

Modelos, límites y alternativas en la evaluación del impacto¹

Marçal Farré

Fundació Pere Tarrés, Universitat Ramon Llull
<marcal.farre@gmail.com>

Joan Cuevas

Fundació Pere Tarrés, Universitat Ramon Llull

Politiken eta gizartearen alorreko esku-hartzeen barruan, eskaera berriak agertu dira, hala nola gardentasuna, kontuak ematea eta emaitzak demostratzea. Nazioarteko eremuan, hainbat joera identifikatzen dira ebaluatzeko jardueri dagokienez, besteak beste: inplementazioan eta berehalako emaitzetan fokalizatzen direnak, eta medikuntzaren eremuko jarduerak antzeratzen dituzten metodologia esperimentalak. Hala ere, eztabaida bizia sortu da metodologia horien aldekoen eta kontrakoen artean. Polarizazio hori gainditzeko, premiazko zat jotzen dugu konsensua bilatzea bi alor hauei dagokienez: eragin-ebaluazioak zer egoeratan erabiltzea komeni den, batetik, eta horien ahalmenari etekin gehien nola atera, bestetik. Ildo horretatik, literaturaren errebasoa egingo dugu, eraginaren ebaluazioaren muga nagusiak jasotzeko eta ebaluazio mota hori zer egoeratan gauzatu litekeen definitzeko helburuarekin.

GAKO-HITZAK:

Ebaluazioa, eragina, erakundeak, aldaketaren teoria, irtenbideak.

Dentro del mundo de las políticas y la intervención social, en los últimos años se ha producido una emergencia de nuevas demandas, como la transparencia, la rendición de cuentas y la demostración de resultados. En el ámbito internacional, se identifican diferentes tendencias en lo que se refiere a las prácticas de evaluación, desde las que se focalizan en la implementación y los resultados inmediatos, hasta las metodologías experimentales que emulan prácticas del mundo médico. Sin embargo, se produce un debate intenso entre partidarios y detractores de estas metodologías. Para superar esta polarización, planteamos la necesidad de buscar un consenso sobre en qué situaciones es apropiado utilizar evaluaciones del impacto y cómo se puede sacar el máximo provecho de su potencial. En este sentido, hacemos un repaso de la literatura para recoger las principales limitaciones de la evaluación del impacto y definir los diferentes escenarios donde ésta se puede llevar a cabo.

PALABRAS CLAVE:

Evaluación, impacto, organizaciones, teoría del cambio, alternativas.

¹ Una versión previa de este trabajo se presentó en el V Congreso de la Red Española de Política Social (REPS) 'Desigualdad y democracia: política públicas e innovación social', celebrado el 5 y 6 de febrero de 2015 en Barcelona.

1. Introducción y contexto

En la última década han aparecido con fuerza los paradigmas de la transparencia y la rendición de cuentas en el mundo de las políticas y de la intervención social. Paralelamente se ha vivido una progresiva incorporación de sistemas de calidad y evaluación, tales como las normas ISO o EFQM, o la incorporación de cuadros de mando integral. En este contexto, diferentes actores clave del ámbito social han incrementado la demanda de evidencia en torno a la utilidad e impacto de las políticas. Esta demanda aparece tanto por parte de financiadores y donantes de programas de entidades del tercer sector como de la ciudadanía en general, del mundo académico e incluso de los mismos profesionales. En muchos casos, aún no existen referentes claros y con recorrido suficiente en este ámbito, o los que existen son poco conocidos.

El objetivo del presente artículo es aportar luz sobre la llamada evaluación del impacto, sobre sus modelos, conceptos clave y las finalidades de las diversas metodologías de evaluación. Más concretamente, se hace referencia a la evaluación del impacto en el ámbito de intervención compartido por las organizaciones del tercer sector y las administraciones públicas, es decir, el ámbito de las políticas sociales y comunitarias. Se trata de un sector de organizaciones muy heterogéneas e intervenciones muy diversas, en un contexto en que deben atender a unas crecientes necesidades sociales con unos recursos decrecientes.

A lo largo del presente artículo, se analizarán las diferentes modalidades de evaluación del impacto existentes, sus principales límites, los aspectos de debate o controversia, y finalmente se concluye con un capítulo de recomendaciones. De manera continuada, se hará alusión a la teoría del cambio, aludiendo al mapa conceptual que se puede crear para un programa social, y que recoge los principales aspectos de dicha intervención: los recursos económicos y de personal, los procesos básicos, las actuaciones, los resultados (*outputs*), los impactos inmediatos (*outcomes*) y los impactos finales. En las presentaciones y valoraciones de metodologías y modelos de evaluación, también se hará referencia

a la utilidad o finalidad de ésta, pues muchos de los responsables o profesionales de organizaciones con programas sociales o comunitarios necesitan analizar las diferentes estrategias de evaluación basadas en dichas finalidades. Esta utilidad puede bascular entre tres opciones: la transparencia y rendición de cuentas, el aprendizaje organizativo y para la toma de decisiones estratégicas sobre el modelo de intervención, y la comunicación y la captación de financiación.

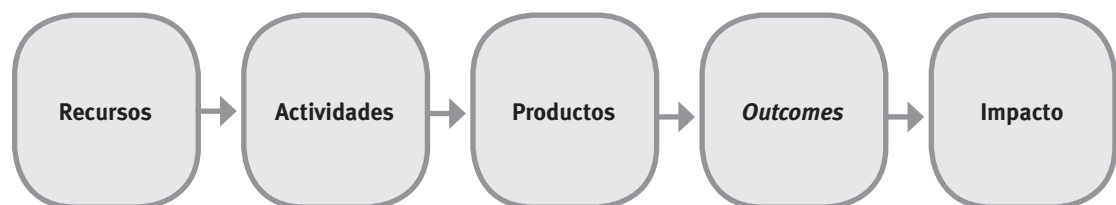
2. ¿Qué es la evaluación del impacto?

En la literatura acerca de la evaluación, podemos encontrar numerosas definiciones de impacto; la gran mayoría hacen referencia a los cambios en las vidas de las personas, generados por una acción, intervención, política o programa. Cualquier intervención, ya sea desarrollada por el sector público o el tercer sector social, se basa en una teoría más o menos explícita, que nos explica por qué la intervención produce un impacto sobre la realidad que queremos transformar. Esta teoría la podemos entender como una secuencia lógica de causalidades que se inicia con los recursos disponibles, pasa por las actividades que se desempeñan con dichos recursos, los productos generados, los resultados o *outcomes* y, por último, el impacto final. Dicho modelo, a menudo conocido como teoría del cambio, es la herramienta de referencia como paso previo al desarrollo del diseño de la evaluación.

En la Figura 1, se especifican los elementos que podrían llegar a conformar el conjunto de la teoría del cambio de una intervención y que se describen seguidamente de forma breve:

- **Recursos:** incluyen el conjunto de recursos de personal, económicos, de material y de equipamiento con los que cuenta la intervención o programa. Un ejemplo podría ser el número de aulas disponibles, el número de profesionales para un programa social determinado, o el grado de capacitación del personal.
- **Actividades:** como actividades se pueden entender tanto las acciones como las tareas o los

Figura 1. La teoría del cambio



Fuente: Elaboración propia.

procesos que serán ejecutados con el objetivo de generar los productos deseados. Ejemplos de estos elementos pueden ser las tareas de difusión de un curso, el procedimiento de inscripción de los usuarios, su acogida y las tareas de docencia.

- **Productos** (en inglés, *outputs*): se refieren al conjunto de resultados inmediatos de las acciones o actividades realizadas. Por ejemplo, en el caso de un curso formativo los productos podrían ser el número de asistentes o la satisfacción media de éstos.
- **Outcomes**: impactos más inmediatos, es decir, a corto plazo. Por ejemplo, en el caso de un programa educativo podría ser una mejora en el grado de inteligencia emocional de los usuarios, o en un programa de inserción laboral podría ser el grado de mejora de las competencias transversales.
- **Impacto**: debería hacer referencia a cambios observados a largo plazo en los destinatarios del programa o intervención. Un ejemplo podría ser una reducción del fracaso escolar en el conjunto de destinatarios de un programa socioeducativo, o el hecho de conseguir un trabajo de calidad y duradero en el caso de un programa de inserción laboral.

De hecho, si hablamos específicamente de evaluación del impacto, podemos entender el conjunto de la intervención como una combinación de dos teorías: la teoría de implementación y la teoría de impacto. La primera sostiene la hipótesis de que, con unos determinados recursos, la intervención va a ser capaz de generar unas actividades y productos concretos, mientras que la segunda asume que estos productos van a tener un impacto en la realidad sobre la que se quiere incidir. De esta forma, se trata de dos preguntas separadas, pero sólo si hemos respondido a la primera tiene sentido plantearnos la segunda, ya que la ausencia de impacto se podría deber al fracaso de cualquiera de las dos teorías. Por lo tanto, la evaluación del impacto identifica los cambios observados y responde a la pregunta de si éstos son resultado de la iniciativa evaluada, o bien si habrían ocurrido de cualquier modo (HM Treasury, 2011). De esta forma, no se valora el mérito o valor de la intervención, ya que ésta es una tarea del decisor (*policymaker*), pero se aporta evidencia para la toma de decisiones.

Actualmente podemos encontrar una multitud de modelos teóricos y prácticas evaluativas que provienen de tradiciones diferentes. En el siguiente apartado repasamos las principales tendencias.

2.1. Principales tradiciones en el ámbito internacional

Alnoor Ebrahim y V. Kasturi Rangan (2010) hacen un repaso a la literatura acerca de la evaluación del impacto, identificando tres corrientes o tradiciones, cada una de las cuales responde a necesidades diferentes. Por un lado, identifican a las grandes fun-

daciones filantrópicas norteamericanas (Robin Hood Foundation, Acumen), que optan por herramientas importadas del mundo de los negocios y la empresa, que les proporcionan información que permite comparar diferentes tipos de proyectos y escoger aquellos que tienen un mayor retorno social. Muchas de estas fundaciones intentan integrar diferentes métricas en sus procesos de decisión, pero la variedad de metodologías dificulta la comparabilidad entre los proyectos evaluados. Al mismo tiempo, los programas se evalúan de forma individual, y nunca se plantean un análisis conjunto, ni tampoco coordinar las ‘inversiones’ en una estrategia común entre las diferentes fundaciones. Un ejemplo paradigmático de la importación de herramientas y métricas del sector privado es la metodología SROI, inspirada en el tradicional indicador financiero ROI (retorno sobre la inversión), que pretende cuantificar el valor monetario del retorno social de las intervenciones evaluadas. Esta metodología fue desarrollada durante los años noventa por la norteamericana Roberts Enterprise Development Fund, con el objetivo de evaluar el valor generado por las iniciativas que ella misma había impulsado (Hall, 2012). Las evaluaciones SROI dan como resultado final un indicador clave en el que se defiende que, por cada unidad monetaria invertida en el programa, la sociedad se ve beneficiada por otra determinada cantidad. Por ejemplo, se podría decir que por cada euro invertido en un programa de refuerzo educativo, se genera un retorno social de 3,5 euros.

En segundo lugar, los autores identifican a las organizaciones del ámbito no lucrativo, que tienden a centrarse en aquello que queda bajo su control, es decir, en los recursos disponibles, las actividades desarrolladas y los productos resultantes, aunque ocasionalmente incluyen también indicadores de *outcomes* e impacto. Sin embargo, una de las críticas que se suele hacer a este tipo de evaluaciones es que, cuando intentan medir *outcomes* e impacto social, no son capaces de establecer si los impactos identificados se deben al programa evaluado, o si se habrían producido igualmente en su ausencia. En este sentido, cabe destacar propuestas como la *actionable evaluation*, desarrollada por la Bill and Melinda Gates Foundation, que se centra en medir sólo “aquello sobre lo que se puede actuar” (Bill & Melinda Gates Foundation, 2010). Esta propuesta intenta mantener la evaluación como una herramienta que aporte valor a las organizaciones en su funcionamiento operativo y su orientación estratégica, evitando invertir en conseguir información no relevante.

Finalmente, los autores citados identifican a las agencias de desarrollo internacional, que optan por metodologías experimentales, las cuales miden indicadores de impacto incorporando un contrafactual, o grupo de control, para representar lo que habría sucedido en ausencia de la intervención evaluada. De este modo, mediante la comparación de los indicadores de impacto del grupo de intervención y los del grupo de control antes y después de la actuación, se puede establecer si existe un efecto causal entre la intervención evaluada y el impacto observado se

puede establecer si existe un efecto causal entre la intervención evaluada y el impacto observado. Dentro de esta corriente, los ensayos controlados aleatorios², aparecen como la metodología de referencia (*gold standard*) para evaluar el impacto. Estos ensayos son una metodología importada del campo de la medicina, conocida por su uso para evaluar nuevos medicamentos y tratamientos médicos; en ese ámbito, resulta fácil seleccionar grupos de individuos y separar de forma aleatoria aquellos que reciben o que están expuestos al tratamiento de los que no lo están, asegurando así su comparabilidad.

3. Debate acerca de las metodologías experimentales

Investigar la causalidad con garantías exige disponer de un contrafactual o grupo de control, que nos indique qué es lo que habría sucedido en ausencia de la iniciativa evaluada, y que además los resultados obtenidos sean suficientemente robustos para considerarse válidos y fiables. Sin embargo, la disponibilidad de un contrafactual válido requiere que la selección de los dos grupos (intervención y control) sea aleatoria y previa a la intervención evaluada, de modo que se asegure su equivalencia estadística y se evite cualquier sesgo.

Teniendo en cuenta estas premisas, los ensayos controlados aleatorios o los diseños experimentales son los que tienen una mayor capacidad de estimar el impacto y establecer si éste es consecuencia de la intervención evaluada, lo que evita sesgos y relaciones espurias. Sin embargo, cuando estos ensayos no son posibles a menudo se opta por metodologías cuasiexperimentales, en las que se simula una selección aleatoria del grupo de intervención y del grupo de control. En estos casos, se llevan a cabo ciertas pruebas para asegurar su comparabilidad, y aunque aportan menos robustez que los ensayos controlados aleatorios, los métodos cuasiexperimentales también ofrecen un grupo comparable que sirve como contrafactual. Finalmente, los métodos no experimentales, al no disponer de un contrafactual, a pesar de que pueden identificar asociación y correlación entre los resultados y la intervención evaluada, no pueden afirmar que esta asociación se deba a una causalidad entre ambos.

A pesar del entusiasmo que suscitan las metodologías experimentales, y del hecho de poder ver si una intervención realmente ‘funciona’, existe un intenso debate entre partidarios y detractores de dichas metodologías. Mientras que sus partidarios destacan que el rigor metodológico de estos métodos permite demostrar si lo que se está haciendo funciona realmente o no, algunos académicos alertan de su falta de pluralismo, y recomiendan incorporar otras metodologías que permitan comparar y triangular la información.

Esta confrontación es especialmente evidente en el mundo de la cooperación al desarrollo, donde podemos encontrar visiones muy polarizadas sobre la conveniencia o no de usar ensayos controlados aleatorios (Cohen y Easterley, 2009). Un ejemplo de sería la profesora del MIT Esther Duflo (2010), que argumenta que, si en el siglo XX esta metodología ha revolucionado la medicina, puede hacer lo mismo con las políticas sociales. En el otro extremo, encontramos críticas como las de Ravallion (2009a), del Banco Mundial, que nos alerta de las consecuencias de la creciente influencia de lo que él llama “los randomistas”, que provoca que sólo se preste atención a aquello que es susceptible de ser evaluado mediante métodos experimentales: “Los investigadores rechazan oportunidades de evaluar programas públicos cuando la aleatorización no es posible. Estudiantes de doctorado buscan algo que aleatorizar. Agencias filantrópicas son reticentes a financiar evaluaciones no experimentales”.

4. Límites de la evaluación del impacto

Tal como hemos comentado, a pesar de la creciente popularidad de las evaluaciones del impacto y los métodos experimentales, levantan un intenso debate entre sus más fervientes partidarios y los críticos. Creemos que, para poder determinar cuándo conviene usar cada tipo de evaluación, hace falta primero conocer sus limitaciones. A continuación, repasamos brevemente las limitaciones más importantes de dichas evaluaciones e intentamos ver cómo pueden superarse.

En primer lugar, exponemos algunas críticas recibidas por el enfoque empirista de las evaluaciones del impacto y del uso de instrumentos como la teoría del cambio; después entramos a analizar limitaciones de la medida de *outcomes* e impactos y el uso del contrafactual, así como su relación con la toma de decisiones; y finalmente repasamos diferentes escenarios, según la naturaleza de la intervención, que determinan diferentes necesidades y limitaciones.

4.1. El establecimiento de un relato cerrado

La definición del impacto y de la teoría del cambio constituyen la base de lo que se va a evaluar y de la identificación de indicadores, al establecer una cadena lógica de supuestos e hipótesis muy concretos sobre por qué una política o programa funciona y llega a tener impacto social. De esta forma, las evaluaciones del impacto tienden a cerrar la visión para permitir así seleccionar y transformar en indicadores lo que se considera más relevante. Sin embargo, algunos autores consideran que esta práctica simplifica en exceso la compleja realidad en que las políticas y los programas sociales operan: estableciendo un único relato sobre cómo funciona la intervención, se presenta la evaluación como un acto exento de conflictos y neutral, ignorando el proceso político implícito en la evaluación (Hall, 2012).

² *Randomized control trials* (RCT) en inglés.

Cabría considerar que pueden existir multitud de relatos acerca del funcionamiento de una política o programa social. Si bien las evaluaciones del impacto aportan información contrastable en relación a qué funciona y qué no, lo hacen reduciendo a una las posibles visiones que puedan existir acerca de la intervención.

A pesar de que las evaluaciones del impacto, debido a sus complejidades técnicas, a menudo no sean las mejor situadas para fomentar la participación de los actores, no hay ninguna razón para pensar que no puedan incorporarla, o intentar incluir de alguna manera sus valores y perspectivas en su diseño. En este sentido, cabe mencionar metodologías alternativas como el *most significant change* o el *process tracing* (Collier, 2011), que parten de tradiciones totalmente alejadas del empirismo de las metodologías experimentales, pero que son muy útiles para identificar secuencias de hechos, y pueden ser un complemento para ofrecer visiones mucho más amplias y enriquecedoras respecto al enfoque de la teoría del cambio. Asimismo, algunas críticas alertan de una reducción de información narrativa a favor de información numérica: las categorías estandarizadas de los cuestionarios a menudo no capturan detalles importantes y pierden profundidad narrativa. En este sentido, algunos autores apuntan a la necesidad de potenciar los métodos mixtos, que permiten integrar técnicas cualitativas y cuantitativas, aprovechando las fortalezas y minimizando las debilidades de ambas tipologías (Bamberger, 2012).

4.1.1. Tipologías de impacto y temporalidad

En la misma línea del apartado anterior, si hablamos específicamente de la definición del impacto, hay que considerar que pueden producirse impactos previstos en la teoría del cambio, así como otros no previstos que quizás ni siquiera estén relacionados con los objetivos de la intervención, pero que son relevantes de cualquier forma.

Al mismo tiempo, existe también una dimensión temporal del impacto. Entendiendo el impacto como un proceso dinámico inducido por una acción, éste puede tener diferentes ritmos, no tiene por qué ser lineal e incluso pueden existir puntos de inflexión que lo precipiten. Sin embargo, cuando hacemos evaluación del impacto no disponemos de esta información, solamente tenemos los indicadores en dos puntos temporales diferentes, lo cual puede darnos a entender una dimensión equivocada del impacto o, en el mejor de los casos, puede resultar insuficiente para comprender cómo funciona el mecanismo causal.

4.2. La generalización fuera del experimento

Si bien las metodologías experimentales y cuasiexperimentales son fuertes en cuanto a validez interna, de modo que pueden investigar la causalidad de los

cambios observados gracias a la disponibilidad de un grupo de control, no se puede decir que lo sean en cuanto a la validez externa: sus resultados no pueden generalizarse fuera del contexto del experimento con tanta facilidad, ya que el resultado del experimento no nos revela los factores que pueden estar influyendo en él. De esta forma, lo que funciona en un contexto, puede no hacerlo en otro. La validez externa depende en gran medida de que el contexto al que se quieren extender los resultados sea lo más similar posible a aquel donde se ha realizado el experimento, pero es posible que existan diferencias, conocidas o desconocidas, ya sea por parte de los beneficiarios, de los profesionales que desarrollan la intervención o de cualquier otro actor que intervenga en el proceso. Una respuesta a este problema por parte de los defensores de los ensayos controlados aleatorios es que el experimento debería replicarse en otros contextos, y hacerlo suficientes veces hasta que podamos sacar conclusiones generalizables. Sin embargo, sus críticos responden que, en ausencia de una teoría sobre qué factores contextuales importan, resulta más bien arbitrario decidir en qué contextos se debería repetir el experimento y cuántas veces habría que hacerlo para obtener un resultado satisfactorio (Cohen y Easterley, 2009).

Un claro ejemplo de dicho problema de validez externa lo podríamos encontrar en un programa piloto que dispone de un equipo motivado y bien capacitado que consigue superar todos los retos de implementación y generar un impacto significativo. Sin embargo, cuando se decide escalar la intervención a un nivel superior, nos encontramos con importantes cambios en la forma de implementar el programa, debido a cuestiones políticas, y a una plantilla preseleccionada que tiene motivaciones y capacidades diferentes al grupo de profesionales que implementó el piloto. En este caso, aunque conocemos el potencial de la idea de intervención, no podemos estar seguros que los resultados vayan a ser los mismos.

4.2.1. Relación entre evaluación y decisión

Siguiendo el hilo del apartado anterior, a pesar de la resonancia de conceptos como el de toma de decisiones basada en la evidencia (*evidence-based policy*), la relación entre evaluación y decisiones no es clara, y algunos autores cuestionan el uso que se hace de la evidencia en la toma de decisiones, sugiriendo que ésta no depende de un solo indicador empírico, sino que hay una diversidad de criterios políticos. En este sentido, los decisores suelen estar interesados en otras preguntas más allá del impacto que ha tenido una intervención en un contexto determinado (Vaessen, 2010, citando a Heckman, 1992; a Heckman *et al.*, 1997; y a Ravallion, 2009b), como si la intervención podría funcionar en un contexto diferente al actual, si se podría escalar a niveles superiores, o qué es exactamente lo que hace que funcionen, y cómo se podría aprovechar ese conocimiento para otras políticas o intervenciones.

Esto no significa que reportar el impacto no sea una información de gran valor. De hecho, sería raro el caso en que este dato no resultara de interés. Sin embargo, el valor de la evaluación del impacto y su influencia en la toma de decisiones puede ser mayor si, además de responder al ‘qué’ (impacto), se puede aportar información sobre el ‘cómo’. Esto se puede hacer en buena medida abriendo la ‘caja negra’ de los mecanismos causales, que se sitúa entre los productos y los *outcomes*, mediante diferentes acciones, como el análisis de contexto, el trabajo de campo cualitativo y la recolección de datos que permita comprobar las conexiones causales, la interacción entre variables y el análisis de factores (White, 2009). Esto nos debe permitir entender mejor cómo funciona la intervención y sus determinantes, de forma que podamos imaginar cómo funcionaría en un contexto diferente y cambiante.

4.3. Adecuación de la evaluación a diversos escenarios

Un primer obstáculo para la evaluación del impacto con contrafactual pueden ser los condicionantes éticos y morales que comporta la propia metodología. Es posible que la política o programa a evaluar esté interviniendo en una necesidad o contexto donde sea difícil constituir un grupo de control al que, aunque muestre una necesidad social clara, se le pida su participación en un estudio sin que reciba intervención o servicio. Es decir, puede ser complicado desde el punto de vista ético y moral negar el servicio a un grupo de personas para poder realizar el estudio.

Más allá de este primer escollo, los diferentes escenarios donde se quiera aplicar la evaluación pueden plantear otras complicaciones. Si analizamos la presencia de evaluaciones del impacto entre diferentes ámbitos o áreas de intervención, podemos observar cómo su presencia es muy desigual. Autores como Ravallion (2008: 6), refiriéndose a proyectos de desarrollo, hablan de un “sesgo miope” que favorece las evaluaciones de proyectos que dan resultados rápidos. Asimismo, Blattman (2008, cit. en Vaessen, 2010) habla de una “sobreevaluación” (*overevaluation*) de ciertas intervenciones económicas, educativas y sanitarias, y de una “infraevaluación” de intervenciones relacionadas con la promoción de la paz, la reducción del crimen o la gobernanza.

Si bien los ensayos controlados aleatorios pueden ser ideales para evaluar cierto tipo de actividades, como vacunas o transferencias financieras condicionadas, resultan más difíciles de aplicar en actuaciones donde el grupo de intervención y control son difíciles de aislar, como programas de derechos humanos, incidencia política o de infraestructuras.

A continuación intentamos describir diferentes escenarios, según el alcance de la intervención y la complejidad de la teoría del cambio.

4.3.1. Delimitación de la intervención

Una de las principales dificultades de las evaluaciones experimentales es la delimitación del objeto de estudio, y por consiguiente, del propio experimento o evaluación. Algunos autores apuntan a lo que llaman un “continuo de intervenciones” (Vaessen, 2010), en el cual resulta difícil analizar una iniciativa de forma aislada del resto. Los diseños experimentales o cuasiexperimentales resultan muy apropiados para iniciativas relativamente simples, con objetivos claramente definidos y llevados a cabo en un período de tiempo relativamente corto, lo cual facilita aislar y evaluar la intervención y descubrir la causalidad. Pero cuando la intervención forma parte de un conjunto mayor, o de un “continuo de intervenciones”, puede no tener mucho sentido analizarla de forma separada, ya que los cambios esperados dependen de otros factores que no controlamos.

Podemos distinguir entre diferentes niveles de intervención (micro, meso y macro): a mayor nivel de intervención, mayor es el alcance, los instrumentos, el número de actores, la interrelación con otras intervenciones y, en definitiva, la complejidad de lo que se quiere evaluar.

4.3.2. Complejidad de la teoría del cambio

La complejidad de la teoría del cambio resulta otro factor muy relevante. La cadena de causalidades que liga la intervención con los impactos deseados puede ser muy sencilla, lineal y fácil de entender, como sería en el caso de las vacunaciones, en que una sola acción consigue el impacto; o puede ser complicada, difusa y difícil de entender, al combinar múltiples casuísticas y factores, algunos de ellos externos, y con un peso que no está claro. Un ejemplo de teoría del cambio compleja serían las campañas de sensibilización, de incidencia política o de cambio de las normas sociales, en las que no sólo resulta difícil medir los cambios observados, sino donde también es difícil que haya una teoría del cambio clara, consensuada y bien comprendida de cómo se consiguen los cambios que perseguimos.

4.3.3. Definición de escenarios y necesidades

Una vez repasados las principales características de las evaluaciones del impacto, y teniendo en cuenta las limitaciones expuestas, nos planteamos en qué casos es deseable hacer una evaluación y cómo delimitarla: ¿nos interesa conocer sólo los productos y resultados inmediatos o queremos adentrarnos a medir *outcomes* e impactos? Probablemente querremos saber el máximo posible sobre la intervención evaluada. Sin embargo, no siempre tiene sentido hacerlo así, ya que cada tipo de organización tiene unas necesidades y unas capacidades diferentes, y diferentes actuaciones nos plantean diferentes posibilidades y limitaciones.

En este sentido, Alnoor Ebrahim y V. Kasturi Rangan (2010) proponen cuatro escenarios básicos combinando los dos aspectos que hemos planteado anteriormente, la delimitación de la intervención –a la que ellos llaman “estrategia operativa”– y la complejidad de la teoría del cambio:

- **Estrategia operativa:** los autores diferencian entre intervenciones con una estrategia operativa centrada (*focused*) y compleja. La primera se refiere a organizaciones o intervenciones que se centran únicamente en tareas muy específicas, como podría ser el servicio de ambulancias, mientras que la segunda se refiere a organizaciones o intervenciones integrales, que combinan diferentes funciones dirigidas a un mismo impacto, como podría ser el caso de un hospital. En otras palabras, hay que diferenciar entre organizaciones que tienen control sobre el conjunto de actividades que generan el impacto, y organizaciones que tienen el control sólo de una pequeña parte de las actividades.
- **Complejidad de la teoría del cambio:** los autores diferencian aquellos casos en los que la teoría del cambio es muy simple y lineal, donde existe un único camino que conduce al impacto, como podría ser el caso de las vacunaciones, frente a los casos en que interactúan múltiples factores y donde varios caminos que conducen al impacto, como sería el caso de la lucha por modificar valores sociales, la promoción de los derechos humanos o la seguridad.

Combinando estos dos factores, establecen cuatro escenarios (Cuadro 1):

- **Resultados de nicho:** se trata principalmente de la provisión de servicios para dar respuesta a necesidades muy tangibles e inmediatas, pero que forman parte de una cadena mayor de intervenciones. En estos casos, tiene sentido medir las actividades y productos, así como la calidad del servicio, pero no evaluar *outcomes* e impactos. Un ejemplo podría ser el servicio de ambulancias.

- **Resultados integrados:** se trata de casos en que la teoría del cambio es simple y la estrategia operativa, compleja, ya que incluye funciones muy diversas. Generalmente se trata de prestar servicios básicos, como sanidad o formación profesional, pero se diferencian de los resultados nicho en el sentido de que la organización o intervención evaluada ocupa varias funciones dentro de la teoría del cambio, y gana peso sobre el conjunto de los productos que generan impacto. En estos casos, debido a la integración de diversas áreas, sería posible y tendría sentido medir *outcomes* e impactos mediante RCT o métodos cuasiexperimentales. Como ejemplo, podríamos señalar una campaña de vacunación, que incluye ciertas dificultades en la estrategia operativa, como la distribución de las vacunas, la sensibilización de la población y la coordinación con los diferentes servicios sanitarios.
- **Resultados institucionales:** sería el caso de intervenciones con estrategia operativa muy simple, pero teoría de cambio compleja, como campañas de sensibilización, de incidencia política o que buscan cambiar las normas sociales. Deberían medir productos e influencia social, pero no necesariamente impacto. Ejemplos de ello serían Amnistía Internacional o Greenpeace.
- **Resultados de ecosistema:** se trata de intervenciones que tratan problemas complejos y que al mismo tiempo tienen una estrategia operativa compleja, pues involucran a muchos actores, como sería la lucha contra la pobreza o procesos de empoderamiento colectivo. Además, los impactos se suelen producir a largo plazo. Se podría medir el impacto de forma conjunta, entre los diferentes actores que intervienen. Oxfam y Save the Children serían buenos ejemplos.

En resumen, a mayor complejidad de la teoría del cambio, más difícil resulta atribuir el impacto, y cuando la estrategia operativa es centrada, no tiene sentido medir el impacto, ya que hay elementos de la teoría del cambio que quedan fuera de la evaluación.

Cuadro 1. Marco de contingencia de Ebrahim y Rangan			
		Estrategia operativa	
		Centrada	Compleja
Teoría del cambio	Centrada	Resultados de nicho	Resultados integrados
	Compleja	Resultados institucionales	Resultados de ecosistema

Fuente: Elaboración propia a partir de Ebrahim y Rangan (2010).

5. Conclusiones

Como hemos visto, hay un debate intenso entre los que piden medidas claras y cuantificables de impacto para dirigir la toma de decisiones hacia lo que realmente funciona, frente a los que piensan que la complejidad en la que se desarrollan las intervenciones no reducirse a un cúmulo de indicadores. Para superar esta polarización, planteamos la necesidad de buscar un consenso sobre bajo qué escenarios es apropiado utilizar evaluaciones del impacto y cómo se puede obtener el máximo provecho de su potencial. Con estos propósitos, lanzamos unas ideas a modo de conclusión.

5.1. ¿Quién debe medir qué?

Teniendo en cuenta las diferentes necesidades y capacidades respecto a la evaluación del impacto, tal como hemos repasado en los apartados anteriores, pensamos que no es factible promover las evaluaciones del impacto a largo plazo para todo tipo de intervenciones. Si bien es cierto que no medir el impacto conlleva el riesgo de no poder demostrar el valor de lo que se está haciendo, medirlo sin suficientes garantías tiene el riesgo de atribuir a la intervención cambios sociales que se deben a otros factores, generar debates poco informados y gastar recursos que podrían dedicarse a otros fines. Consecuentemente, cada intervención debería evaluarse midiendo aquellos elementos que quedan bajo su control. En este sentido, el modelo y los cuatro escenarios definidos por Ebrahim y Rangan (2010) resultan de gran utilidad para discernir en qué casos la medición de indicadores de impacto con métodos experimentales y grupo de control es factible y puede aportar información de valor, y en cuáles nos deberíamos limitar a medir *outcomes*, actividades y productos.

5.2. Evaluación de actividades, productos y *outcomes* como herramienta de aprendizaje organizativo

De esta forma, en intervenciones con una teoría del cambio muy compleja o que sólo realizan funciones concretas dentro de una mayor cadena de actividades, el valor que la evaluación puede aportar tiene más que ver con un proceso continuo de aprendizaje que con la selección de un modelo de intervención. En estos casos, no merece la pena usar métodos experimentales o cuasiexperimentales, pero sí que tiene sentido hacer un seguimiento de las actividades o productos generados y, cuando sea posible, *outcomes* del beneficiario. Medir *outcomes* puede ayudar a la organización a detectar tendencias y a reflexionar sobre su contribución más allá del servicio que se ofrece, asegurando que los recursos y esfuerzos estén alineados con la misión que se propone. En cualquier caso, es importante mantener indicadores fáciles y simples de comunicar e implementar, que sean lo más compartidos posibles entre las diferentes partes involucradas, lo que contribuye

a alinear o integrar el trabajo de diferentes unidades, departamentos o hasta organizaciones.

Al mismo tiempo, es importante que el sistema de evaluación contribuya al aprendizaje y a la incorporación de mejoras más que a la cultura del control. El reto principal desde el punto de vista de este tipo de intervenciones es implementar sistemas ágiles que midan aquello que pueden controlar o influenciar, aportando información de valor sin suponer una carga excesiva. Se tiene que entender la evaluación como un camino progresivo de mejora, que empiece clarificando el modelo de intervención (teoría del cambio), avance hacia la recolección de datos de implementación (actuaciones y productos) y, más adelante, se plantee medir *outcomes*. Para garantizar el éxito de la evaluación, será importante que la teoría del cambio equilibre bien el objetivo de incorporar diferentes perspectivas con el de mantener una teoría del cambio no excesivamente compleja. En este sentido, será importante incorporar a los distintos actores implicados en la intervención y que aporten diferentes perspectivas, así como realizar una buena dinamización para crear el modelo y conseguir un constructo sintético y consensuado.

Por lo que atañe a los profesionales, los sistemas de evaluación también les pueden aportar información y conocimiento de valor. Si la evaluación permite disponer de información tratada que responda a necesidades de los profesionales de intervención, y este flujo de información es periódico y de fácil acceso, el profesional podría tomar decisiones como incidir más en determinados perfiles donde los resultados son más reducidos o proponer cambios en los procesos de atención que den respuesta a debilidades que se han detectado que reducen los resultados.

5.2.1. Evaluaciones del impacto con métodos experimentales como dirección estratégica

Las metodologías experimentales son herramientas que ofrecen una información única, que descubre los potenciales impactos de diferentes modelos de intervención. Esta información puede ser de gran utilidad no sólo para la propia organización que implementa la intervención, sino para una gran diversidad de actores, especialmente para la administración pública y las entidades del tercer sector social, que buscan cómo maximizar su impacto social. Sin embargo, la capacidad de generar evidencia científica sólo está al alcance de unas pocas instituciones, por razones económicas o capacidad de disponer de una muestra suficientemente grande. En nuestra opinión, siendo éste un bien tan escaso, la administración pública, conjuntamente con las entidades del tercer sector más capacitadas, deberían promover las evaluaciones del impacto en determinados ámbitos, con el fin de desvelar los impactos potenciales de diferentes modelos de intervención, y dotar así a las políticas públicas y al sector social de una dirección estratégica hacia donde dirigir sus esfuerzos.

Este tipo de iniciativas permitirían agrupar resultados en lo que podríamos llamar 'bancos de evidencias', lo que a escala internacional ya se promueve a través de proyectos como la International Initiative for Impact Evaluation.

Por otro lado, en lo que se refiere a servicios que proveen de forma fragmentada un amplio abanico de actores, cabe la posibilidad de incrementar el control sobre los impactos haciendo una evaluación integrada del impacto colectivo, ya sea a través de partenariados, alianzas o de las propias federaciones de organizaciones y entidades sociales. Éste es, sin embargo, un recorrido lento, que debe pasar primero por una agenda consensuada, una comprensión del problema común, sistemas de medición compartidos, una hoja de ruta conjunta, una comunicación constante y una secretaría bien equipada, con medios y competencias operativas (Fundación Bertelsmann, 2015).

5.3. Abrir la caja negra de la causalidad

Como hemos visto, las evaluaciones del impacto han sido criticadas por no aportar suficiente conocimiento sobre cómo y por qué se generan los impactos observados. Algunos autores, han insistido en la necesidad de combinar las metodologías experimentales con otras técnicas, lo que permitiría aportar un poco de luz a la caja negra de la causalidad. En este sentido, Howard White (2009) nos habla

de la *theory-based impact evaluation* (TBIE) como solución a estos problemas. Defiende que una buena evaluación del impacto debe cubrir tanto el impacto como la implementación, partiendo de una teoría del cambio bien desarrollada, con especial atención al nexo entre productos y *outcomes*. Además, pone énfasis en la recolección de datos, que debe permitir examinar los diferentes caminos causales, analizar los factores en juego, definir subgrupos que permitan anticipar la heterogeneidad de los potenciales beneficiarios y emplear métodos mixtos. En este sentido, el autor aboga por enriquecer las fuentes de información, incluyendo revisión de literatura proveniente de diversas disciplinas que puedan ser relevantes para el estudio (como la antropología y la economía política) e interactuando directamente con la intervención evaluada. Alternativamente, en cierto tipo de intervenciones en las que la teoría del cambio es muy compleja, con múltiples caminos causales, metodologías cualitativas como el *outcome mapping*, *most significant change* y el *process tracing* pueden ser una buena ayuda. La combinación de estas herramientas, además de la disponibilidad de un contrafactual, debe permitir entender mejor el impacto y por qué se produce. Aunque en teoría estas ideas tengan una buena aceptación, pocos estudios combinan la rigurosidad científica de los métodos experimentales con la capacidad explicativa de las herramientas descritas, y ésta supone una clara vía para incrementar el valor y la capacidad de influencia de las evaluaciones del impacto en la toma de decisiones.

Bibliografía referenciada

- BAMBERGER, M. (2012): *Introduction to Mixed Methods in Impact Evaluation*, serie Impact Evaluation Notes, nº 3, Interaction; The Rockefeller Foundation.
- BAMBERGER, M.; CARDEN, F.; y RUGH, J. (2009): *Alternatives to the Conventional Counterfactual*, Orlando, American Evaluation Association.
- BILL & MELINDA GATES FOUNDATION (2010): *A Guide to Actionable Measurement*, Bill & Melinda Gates Foundation.
- BLATTMAN, C. (2008): "Impact evaluation 2.0. Presentation to the Department of International Development, London, UK".
- CANYELLES, J. M. (dir.). (2008): "Gestió estratègica d'ONL: desenvolupar un model de gestió estratègica amb Quadre de Comandament Integral adequat al tercer sector", Barcelona, Fundació Pere Tarrés.
- CARRERAS, I.; IGLESIAS, M.; y SUREDA, M. (2011): *Liderazgo orientado a resultados en las ONG: estrategia, sistemas de medición y cuadros de mando*, Barcelona, Instituto de Innovación Social de Esade.
- COHEN, J.; y EASTERLEY, W. (2009): *What Works in Development? Thinking Big and Thinking Small*, Brookings Institution Press.
- COLLIER, D. (2011): "Understanding process tracing", *Political Science and Politics*, vol. 44, nº 4, págs. 823-830.
- DAVIES, R.; y DART, J. (2005): *The 'Most Significant Change' (MSC) Technique*.
- DUFLO, E. (2010): "Esther Duflo: Social experiments to fight poverty", serie TED Talks, TED [http://www.ted.com/talks/esther_duflo_social_experiments_to_fight_poverty].
- DUFLO, E.; GLENNERSTER, R.; y KREMER, M. (s/d): "Randomized Evaluations of Interventions in Social Service Delivery".
- EBRAHIM, A.; y RANGAN, V. K. (2010): *The Limits of Nonprofit Impact: A Contingency Framework for Measuring Social Performance*, serie Working Papers, nº 10-099, Harvard Business School.
- FUNDACIÓN BERTELSMANN (2015): *Trabajo colaborativo. Soluciones sistemáticas para problemas complejos*, Barcelona, Fundación Bertelsmann.
- HALL, M. (2012): "Evaluation Logics in the Third Sector", *Voluntas. International Journal of Voluntary and Nonprofit Organizations*, vol. 25, nº 2, págs. 307-336.
- HM TREASURY (2011): *The Magenta Book. Guidance for Evaluation*, Londres, HM Treasury.
- THE LANCET (2004): "The World Bank is finally embracing science" [editorial], *The Lancet*, vol. 364, nº 9.436, págs. 731-732.
- RAVALLION, M. (2009a): "Should the randomistas rule?", *Economists' Voice*, febrero.
- (2009b): "Evaluating three stylised interventions", *Journal of Development Effectiveness*, vol. 1, nº 3, págs. 227-236.
- (2008): *Evaluation in the Practice of Development*, serie Policy Research Working Papers, nº 4.547, Washington D.C., Banco Mundial.
- VAESSEN, J. (2010): *Challenges in Impact Evaluation of Development Interventions: Opportunities and Limitations for Randomized Experiments*, serie Discussion Papers, nº 2010.01, Amheres, Institute of Development Policy and Management.
- WHITE, J. (2009): *Theory-Based Impact Evaluation: Principles and Practice*, serie Working Papers, nº 3, International Initiative for Impact Evaluation.