulso científico para salir de la trampa de la pobreza”, *El Mundo*, 14 de mayo de 2015.

⁵³ Esther Duflo, Rachel Glennerster y Michael Kremer, “Using Randomization in Development Economics Research: A Toolkit”, *Discussion Paper Series* 6059, Centre for Economic Policy Research, 2007; y J-PAL, “Introduction to Evaluations” [https://www.povertyactionlab.org/sites/default/files/resources/Introduction%20to%20Evaluations.Final_.pdf], fecha de consulta: 31 de mayo de 2016.

⁵⁴ IEG-World Bank, *Institutionalizing Impact Evaluation within the Framework of a Monitoring and Evaluation System*, Washington DC., 2009; e IEG-World Bank, *World Bank Group Impact Evaluations...*, *op. cit.*

⁵⁵ General Accounting Office hasta 2004. GAO, *Performance Measurement and Evaluation. Definitions and Relationships*, Washington DC., 2011; y GAO, *Designing Evaluations. 2012 Revision*, Washington DC., 2012.

⁵⁶ Martin Ravallion, “Evaluating Anti-Poverty Programs”, Paul Schultz y John Strauss (eds.), *Handbook of Development Economics*, núm. 4 [<http://www.sciencedirect.com/science/article/pii/S1573447107040594>], fecha de consulta: 15 de mayo de 2016.

⁵⁷ Judy L. Baker, *Evaluación del impacto de los proyectos de desarrollo en la pobreza. Manual para profesionales*, Washington, DC., Banco Mundial, 2000.

⁵⁸ Shahidur R. Khandker, Gayatri B. Koowal y Hussain A. Samad, *Handbook of Impact Evaluation. Quantitative Methods and Practices*, Washington DC., Banco Mundial, 2010.

⁵⁹ Paul J. Gertler, Sebastian Martinez, Martin Premand, Laura B. Rawlings y Christel M.J. Vermeersch, *Impact Evaluation in Practice*, Washington DC., Banco Mundial, 2011.

⁶⁰ 3ie (International Initiative for Impact Evaluation), *Impact Evaluation Glossary*, 2012 [http://www.3ieimpact.org/media/filer_public/2012/07/11/impact_evaluation_glossary_-_july_2012_3.pdf], fecha de consulta: 15 de mayo de 2016.

b) Por su parte, quienes ponen el acento en la idea de contribución estiman que resulta imposible (por razones ontológicas y epistemológicas) y/o injustificadamente caro, distinguir los efectos netos que se derivan de un programa de los provocados por otras causas que inciden sobre el mismo asunto y población. El trabajo del evaluador consiste entonces en examinar el cambio producido por diversas causas (impacto de ese conjunto) y aproximarse a valorar y explicar la contribución de una de ellas (la intervención evaluada) al impacto del conjunto. Ante la imposibilidad de tener una certeza absoluta sobre la atribución de los impactos al programa, este tipo de análisis pretende disminuir al menos la incertidumbre. Para ello puede apoyarse en diferentes aportes teóricos, como la evaluación basada en la teoría o el enfoque de la complejidad.

La propuesta contributiva fue desarrollada durante la década de 1990 en el seno de la Oficina del Canadian Auditor General. Actualmente se encuadran en esta posición los trabajos de Mayne,⁶¹ su más destacado exponente, Kotvojs,⁶² Kotvojs y Shrimpton,⁶³ Dybdal, Nielsen y Lemire⁶⁴ o el manual sobre el tema publicado por el gobierno escocés.⁶⁵

c) Un tercer conjunto de aproximaciones parte de una idea menos restrictiva de atribución pero sin limitarse a una metodología experimental, como en el primer grupo; acepta que los impactos puedan tener otras causas además del programa y, por tanto, asume también el concepto de contribución. Insisten en la utilidad de los enfoques basados en la teoría, los estudios de caso o los métodos participativos para detectar y comprender los vínculos entre causas (programa) e impactos. Los trabajos de Rogers,⁶⁶ Stern *et al.*⁶⁷ o White⁶⁸ serían

⁶¹ John Mayne, "Addressing Attribution Through Contribution Analysis...", *op. cit.*

⁶² Fiona Kotvojs, "Contribution analysis: a new approach to evaluation in international development", trabajo presentado en la *Australian Evaluation Society International Conference*, Darwin, 2006 [<http://www.aes.asn.au/conferences/2006/papers/022%20Fiona%20Kotvojs.pdf>], fecha de consulta: 2 de junio de 2016.

⁶³ Fiona Kotvojs y Bradley Shrimpton, "Contribution analysis: A new approach to evaluation in international development", *Evaluation Journal of Australasia*, 7(1), 2007, pp. 27-35.

⁶⁴ Line Dybdal, Steffen Bohni Nielsen y Sebastian Lemire, "Contribution Analysis Applied: Reflections on Scope and Methodology", *The Canadian Journal of Program Evaluation*, 25(2), 2011, pp. 29-57.

⁶⁵ Gobierno de Escocia, "Guide 6: Contribution analysis", *Social Science Methods Series*, 2009.

⁶⁶ Patricia Rogers, "Overview: Strategies for Causal Attribution", *Methodological Briefs. Impact Evaluation 6*, Florencia, Unicef-Oficina de Investigación, 2014.

⁶⁷ Elliot Stern, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies y Barbara Befani, *Broadening the range of designs...*, *op. cit.*, p. 14.

⁶⁸ Howard White, "A Contribution to Current Debates in Impact Evaluation", *op. cit.*

los ejemplos más representativos. El último de los citados presenta una de las propuestas más equilibradas: considera que siempre que sea posible se debe recurrir al método experimental para atribuir los impactos a un programa; sin embargo, aboga por el uso de los enfoques basados en la teoría cuando se trata de explicar por qué una intervención genera un impacto.

d) Finalmente corresponde referir los trabajos que se desmarcan de la polémica atribución frente a contribución, algunos de los cuales ofrecen novedosas e interesantes aportaciones al tema. Cabría incluir aquí al United Nations Evaluation Group (UNEG), la red que agrupa a las unidades de evaluación de distintos departamentos, agencias, fondos, programas y organizaciones asociadas de la ONU. El propósito central de la evaluación de impacto para el UNEG es identificar las relaciones de causalidad que genera una intervención y mostrar, en su caso, cómo se han producido.⁶⁹

En forma paralela al debate presentado, se desarrolla el de los métodos experimentales y no experimentales retomados por la evaluación de impacto:

a) Como se ha mencionado, los autores que optan por el enfoque atributivo suelen considerar que los diseños experimentales son los únicos que aportan la evidencia científica que se requiere para el mejor diseño de las políticas (*evidence-based policy*). Se trata de medir la evolución del grupo experimental (atribuible al programa y otras causas) y la del grupo de control (sólo atribuible a otras causas) para calcular el impacto neto del programa como la diferencia de ambas mediciones. Los sujetos de uno y otro grupo han de ser asignados de forma aleatoria y ambos han de ser estadísticamente equivalentes en las variables que pueden interferir si se pretende afirmar con certeza que los impactos son atribuibles sólo a la intervención. Cuando la selección aleatoria no es posible puede sustituirse por métodos menos rigurosos como la integración por pares similares, pareo o emparejamiento (*matching methods*), y asumen la denominación de cuasiexperimentales.⁷⁰

Su ventaja comparativa es la medición rigurosa del impacto alcanzado y atribuido exclusivamente a la intervención evaluada, en aquellos casos en

⁶⁹ Algunas de las entidades del UNEG tienen su propia definición de impacto. UNEG, *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*, 2013.

⁷⁰ Donald Campbell y Julian Stanley, *Experimental and Quasi-Experimental Designs for Research*, Chicago, Rand McNally, 1973; Carol Weiss, *Investigación evaluativa*, México, Trillas, 1982.

que este diseño es factible.⁷¹ El problema es que son pocas las situaciones (pruebas piloto o programas con lista de espera que permiten diseñar controles adecuados) en que es posible contar con un grupo de control equivalente, que éste y el grupo experimental se mantengan completos y sin contaminarse durante el tiempo que el experimento requiera y que se disponga de las cuatro mediciones necesarias para calcular el impacto.⁷²

Además son diseños muy costosos, que predefinen los indicadores a medir con alto riesgo de no incluir los efectos inesperados, que no recuperan variables cualitativas ni explican cómo operan los nexos causales que conectan el programa y los niveles de impactos medidos (caja negra), y que enfrentan problemas éticos importantes como la exclusión de los beneficios del programa en el grupo testigo.⁷³

Por otra parte, los experimentos que se llevan a cabo en la realidad social no cumplen, en sentido estricto, con los criterios de los que se consideran *verdaderos experimentos* en ciencias naturales, a las que se ha tratado de emular. El contrafactual no remite a una realidad idéntica (comparable) a la que se da con la intervención,⁷⁴ desde el momento en que los mismos individuos no pueden recibir y no recibir el programa simultáneamente. Además, los destinatarios de los programas (en tanto que seres humanos con valores, creencias e intereses) introducen factores de variabilidad y particularidad que

⁷¹ Sobre los equívocos en torno al método experimental, Dean Karlan, "Thoughts on Randomized Trials for Evaluation of Development: Presentation to the Cairo Evaluation Clinic", Robert Chambers, Dean Karlan, Martin Ravallion y Patricia Rogers, *Designing impact evaluations: different perspectives...*, *op. cit.*, pp. 8-13.

⁷² Línea basal y medición al momento de evaluarse, tanto en el grupo experimental como en el de control.

⁷³ No faltan intentos para justificar ese tipo de decisiones. William D. Savedoff, Ruth Levine y Nancy Birdsall, "When Will We Ever Learn. Improving Lives through Impact Evaluation". Report of the Evaluation Gap Working Group, Washington DC., The Center for Global Development, 2006; Thomas D. Cook, Michael Scriven, Chris L.S. Coryn y Stephanie D.H. Evergreen, "Contemporary Thinking About Causation in Evaluation: A Dialogue With Tom Cook and Michael Scriven", *American Journal of Evaluation*, 31(1), pp. 105-117; José Ma. Larrú, "La evaluación de impacto: qué es, cómo se mide y qué está aportando en la cooperación para el desarrollo", José Ma. Larrú y Jorge Lugrís, *Dos modalidades de evaluación: evaluaciones de impacto aleatorias y evaluaciones participativas*, Documento de Trabajo CECOD, 1, 2008, pp. 5-24. Sobre los vínculos entre el experimentalismo y el pensamiento neoliberal, Sandra Mathison, "Serving the Public Interest through Educational Evaluation...", *op. cit.*, p. 532.

⁷⁴ Cook recuerda que los propios defensores de la experimentación admiten que el contrafactual perfecto no existe. Thomas D. Cook, Michael Scriven, Chris L.S. Coryn y Stephanie D.H. Evergreen, "Contemporary Thinking About Causation in Evaluation...", *op. cit.*, p. 111.

difícilmente los experimentos pueden controlar, por más que los dos grupos sean equivalentes (desde el punto de vista estadístico) en ciertas variables mensurables. Por otra parte, resulta casi imposible que, en la práctica, funcione el principio de doble ciego que ha de caracterizar a los experimentos o que no se produzca el llamado *efecto Hawthorne*.

b) El resto de los autores se inclina por otras alternativas metodológicas (no experimentales), que permiten identificar y explicar sólidamente las relaciones de causalidad apoyándose en la teoría del programa (*theory-based y theory-driven evaluations*), en enfoques participativos y en estudios de caso, que combinan técnicas cuantitativas y cualitativas como series de datos longitudinales disponibles con la recolección de datos complementarios por medio de encuestas de percepción, grupos focales o entrevistas. Es importante que se identifique el cambio con claridad, ya sea en forma contributiva o se analicen en profundidad otros factores diferentes al programa en estudio que hayan podido incidir en los resultados observados, si se busca la atribución a éste.

Las ventajas y desventajas de estas alternativas son las opuestas del anterior: la medición no es tan precisa porque su aporte no se cuantifica y parte de su información proviene de percepciones subjetivas, de ahí la importancia de su triangulación; pero son más factibles, incorporan información cualitativa con base en las percepciones de distintos actores (no sólo en las de los decisores que son las que se plasman en los objetivos de la intervención), permiten avanzar en la explicación necesaria para tomar decisiones (causas del impacto identificado), y evitan los problemas éticos. Utilizadas en sustitución de los diseños experimentales, permiten evaluar impactos cuando los requisitos de éstos no pueden cumplirse o se consideran una simplificación exagerada; empleadas como complemento de ellos, agregan información cualitativa que ayuda a identificar las causas de los impactos y permiten hacer las correcciones pertinentes.

Como apuntamos en otro lugar, “A pesar de los esfuerzos de la última década por reconciliar estas posturas, la controversia todavía muestra la presencia de comunidades científicas empeñadas en una especie de dogmatismo metodológico”;⁷⁵ sin embargo, sobresalen algunas iniciativas que toman lo mejor de cada corriente y lo integran en un enfoque mixto.⁷⁶

⁷⁵ Myriam Cardozo, *Evaluación y metaevaluación en las políticas y programas públicos. El estado del arte*, México, UAM-Xochimilco, 2012, p. 67.

⁷⁶ Frans Leeuw y Jos Vaessen, *Impact Evaluations and Development...*, *op. cit.*

EXPERIENCIAS ALTERNATIVAS DE EVALUACIÓN DE IMPACTO

La evaluación de impacto experimental y cuasiexperimental ha sido la más difundida tanto por organismos internacionales (Banco Mundial) como a nivel de los gobiernos nacionales (los casos nacionales presentados sirven de ejemplos, entre muchos más). Son también las más accesibles a través de las páginas web institucionales.

En cambio, las evaluaciones encargadas por organizaciones dispersas y especialmente las realizadas con enfoques alternativos, se vuelven mucho más difíciles de localizar y muchas veces sólo las conocemos cuando son relatadas en artículos de revistas especializadas. A pesar de esta dificultad, se han reunido algunas experiencias que ejemplifican su utilización en diferentes instancias, sectores y países.

Son muchos los ejemplos de casos exitosos reportados que se han basado en la teoría del programa. Merece la pena destacar la evaluación del Spring Sun Program, que pretendía abordar el problema del consumo de drogas entre los estudiantes de secundaria (*middle schools*) de Taiwan, y fue realizada por Chen;⁷⁷ la que tuvo como objeto el Bangladesh Integrated Nutrition Project (BINP), un programa piloto cuyo propósito era ofrecer orientación y alimentación suplementaria a mujeres embarazadas y a niños con problemas de crecimiento y/o malnutrición;⁷⁸ y la evaluación del programa conjunto de UNFPA-UNICEF sobre mutilación genital femenina.⁷⁹

A su vez, entre los casos ilustrativos que recogen las perspectivas de la contribución y la complejidad se encuentran la evaluación de la estrategia libre de humo en Ontario (Canadá), la vinculada al impacto en la pobreza de la construcción de infraestructura carretera en Tanzania, contratada por la Comisión Europea (CE) a Euréval (2007) y la relativa al crecimiento y el empleo, realizada por Euréval y Ramboll Management (2008). La primera combinó *Intervention Path Contribution Analysis* (IPCA) con la recolección sistemática de información de campo y la revisión de la literatura. La segunda retomó la

⁷⁷ Huey-T Chen, "Normative Evaluation of an Anti-Drug Abuse Program", *Evaluation and Program Planning*, 20(2), 1997, pp. 195-204.

⁷⁸ Howard White y Edoardo Masset, "Assessing Interventions to Improve Child Nutrition: A Theory-Based Impact Evaluation of the Bangladesh Integrated Nutrition Project", *Journal of International Development*, 2007, 19(5), pp. 627-652.

⁷⁹ United Nations Population Fund and United Nations Children's Fund, "Joint Evaluation, UNFPA/UNICEF", *Joint Programme on Female Genital Mutilation/Cutting: Accelerating Change*, 2008-2012, Nueva York, 2013.

compleja cadena de causas-efectos, reconociendo como antecedentes otros dos trabajos realizados en Ruanda (2006) y Jordania (2007), relacionados con la evaluación de los apoyos de la Comunidad Europea en el desarrollo y la pobreza rural, la reconstrucción de instituciones judiciales y el acceso de las empresas medianas y pequeñas al mercado de la Unión Europea. Una tercera, con un análisis muy similar al del enfoque contributivo, fue realizada por Lawson et al. también en Tanzania (2005).⁸⁰ Más recientemente se localizaron dos artículos: uno que relata el análisis contributivo realizado en el sector de salud pública australiano siguiendo el procedimiento enunciado por Mayne, y otro que lo vincula con la teoría del programa y la complejidad.⁸¹

CONCLUSIONES Y SUGERENCIAS

La evaluación de impacto realizada desde las instituciones en los casos nacionales revisados ha sido prácticamente identificada con el uso de estos modelos cuantitativos que, ante la imposibilidad de emplear el enfoque experimental, han recurrido al *second best* constituido por la aproximación cuasiexperimental y, excepcionalmente, la han complementado con métodos cualitativos para profundizar en la explicación de aquello que había sido simplemente medido. En el caso de España, la apuesta por la experimentación se advierte de forma clara no sólo en la práctica evaluativa regional sino también en el discurso de algunas instituciones y en foros diversos, incluidos los científicos y profesionales. En los tres casos latinoamericanos resalta también la poca presencia de programas de formación y bibliografía que difunda enfoques alternativos o complementarios, finalidad a la que el presente artículo ha intentado contribuir.

Si bien esta opción más difundida se limita a intentar medir el impacto en una relación de atribución “causa-efecto”, mediante métodos cuantitativos

⁸⁰ Robert Schwartz y John Garcia, “Intervention Path Contribution Analysis (IPCA) for complex strategy evaluation: Evaluating the Smoke-Free Ontario Strategy”, Kim Forss, Mita Marra y Robert Schwartz (eds.), *Evaluating the complex...*, *op. cit.* pp. 187-208; y Jacques Toulemonde, Douglas Carpenter y Laurent Raffier, “Coping with the evaluability barrier: Poverty Impact of the European Support at Country Level”, Kim Forss, Mita Marra y Robert Schwartz (eds.), *Evaluating the complex...*, *op. cit.*, pp. 123-144.

⁸¹ Janice Biggs, Louise Farrell, Glenda Lawrence y Julie Johnson, “A practical example of contribution analysis to a public health intervention”, *Evaluation*, 20(2), 2014, pp. 214-229; y Johannes Schmitt y Derek Beach, “The contribution of process tracing to theory-based evaluations of complex aid instruments”, *Evaluation*, 21(4), 2015, pp. 429-447.

y diseños experimentales y cuasiexperimentales, se ha mostrado que no existe un único concepto de impacto, un solo enfoque teórico ni un método indiscutiblemente mejor en toda ocasión, sino una amplia variedad de éstos. La revisión de las principales alternativas teórico-metodológicas y de casos en que éstas se han aplicado con éxito, aunque breve por las razones ya comentadas, ha evidenciado que existen enfoques y métodos alternativos sólidos y de gran potencialidad para el contexto analizado, que son mucho menos conocidos y casi no se utilizan, pero podrían aportar a la realización de un número mayor de este tipo de evaluaciones, imprescindibles para conocer si se han logrado mejoras en la calidad de vida de los beneficiarios, incluyendo su propia percepción al respecto.

No se aboga, en modo alguno, por sustituir una hegemonía por otra: tan conveniente es superar la actual equiparación de los métodos cuantitativos y/o experimentales con la evidencia científica “verdadera” como erróneo sería querer otorgarle a lo “no cuantitativo” esa misma posición de superioridad; defendemos, en cambio, la clarificación de los conceptos, el debate sobre los distintos métodos y enfoques para reconocer y admitir sus debilidades y fortalezas, el intercambio de experiencias y la difusión de alternativas que pueden ser provechosas –y que apenas se mencionan, al menos en el contexto latinoamericano.

Se trata, en definitiva, de introducir la diversidad presentada en este artículo en los cursos de evaluación impartidos en nuestras universidades latinoamericanas y españolas y de generar una discusión abierta e informada que permita valorar en cada caso el enfoque y método más apropiado.