

Diseños y métodos para la evaluación del impacto



**PABLO RODRIGUEZ
BILELLA**

Documento de trabajo
para EvalParticipativa
Número 2, Agosto de 2025

Rodríguez Bilella, Pablo

Diseños y métodos para la evaluación de impacto / Pablo Rodríguez Bilella. - 1a ed. – Rivadavia; 2025.
Libro digital, DOC

Archivo Digital: descarga

ISBN 978-631-01-0869-8

1. Evaluaciones. 2. Metodología de la Investigación. I. Título.
CDD 001.4

DISEÑOS Y MÉTODOS PARA LA EVALUACIÓN DE IMPACTO

Autor: Pablo Rodríguez Bilella

Diseño: María Clara Graffigna

Ilustraciones: Ana Clara Bustelo

Esta publicación es el resultado del proyecto “Fortalecimiento de la teoría y práctica de la Evaluación Participativa en América Latina y el Caribe: generación de una comunidad de práctica y aprendizaje”, desarrollado entre el Programa de Estudios del Trabajo, el Ambiente y la Sociedad (PETAS), Instituto de Investigaciones Socioeconómicas, Facultad de Ciencias Sociales, Universidad Nacional de San Juan (Argentina) y el proyecto Focelac+ para el fomento de capacidades y articulación de actores de la evaluación en América Latina del Instituto Alemán de Evaluación para la Cooperación al Desarrollo (DEval).

Esta publicación contó con el apoyo de:



© Pablo Rodríguez Bilella

© Programa de Estudios del Trabajo, el Ambiente y la Sociedad
www.petas-unsj.org / www.evalparticipativa.net

Instituto de Investigaciones Socioeconómicas
Facultad de Ciencias Sociales - UNSJ
Complejo Universitario 'Islas Malvinas'
Av. Ignacio de la Roza 590 (oeste)
Rivadavia - San Juan - Argentina.
CPA: J5402DCS
Email: petas.unsj@gmail.com
Tel: +54 (264) 4231949

ISBN 978-631-01-0869-8 (PDF/electrónica)

Se permite la reproducción total o parcial, citando apropiadamente su fuente. 2025

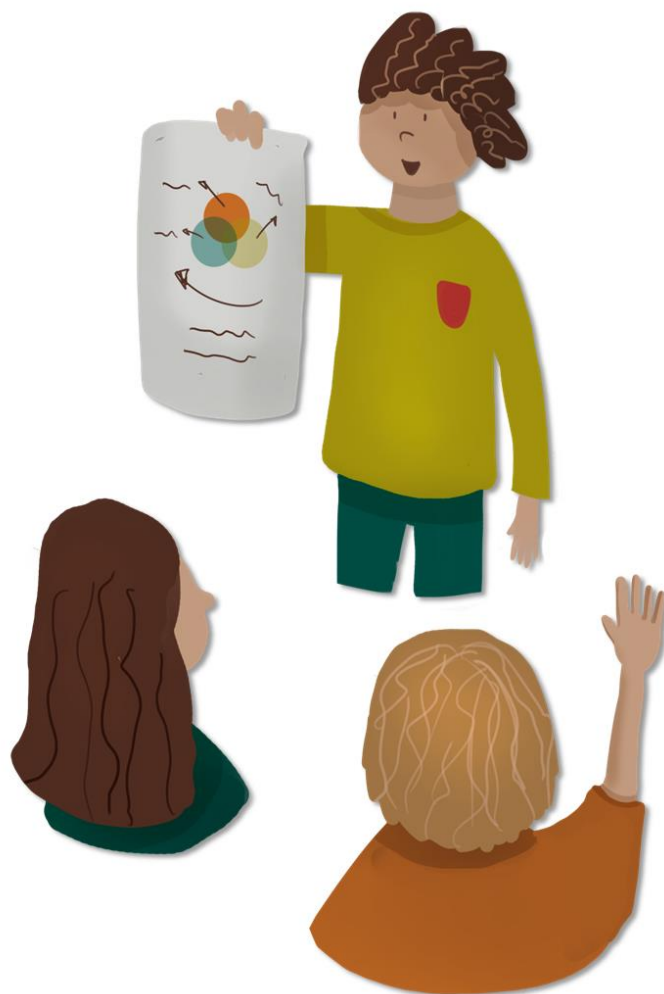
1. INTRODUCCIÓN

Durante las últimas dos décadas ha crecido en importancia y presencia el movimiento tendiente a alcanzar políticas basadas en la evidencia. Éste sostiene que los responsables de política deberían tomar sus decisiones sobre la base de la mejor evidencia disponible respecto a “qué funciona”, y no sobre la base de ideologías o en respuesta a intereses particulares. Uno de sus supuestos base es que no toda la evidencia ha sido lo suficientemente rigurosa como para brindar certezas en la toma de decisiones, por tanto se acentuó la orientación en pos de apuntalar determinados enfoques y metodologías que, dado su planteo *riguroso*, llevaran a resultados superadores. Este contexto general ha favorecido el debate y la discusión sobre la evaluación de impacto desde diferentes ámbitos y espacios (políticos, académicos, sociales, etc.).

Es así que el crecimiento en número de las evaluaciones de impacto en los últimos quince años ha sido notable, acompañado con la reflexión (y la preocupación) por la calidad de sus métodos, conclusiones y recomendaciones. Un elemento crítico recurrente da cuenta del alto costo que dichas evaluaciones suelen conllevar, junto a un contexto de realización donde son comunes las limitaciones de tiempo, información y, en ocasiones, respaldo político. Dadas dichas limitaciones, corresponde entonces reflexionar a fin de optar por el diseño más riguroso posible en función de las limitaciones existentes. Una buena evaluación de impacto sería aquella que alcanza el equilibrio entre los imperativos de utilidad y orientación práctica, rigurosidad, y respeto a los principios éticos.

Este documento de trabajo tiene por objetivo brindar una visión general e introductoria a la evaluación de impacto, ubicándola en el contexto de los debates contemporáneos que alrededor de ella se han suscitado. Para ello, se da cuenta de elementos que justifican su realización, en articulación con los aspectos de su oportunidad y de los actores relevantes para su desarrollo.

Especial atención se brinda a ciertos aspectos claves a ser contemplados en toda evaluación de impacto: explicitación de los valores; el desarrollo de una teoría sobre el cómo funciona la intervención; la medición o descripción de los impactos y otras variables importantes; explicación de la incidencia de los impactos; síntesis de los resultados; e informe y respaldo del uso de la evaluación. Una vez presentado este marco general, en la sección final se caracterizan brevemente distintos enfoques de diseño de evaluación de impacto.



2. DEFINIENDO LA EVALUACIÓN DE IMPACTO

La relevancia creciente que ha ganado la evaluación de impacto se corresponde con las razones que dan cuenta de su importancia. Conocer el impacto de una intervención particular (un programa, un proyecto) facilita la toma de decisiones respecto a su continuidad, su modificación, o bien su finalización. Cualquiera sea el curso a tomar, la evaluación de impacto debería facilitar el aprendizaje respecto a cómo replicar o aumentar la escala de una experiencia piloto, incluso brindando criterios para su adaptación a otro contexto. Junto con este componente de aprendizaje está el componente de rendición de cuentas de la intervención, tanto para los que la financian (donantes, recursos públicos, etc.) a fin de mostrar que los recursos se están invirtiendo prudentemente, como también hacia distintos actores sociales (comunidades, beneficiarios, organizaciones de la sociedad civil).

Un acercamiento general a la conceptualización de la evaluación de impacto señala que ella investiga los cambios provocados por una intervención. La escala de este tipo de intervenciones puede ser de distinto tipo: desde un proyecto pequeño y local sobre educación de adultos; un programa encarado por una ONG para el fortalecimiento de la sociedad civil; una secuencia de proyectos de administración de recursos naturales ejecutada en un área geográfica; o un conjunto de actividades concurrentes realizadas por diferentes organizaciones con el fin de mejorar la capacidad de una comunidad. La diversidad de definiciones existentes sobre la evaluación de impacto manifiesta distintos énfasis, lo cual tiene a su turno consecuencias en la selección de diseños y métodos para llevarla adelante.

Para el [Comité de Asistencia para el Desarrollo de la OECD](#), impactos son los efectos de largo plazo, primarios y secundarios, producidos directa o indirectamente e intencionalmente o no, por una intervención para el desarrollo.¹ Esta definición apunta a revelar diversos efectos, no sólo aquellos buscados explícitamente por la intervención, los cuales pueden ser tanto positivos como negativos para la situación en que se opera. Se reconoce que dichos efectos son *producidos* (de alguna manera causados) por la intervención, y se sugiere que pueden existir distintos tipos de vínculos entre dichos efectos y las diversas formas de intervenciones: proyectos, programas, políticas. Finalmente, hay un énfasis en los efectos duraderos o de largo plazo de las intervenciones.

En un sentido similar, la [Federación Internacional de Sociedades de la Cruz Roja y de la Media Luna Roja](#) define al impacto² como un examen de los efectos amplios y a largo plazo en individuos, comunidades e instituciones para los que contribuye una acción -efectos sociales, económicos, técnicos, ambientales-. Los impactos pueden ser inmediatos y de largo plazo, intencionales o no, positivos y negativos, de carácter macro y micro. Los estudios de impacto buscan responder la pregunta ¿qué diferencia real tuvo la acción realizada en la mejora de la capacidad de las comunidades para reducir su nivel de vulnerabilidad? ¿Cuántas personas han sido afectadas, y cómo se han beneficiado de la acción?

En la misma línea, la [Sociedad de Evaluación del Reino Unido](#) sostiene que impacto son “los efectos de un programa en la sociedad. Los impactos pueden ser positivos o negativos, previstos o imprevistos”.



¹ OECD-DAC. 2002. *Glosario de los principales términos sobre evaluación y gestión basada en resultados*. Paris.

² The International Federation of the Red Cross and Red Crescent. 2001. *Evaluation Operational Framework*.

El manual del Banco Mundial [La Evaluación de Impacto en la Práctica](#) sostiene que *las evaluaciones de impacto son un tipo particular de evaluación que intenta responder a preguntas sobre causa y efecto. A diferencia de las evaluaciones generales, que pueden responder a muchos tipos de preguntas, las evaluaciones de impacto se preocupan por saber cuál es el impacto (o efecto causal) de un programa sobre un resultado de interés. Sólo interesa el impacto del programa: el efecto directo que tiene en los resultados. Una evaluación de impacto analiza los cambios en el resultado directamente atribuible al programa.* (p.7)

En el año 2008, la [Iniciativa Internacional para la Evaluación de Impacto \(3ie\)](#) definía en su documento fundacional a la evaluación de impacto rigurosa como *el análisis que mide el cambio neto en los resultados (outcomes) para un grupo particular de personas que pueden ser atribuidos a un programa específico, utilizando la mejor metodología disponible, factible y apropiada tanto para la pregunta de evaluación que se está investigando como al contexto específico.* (p.2)

El trabajo de monitoreo y evaluación de proyectos, programas o políticas públicas ha tendido durante mucho tiempo a centrarse en dar cuenta de ciertos procesos o bien en la utilización de recursos (¿estamos ejecutando el programa en los tiempos planificados, con los costos pensados?), así como también en aquello que se hacía o realizaba (los llamados *outputs*: ¿cuántos talleres se dictaron? ¿cuánta gente participó?). Sin embargo, **el punto central en la evaluación de impacto apunta a conocer si la situación que dio origen a la intervención ha cambiado, y si la intervención ha tenido algo que ver con dicho cambio.** En otras palabras, si puede *atribuirse* o no -y en qué medida- a dicha intervención el cambio en la situación problemática.

Para algunas líneas o escuelas de evaluación de impacto la cuestión de la atribución se dirime buscando responder al interrogante ¿qué hubiera pasado si la intervención no hubiera tenido lugar?, lo cual va más allá de un análisis de la situación previa y posterior a la intervención. Estos interrogantes son los que justifican la aparición del concepto del *contrafactual* o escenario contrafáctico: ¿qué hubiera pasado con la situación de intervención de no haberse desarrollado el programa o proyecto?

Colocando a este concepto en el centro, otras definiciones manifiestan un foco más acotado en su concepción de la evaluación de impacto. Una de ellas es la adoptada por [USAID-Agencia para el Desarrollo Internacional de EE.UU.](#)-, la cual sostiene que *las evaluaciones de impacto miden el cambio en un resultado de desarrollo que se puede atribuir a una intervención definida; las evaluaciones de impacto se basan en modelos de causa y efecto y requieren un contrafactual creíble y definido rigurosamente para controlar factores distintos a la intervención que podrían explicar el cambio observado.* ([USAID Evaluation Policy: year one](#))

Un manual de la CEPAL ([Aedo, 2005, p.5](#)) sostiene que *La estimación de impacto intenta establecer la diferencia, en alguna variable que se ha escogido como indicador de resultados de un programa, entre la situación que presenta un individuo, o el cambio en ésta, después de haber participado en el programa versus la situación en que se encontraría, o el cambio en ésta, si no hubiese sido beneficiario.*

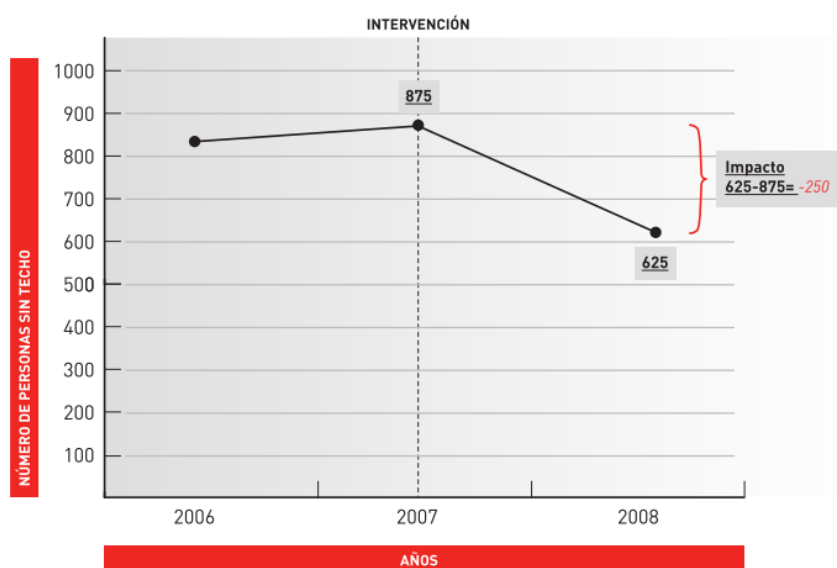
El Banco Interamericano de Desarrollo (BID) sostiene que la evaluación de impacto determina el efecto neto causal de una intervención en un indicador de interés. El abordaje de la noción de *efecto neto causal* requiere incorporar el concepto del contrafáctico, ya que el impacto neto que puede ser atribuido a la intervención es la diferencia entre el valor observado y el contrafáctico. Desde este punto de vista, evaluar el impacto de un programa implica ser capaz de aislar el efecto del programa en relación con otros fenómenos que afectan al problema o situación que la intervención pretende abordar.

El ejemplo desarrollado por [Blasco y Casado \(2009\) en su guía de Evaluación de Impacto](#) resulta muy pertinente para ilustrar este punto. Los autores presentan el caso de un programa de atención a las personas sin techo que pernoctan en las calles, llevado adelante por un ayuntamiento.

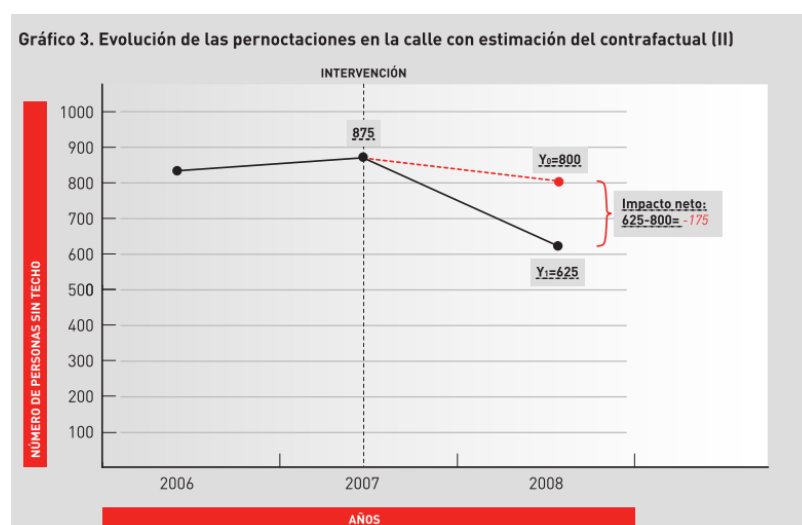
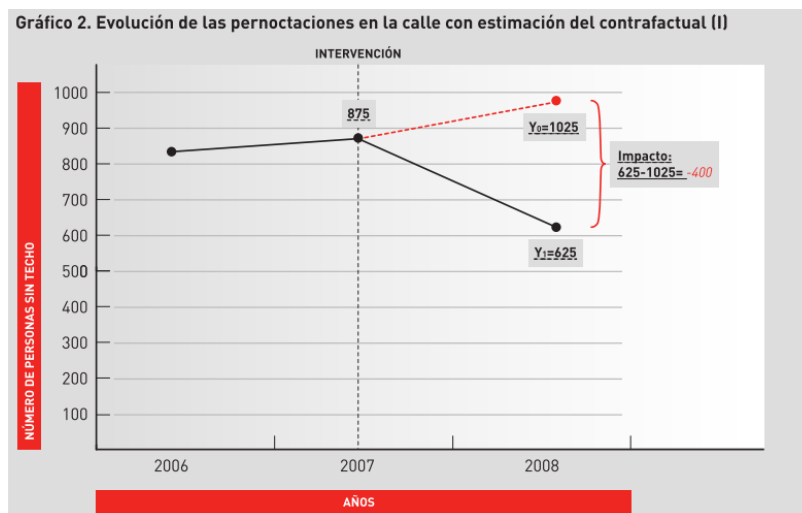
La intervención consiste en ofrecer atención personalizada en la calle a las personas que no utilizan los recursos municipales residenciales, con el propósito de que cada persona reciba siempre la atención del mismo trabajador social. El programa es costoso, ya que implica la contratación de numerosos trabajadores sociales nuevos, pero se espera que ayude a reducir considerablemente el número de personas que tienen que dormir a la intemperie. La teoría consiste en que, mediante este tipo de atención, el trabajador social desarrollará una relación de confianza con la persona sin techo que le permitirá detectar mejor qué problemas sufre, informarla, orientarla y acompañarla al recurso o servicio más adecuado en cada caso, e ir venciendo las barreras y desconfianzas que provocan que las personas sin techo se muestren reticentes a utilizar estos recursos. Se supone que de esta manera se incrementará el número de personas que entran en contacto con el sistema de atención, lo cual constituye un primer paso crítico para poder proporcionarles la asistencia que necesitan y, en último término, permitir que estas personas accedan a soluciones residenciales estables en las que puedan desarrollar su proyecto de vida con la mayor autonomía personal posible.

Imaginemos que este ayuntamiento, para medir el impacto de sus programas para las personas sin techo, realiza recuentos anuales de las personas que pernoctan en la calle. Tal y como muestra el gráfico 1, los recuentos realizados con anterioridad al nuevo programa indicaban que en el año 2007 había 875 personas durmiendo en las calles de la ciudad, lo que representaba una pequeña variación respecto al año anterior. El recuento del año 2008 — esperado con expectación para poder estimar el impacto del nuevo programa — revela que la población de personas sin techo ha bajado hasta las 625 personas. En otras palabras, esto implica una reducción de 250 personas, casi un 30% comparado con la población sin techo del año anterior. A primera vista parece que el impacto del programa ha sido positivo y considerable. Sin embargo, ¿podemos considerar que esta conclusión es suficientemente exacta?

Gráfico 1. Evolución de las pernoctaciones en la calle



Los gráficos 2 y 3 ... representan con una línea roja dos posibles contrafactuales. El primero se basa en la estimación de que, en ausencia de la intervención, el número de personas sin techo habría aumentado (esto correspondería, por ejemplo, a un escenario de más desempleo, peores servicios a las personas con enfermedad mental y más inmigración indocumentada). Por otro lado, en la estimación del contrafactual del gráfico 3 se asume que el número de personas habría disminuido igualmente en ausencia del programa (a causa, por ejemplo, de un escenario de menos desempleo, mejores servicios y menos inmigración). Obsérvese que Y1 no varía en ninguno de los dos gráficos: el programa comenzó, y luego se observó y midió el número de personas que pernoctaban en la calle después de la implementación. Por tanto, la divergencia de la magnitud del impacto en un gráfico y en el otro (400 personas en el gráfico 2 y 175 en el 3) se debe exclusivamente al hecho de que la estimación de Y0 (el contrafactual) es diferente.



Aquí aparece uno de los inconvenientes clave que presentan las evaluaciones de impacto: no es posible observar las mismas unidades (personas, hogares, escuelas, etc.) con y sin la intervención a la vez. Por tanto, hay que identificar o generar el contrafactual, el cual refiere al cálculo de lo que habría pasado si la intervención no hubiese ocurrido. Tal como Blasco y Casado (2009) aseveran, la bibliografía estadística y econométrica está repleta de estrategias para identificar el contrafactual, aunadas por el común reto de encontrar unidades de comparación (colegios, personas, barrios, etc.) que cumplan con la condición de reunir características similares a los que han participado en el programa. Por tanto, si bien no es una tarea libre de dificultades, existe una abundante producción bibliográfica al respecto.

Veíamos entonces que son diversas las organizaciones que sostienen que un elemento definitorio de la evaluación de impacto es la inclusión de algún tipo de **contrafactual**. Dicho planteo, a su vez, ha dado espacio a diversas polémicas las cuales se han reflejado en importantes debates metodológicos, los cuales presentaremos sucintamente en la siguiente sección.

3. DEBATES ALREDEDOR DE LA EVALUACIÓN DE IMPACTO

En círculos académicos y políticos se ha dado un intenso debate respecto a qué debería considerarse una evaluación de impacto rigurosa. Este debate ha tendido a polarizarse entre los que postulan la aplicación de métodos experimentales y cuasi-experimentales, versus quienes postulan una mayor

amplitud y pertinencia en el uso de otros acercamientos metodológicos.³ Los primeros han sido referidos en la literatura como “randomistas”, en referencia al apego a la metodología de las pruebas aleatorias controladas (*randomized control trials*, o RCT por sus siglas en Inglés), sosteniendo que la aleatorización es el único medio capaz de garantizar que el sesgo de selección no observable sea tenido en cuenta. Vale decir, la elección de quiénes serán beneficiarios de una intervención determinada versus la elección de quiénes conformarán un grupo control, debe ser hecha al azar y partiendo de un universo común de actores: todos deberían tener las mismas chances de terminar en un grupo o en el otro.

En la vereda opuesta a los *randomistas* se sitúan aquellos que sostienen que la aleatorización es excepcionalmente apropiada para ser usada en la evaluación de intervenciones de desarrollo, sugiriendo que apenas un 5% de los programas de desarrollo son factibles de ser evaluados mediante un diseño del tipo RCT.⁴ E incluso cuando sí es apropiado el uso de ese tipo de diseño, el uso de los contrafactuales apenas permite responder preguntas contingentes y asociadas a contextos particulares, y que sus hallazgos no pueden ser usados para generalizar en otros escenarios (lo que suele denominarse una baja validez externa⁵), a menos que sean acompañados por un conocimiento más minucioso de los mecanismos causales operantes en el proceso que va de la posible causa al efecto.⁶

En la esfera de la evaluación del desarrollo internacional, los avances realizados en el campo de la evaluación de impacto han sido importantes, si bien ha existido una mirada crítica de distintos actores y organizaciones de evaluación sobre el exagerado énfasis dado a los diseños experimentales y cuasi-experimentales de evaluación de impacto sobre otros. En este sentido, los contrafactuales - íntimamente asociados a la metodología RCT- son una de las formas más discutidas para pensar la causalidad en la evaluación de impacto. De algún modo, sus propulsores han logrado apoderarse del calificativo “riguroso” para caracterizar su metodología, en oposición a otras que usarían acercamientos más “blandos”.

Un acercamiento muy valioso a una reflexión más amplia sobre estos puntos ha sido el estudio encargado por el Departamento para el Desarrollo Internacional (DfID) del gobierno británico, titulado “[Ampliando el rango de diseños y métodos para la evaluación de impacto](#)” (material en Inglés), conducido por el Dr. Elliot Stern junto a un equipo de destacados evaluadores internacionales. Ellos dicen:

Pero los supuestos del pensamiento contrafáctico no siempre se sostienen (por ejemplo, puede ser difícil o imposible la búsqueda de una coincidencia idéntica para el mundo factual, esto es, el mundo en el que han sido la causa y el efecto observados), e incluso cuando lo hacen, los contrafácticos asocian una causa con un efecto determinado, sin proporcionar información sobre lo que ocurre en el medio: cómo es el efecto producido. Esta información puede ser importante para los propósitos de atribución, porque conociendo sólo el principio y el final del proceso

³ Los enfoques experimentales y cuasi-experimentales son muy similares en sus desarrollos, siendo la principal diferencia la perspectiva desde la cual se plantea la identificación de los grupos poblacionales factuales (aquellos beneficiarios de la intervención) y contra factuales (denominados grupos control y comparativo, respectivamente). Mientras que el acercamiento experimental procura una separación efectiva de la población, definiendo grupo factual y tratamiento o control antes de iniciar la intervención, la perspectiva cuasi-experimental busca recrear dicha distinción de beneficiarios y no beneficiarios con características similares que puedan ser comparados de forma unipersonal para establecer diferencias estadísticas en las variables que se desea testear.

⁴ Bamberger, M., Rugh, J. and Mabry, L. (2006) *RealWorld Evaluation*. Thousand Oaks, CA: Sage Publications.

⁵ Ravallion M (2009). *Should the randomistas rule? The Economists' Voice*. 6 (2): 1–5.

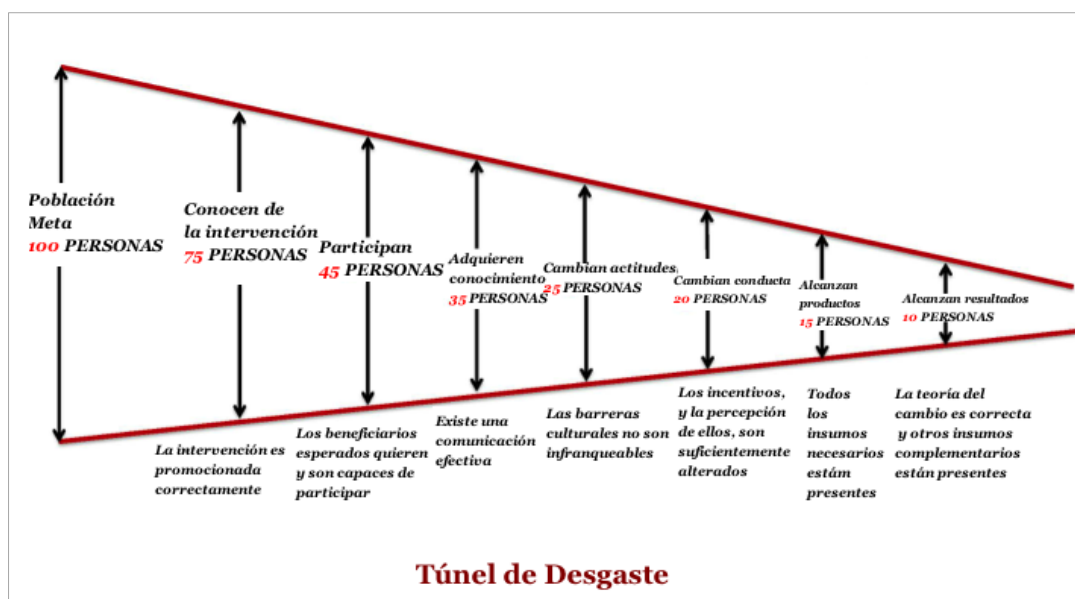
⁶ Un aporte adicional brinda la reflexión de Rick Davies sobre contrafactuales relativos versus absolutos: [Relative rather than absolute counterfactuals: A more useful alternative?](#)

causal puede no ser suficientemente minucioso en algunas situaciones. (Stern et al., 2012:7)⁷

Esta idea de la insuficiencia en conocer apenas los extremos del proceso causal es lo que ha sido referido muchas veces como “caja negra”, en la cual conocemos los insumos por un lado y los resultados finales por otro, pero el cómo se pasa de uno a otro se pierde en el medio, y queda indescifrable. En tal sentido es muy interesante ver algunas confluencias crecientes de parte de instituciones y organismos que han apostado fuerte por la preferencia hacia diseños experimentales y cuasi-experimentales, pero que empiezan a hacer explícito la necesidad de contar con una buena teoría del cambio.

En el caso del BID, [mencionado anteriormente](#), cuya definición de la evaluación de impacto rigurosa considera que esta apunta a determinar el efecto causal neto de una intervención en un indicador de interés, no limita la conceptualización a dichos elementos. Complementan su concepto de la evaluación de impacto sosteniendo que *no se pregunta únicamente si existió un efecto causal neto, sino también la causa por la cual se produjo (o no) dicho efecto. A tal fin, resulta importante explicitar la lógica del proyecto y analizar la existencia de otros mecanismos y variables.*

En el mismo sentido se explaya Howard White, director ejecutivo de 3iE, quien reconoce [la necesidad de una buena teoría del cambio para sostener los diseños de evaluación](#). En muchos casos, sino la mayoría, la teoría del cambio permite hacer explícitos los supuestos en las distintas etapas de la implementación del proyecto, si bien no suele recogerse información de modo sistemático a lo largo de la cadena causal. Para explicar este concepto, White presenta “el túnel del desgaste”.



Supongamos que hay una población meta de 100 personas para un proyecto. ¿Cuántas de estas 100 personas realmente conocen el proyecto? Y de ellas, ¿cuántas llegaron a participar? Y entre aquellas que participan, ¿cuántas realmente entienden lo que se supone deben hacer? ¿Cuántos de los que sí entienden están realmente de acuerdo y entonces ¿cuántos lo hacen? ¿Y haciéndolo permiten que se materialicen los resultados del proyecto? Y si es así ¿cuántas personas en realidad se benefician? La respuesta a todas estas preguntas se hace cada vez más y más estrecha. Afirmaciones

⁷ Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). BROADENING THE RANGE OF DESIGNS AND METHODS FOR IMPACT EVALUATIONS. En *DFID* (N.º 38; Número February 2011). <http://www.dfid.gov.uk/Documents/publications1/design-method-impact-eval.pdf>

grandilocuentes acerca de lo que el proyecto pretende alcanzar se realizan bajo el supuesto de que todas las 100 personas de la población meta experimentarán el cambio deseado en los resultados. Pero estas afirmaciones rápidamente se evaporan en el viaje a través del estrechamiento del embudo de desgaste.

La mirada hacia este embudo de desgaste facilita también visualizar quiebres y discontinuidades en la cadena causal, favoreciendo que emerjan preguntas y “baches” que deberían ser abordados desde la evaluación. Si queda claro que todos los insumos estuvieron presentes, ¿por qué no se alcanzó el producto esperado? O bien, ¿por qué si se llegaron a alcanzar los diferentes productos, no se produjeron los cambios o los resultados hacia las que tendía la intervención? Quizás la visión retrospectiva descubre la ausencia de elementos clave en la intervención, o bien identifica factores presentes que condujeron al éxito de la misma. El punto importante es que la respuesta a estas preguntas, que se derivan de la atenta mirada a la teoría del cambio de la intervención, tiene la potencialidad de conducir a una evaluación relevante y útil, que permita identificar no solamente si una intervención funcionó o no, sino cómo funcionó, para quién sí y para quién no, y porqué fue así. El otro elemento que aparece derivado del uso de una teoría de cambio estrechamente vinculada a la evaluación de impacto es que gran parte del análisis es factual (y no contrafactual), y si bien no hace uso de complejas operaciones estadísticas, postula un tratamiento atento y serio de los datos.

El abordaje de estas cuestiones apunta a destacar el importante rol de contar con un entendimiento adecuado de la evaluación de impacto, la cual no debe ser asociada únicamente con un cierto tipo de diseño o diseños considerados superiores a otros. El estudio antes citado de Elliot Stern y otros, sostiene que **la evaluación de impacto puede ser entendida como una forma de análisis causal que vincula a una intervención con efectos para los beneficiarios**. Reconoce la existencia de una pluralidad de formas para hacer inferencias asociadas a diferentes escuelas metodológicas y paradigmas, por tanto, la asociación de la evaluación de impacto sólo con el uso de los contrafactuales con el fin de atribuir los resultados a los programas, corre el riesgo de asociarla sólo a un método. Un diseño apropiado para la evaluación de impacto debe coincidir o ser coherente con las preguntas de evaluación, así como ser adecuados a los programas que están siendo evaluados. De este modo, es erróneo clasificar a los diseños de evaluación de impacto como experimentales o cuasi experimentales por un lado, y como “otras alternativas metodológicas” por otro.

Vale la pena explicitar que más allá de las limitaciones de los enfoques experimentales y cuasi-experimentales (de las cuales cuenta, al igual que los otros métodos) un aporte sustantivo realizado por quienes abogan por ellos es que han favorecido el surgimiento de interrogantes importantes sobre los métodos de evaluación de impacto en general. De alguna manera, los proponentes de la hipótesis e ideas experimentales han logrado desafiar a todos los evaluadores para hacer frente a las cuestiones de afirmaciones causales y explicar la efectividad de las intervenciones de desarrollo.

Tras dejar en claro lo que ha implicado este debate metodológico, entenderemos que **la evaluación de impacto incluye cualquier tipo de evaluación que investigue de manera sistemática y empírica los impactos que una intervención produce**. Esta visión amplia de la evaluación de impacto permite dar cuenta del interés en el énfasis en el largo plazo y el carácter complejo de las intervenciones de desarrollo. Respecto al énfasis en el largo plazo, se debe reconocer que los impactos han sido generalmente definidos como aquellos resultados alcanzados por una intervención en el largo término, pero sin embargo muchos de los ejemplos de evaluación de impacto actuales están apuntando a resultados intermedios o de corto plazo.⁸ El carácter complejo es reconocido a partir del espacio dado a la incertidumbre y la aparición de efectos no esperados (y no deseados) de las intervenciones de desarrollo. En tal sentido, y en aras al reconocimiento de dicha complejidad, comienza a verse un creciente interés en reconocer y comprender de qué modo intervenciones

⁸ Por ejemplo, las evaluaciones de los proyectos de transferencia condicionales de efectivo han sido realizadas en términos de si han aumentado la asistencia escolar y la retención, lo que es un resultado intermedio. Asistir a la escuela es una manera de lograr el impacto a largo plazo del aumento de las capacidades educativas (como la alfabetización y la aritmética), pero esto no se está mirando en términos de impactos a largo plazo.

particulares “contribuyen” al logro de cierto impacto, antes que centrarse en cuestiones solamente de “atribución”. Esto es particularmente visible en el caso de programas complejos e intervenciones de larga duración centradas en temas de gobernanza, democracia y rendición de cuentas, donde las estimaciones de “efectos neto” se hacen casi imposibles de abordar, reduciendo el énfasis en “medir el impacto”.⁹ Finalmente, el énfasis en el acercamiento sistemático y empírico deja el campo abierto a una pluralidad de diseños y métodos, sin buscar la equivalencia de la evaluación de impacto con uno de ellos en particular.

La discusión sobre la conceptualización de la evaluación de impacto no es un ejercicio especulativo, sino que se corresponde con la importancia notable que ha ido adquiriendo el medir el impacto de las acciones de desarrollo y de política pública en muy variados contextos. De acuerdo a la definición que se sostenga se corresponderá la decisión sobre los métodos o diseños de evaluación a adoptar, los cuales a su turno deberán recibir el visto bueno en cuanto a su calidad y credibilidad por parte de diferentes actores (los usuarios potenciales de la evaluación, los socios o donantes, los que solicitan la evaluación de impacto, etc.).

4. ¿QUÉ PREGUNTAS INTENTA CONTESTAR LA EVALUACIÓN DE IMPACTO?

Como cualquier otro tipo de evaluación, el diseño de la evaluación de impacto debe partir de clarificar cuál es el objetivo de la misma, así como definir quiénes serán sus usuarios primarios. Una vez dirimidas esas cuestiones de base, corresponde concentrarse en una cantidad limitada de **preguntas evaluativas clave**. Estas se denominan preguntas panorámicas, amplias, o de alto nivel, y no deben confundirse con preguntas específicas como las que el evaluador plantea en el uso de un instrumento como la entrevista o el cuestionario. Las preguntas deben estar estrechamente vinculadas con el objetivo de la evaluación de impacto, y para ello parte de la literatura señala que deberían ser ± 5 (vale decir, entre tres y siete preguntas).

Una reflexión pertinente en este punto es respecto al carácter explícitamente *evaluativo* que han de tener las preguntas de evaluación. Ello implica ir más allá de preguntas meramente descriptivas, del tipo *¿fue el programa ejecutado de acuerdo a lo planificado en su diseño?*, para avanzar hacia preguntas evaluativas que indaguen, por ejemplo, sobre cuán bien fue ejecutado, si lo fue de un modo consistente con normas y estándares profesionales, si lo hizo de un modo ético y culturalmente apropiado, etc.

Jane E. Davidson reflexiona lo siguiente en su mini-book [Principios Básicos de la Evaluación para la Acción](#) (2013):

Al mirar el alcance o cobertura del programa, una pregunta descriptiva indagaría a cuántas personas se ha alcanzado y / o cómo fueron alcanzadas. Una pregunta evaluativa indagaría, "¿Cuán adecuado fue el alcance del programa?" En otras palabras, no preguntaríamos solamente a cuántas personas se llegó, sino si se llegó a suficiente gente. ¿Se alcanzaron a las personas adecuadas? No preguntaríamos sólo por la forma en que se hizo, sino cuán bien fue hecho. ¿Utilizamos las mejores vías y métodos que podríamos usar? ¿Qué tan bien llegamos a poblaciones difíciles de alcanzar? ¿Accedieron quienes tienen mayor necesidad? ¿A quién no llegó, y fue eso razonable, ético, justo?

Al hacer preguntas acerca de los resultados, no importa solamente qué ha cambiado o si los objetivos y metas se cumplieron, la pregunta es: "¿Cuán sustanciales y valiosos fueron los resultados?" ¿Qué tan bien se cubrieron las necesidades más urgentes, y si se ayudaron a hacer realidad las aspiraciones más importantes? ¿Deberían los resultados ser

⁹ Valga esta breve referencia para señalar que un componente relevante de los debates sobre la evaluación de impacto ha pivotado alrededor de las nociones de “atribución” y “contribución”.

considerados verdaderamente impresionantes, mediocres, o inaceptablemente débiles? ¿Fueron no solamente estadísticamente significativos, sino también educacional, socio-económico, y prácticamente significativos? ¿Marcaron una diferencia real en la vida de las personas? ¿Valió la pena alcanzar los resultados, dado el esfuerzo y la inversión puesta en obtenerlos?

Ejemplos de preguntas clave de evaluación para la evaluación de impacto

Impacto general

- ¿Funcionó? ¿Produjo [la intervención] [los impactos deseados] a corto, mediano y largo plazo?
- ¿Para quién, de qué maneras y en qué circunstancias funcionó [la intervención]?
- ¿Qué impactos no buscados (positivos y negativos) produjo [la intervención]?

Naturaleza de los impactos y su distribución

- ¿Es probable que los impactos sean sustentables?
- ¿Estos impactos alcanzaron a todos los beneficiarios deseados?

Influencia de otros factores en los impactos

- ¿Cómo funcionó [la intervención] en conjunto con otras intervenciones, programas o servicios para lograr los resultados?

- ¿Qué ayudó o dificultó el logro de estos impactos por parte de [la intervención]?

Cómo funciona

- ¿Cómo contribuyó [la intervención] a los [impactos deseados]?
- ¿Cuáles fueron las características particulares de [la intervención] que marcaron la diferencia?
- ¿Qué variaciones hubo en la implementación?
- ¿Cuál ha sido la calidad de la implementación en diferentes sitios?
- ¿Hasta qué punto las variaciones de la implementación explican las diferencias en el impacto?

Relación de los impactos deseados con las necesidades

- ¿En qué medida los impactos coinciden con las necesidades de los beneficiarios deseados?

Rogers, Patricia J. *Introducción a La Evaluación De Impacto*, 2012.

Las preguntas de investigación deben ser las que orientarán la selección del mejor diseño y métodos para responderlas. Jonas et al¹⁰ señalan que el diseño de tipo RCT es capaz de responder al interrogante “¿funcionó la intervención?”, pero no a la pregunta “¿Cómo podemos mejorar la intervención? A su vez, Cartwright y Munro¹¹ (2010: 265) dan cuenta que la evaluación experimental puede responder si “¿funcionó aquí?”, pero no “¿funcionará en otro contexto?”. Para dicho propósito es que necesitamos teoría para juzgar qué factores tienen capacidades estables (para promover los resultados, y cómo) y para hipotetizar cuándo valen la pena implementarlos.

5. ACTORES Y OPORTUNIDAD PARA LA EVALUACIÓN DE IMPACTO

En relación a los actores capaces de realizar una evaluación de impacto, las posibilidades son básicamente tres:

- una evaluación externa, a cargo de un consultor o equipo contratado para llevarla adelante;
- una evaluación interna, a cargo de alguna unidad interna de la organización –la que no debería ser la misma que la unidad implementadora-; o bien

¹⁰ Jones, Nicola, Harry Jones, Liesbet Steer, and Ajoy Datta. 2009. *Improving impact evaluation production and use*. London.

¹¹ Cartwright N and Munro E (2010) The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice* 16: 260-266.

(c) un equipo combinado de evaluadores externos e internos.

En distintas organizaciones existe la idea generalizada que solamente las evaluaciones externas garantizarían la suficiente independencia y especialización -y por tanto credibilidad- de la evaluación de impacto. Sin embargo, la falta de conocimiento y atención a los matices de los contextos particulares de la intervención pueden en ocasiones dañar la credibilidad y validez de algunas evaluaciones. Ante ello se sostiene que el involucramiento de los actores del programa (funcionarios, técnicos, beneficiarios) ayudaría a subsanar dicho problema, aumentando así su credibilidad. Sin embargo, la simple participación de los involucrados –o el carácter mayormente interno de la evaluación- no son tampoco garantías por sí mismos de la validez de la misma. El factor clave reside entonces en la calidad del trabajo realizado, y los controles que del mismo se puede tener. Una alternativa al respecto es ahondar o articular instancias diversas de **triangulación**.¹²

Uno de los primeros autores en codificar la práctica de la triangulación desde las ciencias sociales fue Norman Denzin, quien la definió como "la combinación de metodologías en el estudio de un mismo fenómeno". Si bien la noción de triangulación ha tendido a estar asociada al calificativo "metodológica", Denzin señala la existencia de cuatro tipos básicos de triangulación, los cuales son relevantes para profundizar en el quién realiza la evaluación de impacto.

La **triangulación de investigadores** (o evaluadores, en nuestro caso) hace referencia al abordaje de una realidad o proyecto a evaluar desde las miradas múltiples de distintos observadores. Por un lado, aporta a la confiabilidad de los hallazgos al permitir que distintas miradas den cuenta del mismo objeto. El objeto construido por un equipo interdisciplinario, vale decir, la evaluación alcanzada de una intervención determinada, debe ser superadora de la suma de interpretaciones parciales (la del sociólogo, la del economista agrario, la del agrónomo, etc.). La incorporación de los beneficiarios en distintas instancias de procesos evaluativos, muchas veces entendidas como modalidades de "evaluación participativa", podría enmarcarse en la concepción de la triangulación de investigadores, siempre que los beneficiarios asuman roles y actitudes de evaluadores. No es la verbalización de opiniones sobre un proyecto lo que convierte a un beneficiario en evaluador, sino la articulación de dichas opiniones junto al análisis de los evaluadores.

La **triangulación de datos** refiere a tres tipos posibles. (a) La **triangulación de personas**, por la cual se accede a la información que brindan distintos actores en un mismo escenario, permite captar múltiples perspectivas de una misma intervención. (b) La **triangulación de espacios** contempla el acceso a un fenómeno en distintos ámbitos geográficos, tales como distintas poblaciones en una misma región, o distintas regiones en un proyecto nacional. (c) La **triangulación de tiempos o temporal** (estudios diacrónicos), si bien puede parecer más emparentada con una perspectiva de seguimiento y monitoreo que con una de evaluación ex post, es posible entender la realización de un diagnóstico de base como el primer acercamiento en el tiempo, y la evaluación final como el segundo.

La **triangulación teórica** contempla la posibilidad de hacer uso de diversos marcos conceptuales para dar cuenta del objeto en estudio. Dichos marcos conceptuales están insertos en paradigmas, los cuales son "una especie de creencia metafísica en que ciertas áreas de la naturaleza, implicando ciertas fronteras de atención y compromiso, son particularmente valiosas para el estudio; y que ellas forman la base para tradiciones específicas de investigación, en las cuales se elabora teoría y se reglamenta (es decir, se estandariza) la metodología para proveer un fundamento reconocido para la práctica de una disciplina particular" (Hewitt de Alcantra, 1982).

La decisión respecto a **cuándo realizar la evaluación de impacto** no debe estar guiada sólo por criterios burocráticos (porque el calendario de la planificación así lo indica) o meramente presupuestarios (pues contamos con los recursos para ello y hay que aprovecharlos). Con (cierta)

¹² Rodríguez Bilella, Pablo. 1999. "[Evaluación de Proyectos y Triangulación: acercamiento metodológico hacia el enfoque centrado en el actor](#)" En *II Taller Electrónico sobre Evaluación de Proyectos de Reducción de la Pobreza Rural*, PREVAL - IICA. San José, Costa Rica.

independencia del diseño y método elegido para realizarla, cualquier evaluación de impacto conlleva importantes esfuerzos, energías y uso de recursos, por lo cual la decisión del cuándo debería ser guiada por dos factores íntimamente relacionados: la necesidad de su realización, y la decisión firme de hacer uso de sus resultados. La elección del momento para la evaluación de impacto es muy importante: adelantarse a ese momento puede resultar en que la evidencia de los impactos ocurridos esté en curso y pueden resultar insuficientes; retrasar dicho momento conlleva dificultades logísticas (puede ser difícil realizar un seguimiento de los participantes) así como estratégicas (ser demasiado tarde para influenciar las decisiones acerca de la futura dirección de la intervención). En todo caso, planificar la evaluación de impacto desde el inicio de la intervención permite recoger evidencia a lo largo de su desarrollo, contar con datos de línea de base, e incluso permitir la utilización de métodos como las pruebas aleatorias controladas.

6. ASPECTOS CLAVE A CONSIDERAR EN LA ELECCIÓN DE MÉTODOS PARA LA EVALUACIÓN DE IMPACTO

Tal como se ha planteado anteriormente, en los últimos años se han producido acalorados debates alrededor de los métodos pertinentes, adecuados, preferibles o bien creíbles para la realización de evaluación de impacto. Conferencias, blogs, artículos académicos, libros, etc., han sido terreno de discusiones y pujas de visiones contrapuestas sobre qué debe ser considerada evidencia creíble, rigurosa y útil, así como también sobre quiénes son los actores que deben estar involucrados en la realización y control de las evaluaciones.

Veíamos previamente que este debate sobre métodos y diseños ha tenido como actor “en el centro del ring” al diseño de las pruebas aleatorias controladas (o RCT por sus siglas en Inglés, *randomized control trials*). Para sus defensores a ultranza, el mismo sería la “bala de oro” de la evaluación de impacto, y el único método o diseño que merecería tal nombre, o al menos el calificativo de “riguroso”. Una postura alternativa ha sido sostener la validez de la *adecuación situacional*, esto es, que los métodos y diseños de evaluación de impacto deben ser elegidos por su **adecuación al objetivo y a los tipos de preguntas** de evaluación que se realizan, a la **disponibilidad de recursos**, y a la **naturaleza de la intervención** (vale decir, si la misma es estandarizada o bien adaptable en relación al contexto de implementación).

Por tanto, la elección del diseño y métodos de la evaluación de impacto debe estar guiado por una reflexión sobre seis aspectos clave de la misma ([Rogers, 2012:6-15](#)):

- 1) **Clarificar valores:** La evaluación de impacto facilita conclusiones acerca del grado de éxito (o fracaso) de una intervención, por lo tanto es importante aclarar a qué se parece el éxito con respecto a:
 - i) lograr impactos deseables y evitar (o por lo menos minimizar) impactos negativos: Por ejemplo, ¿el éxito de un proyecto de desarrollo de carreteras será juzgado en términos de un mayor acceso a los mercados o un mejor acceso a hospitales de maternidad? ¿Qué nivel de pérdida del hábitat y de la biodiversidad sería considerado un costo razonable para la carretera? ¿Cuándo una concesión no sería aceptable?
 - ii) lograr una distribución deseable de los beneficios: Por ejemplo, ¿deberíamos juzgar el éxito en términos del resultado educacional promedio, por mejoras para los más desfavorecidos o colocar un grupo vulnerable o desfavorecido (como niñas) en el mismo nivel que sus contrapartes más favorecidas?

Existen diversas formas de abordaje para este punto de aclarar los valores para una evaluación de impacto, tales como la indagación apreciativa, encuestas en la comunidad, el método de El-cambio-mas-significante, el proceso Delphi, etc. (ver [Rogers, 2012:6-7](#)).

- 2) **Desarrollar una teoría de cambio:** hicimos referencia a este punto al comentar cómo en el caso del BID y de la organización 3iE consideran sumamente valioso contar con una teoría del

cambio como respaldo y sostén de cualquier evaluación de impacto. Una Teoría del Cambio suele incluir a la *teoría del programa*, la cual refiere al conjunto de supuestos sobre la forma en que el programa o proyecto provoca los beneficios que se espera que produzca, incluyendo las estrategias y tácticas que ha adoptado para lograr sus metas y objetivos. Muchas veces no es explícita, vale decir, no ha sido formalmente articulada o establecida, y ante su ausencia como tal uno se encuentra con una “evaluación de caja negra”, asimilable a querer hacer centro en un tiro al blanco con los ojos vendados. Por tanto, una tarea inicial en toda evaluación de impacto es re-constituir o explicitar dicha teoría de cambio. Algunos métodos para representar una teoría de cambio son el marco lógico, la cadena de resultados, el mapeo de resultados (ver [Rogers, 2012:7-8](#), como así también el sitio [Better Evaluation](#), haciendo uso de la opción de traducción al castellano).

Las teorías de cambio, modelos lógicos o teorías de programa pueden mejorar la evaluación de impacto al ayudar a:

- i) identificar los resultados intermedios o impactos que se pueden observar dentro del plazo de la evaluación y que son precursores de los impactos a largo plazo que se pretende alcanzar con la intervención.
 - ii) identificar, si una intervención no fue exitosa, en qué partes del proceso dejó de funcionar o fracasó.
 - iii) distinguir entre el fallo de implementación (donde no se han logrado los impactos porque la intervención no se implementó correctamente) y el fallo de teoría (donde la intervención no lleva al impacto deseado aún habiendo sido implementada correctamente).
 - iv) identificar los aspectos de la intervención que la hizo funcionar y son, por lo tanto, críticos.
 - v) identificar variables importantes de comportamiento y de contexto que se deben abordar en la recopilación de datos, en el análisis y en los informes para comprender variaciones en los impactos.
 - vi) brindar recomendaciones a los clientes sobre la base de juicios que dan cuenta porqué el programa funcionó (o no) de determinada manera.
- 3) **Medir o describir impactos y otras variables importantes**, incluyendo los procesos y el contexto. Las descripciones de los impactos no deben informar solo acerca del efecto promedio de tratamiento, esto es, el cálculo de la diferencia promedio que genera una intervención: por ejemplo, los campesinos participantes en el programa cultivaron 30% más que el grupo control. Lo que importa es también dar cuenta acerca de la multiplicidad de resultados y patrones presentes. La información sobre la implementación es clave para distinguir si lo que existió fue una falla en la implementación del programa, o bien si habiéndose implementado correctamente no se llegó al impacto esperado (una falla en la teoría de cambio). La información sobre el contexto permitirá comprender las posibilidades de replicación, al iluminar las condiciones en las cuales determinadas intervenciones funcionan y cuándo no. En tal sentido, los tomadores de decisiones y formuladores de políticas podrán contar con una gama alta de fuentes para entender cuándo funciona una intervención, para quién funciona, de qué modo lo hace, con qué facilitadores y qué obstructores, etc.
- 4) **Explicar si la intervención fue la causa de los impactos observados**. Sólo excepcionalmente puede considerarse que una intervención social sea la única causa de ciertos cambios producidos. Por lo general, las intervenciones sociales funcionan en combinación tanto con otros programas, con un determinado contexto, o bien con otros factores. En muchas circunstancias esto es olvidado u obviado por los actores institucionales de los programas y proyectos, quienes parecen tener un “espíritu fundacional” (*antes que nosotros, no había desarrollo*) o bien un excesivo énfasis en reconocer (midiéndolo) el aporte puntual y concreto

realizado por la intervención. Dada la complejidad que manifiestan los espacios sociales, hay un renovado énfasis en pensar a la “atribución causal” no tanto como “atribución total” (es decir, la intervención como única causa), sino como “atribución parcial”, llamándolas en ocasiones “posibles contribuciones”. De acuerdo a Patricia Rogers (2012:10):

*Puede ser útil investigar la atribución causal o la posible contribución en lo referente a tres componentes. **El punto de partida es lo factual:** comparar los resultados actuales con aquellos esperados si la teoría de cambio fuera cierta. ¿Cuándo, dónde y para quién ocurrieron los impactos? ¿Los resultados son coherentes con la teoría que la intervención causó o contribuyó a los resultados? **El segundo componente es el contrafactual:** un cálculo de lo que habría sucedido en ausencia de la intervención. **El tercer componente es investigar y descartar explicaciones alternativas.** En algunos casos, es posible incluir los tres componentes en una evaluación de impacto. En situaciones complejas, es probable que no sea posible calcular un contrafactual, y el análisis causal tendrá que depender de los otros componentes.*

Algunos de los métodos disponibles para el abordaje de estos tres componentes son: (a) para examinar lo factual: estudios de casos comparativos, dosis-respuesta, atribución del beneficiario/experto, predicciones, temporalidad; (b) para examinar el contrafactual: diferencia en diferencia, contrafactual construido lógicamente, comparaciones pareadas, diversas o cambiantes líneas de base, puntuaciones de la propensión, RCT, regresión discontinua, contrafactual creado estadísticamente; (c) métodos para identificar y descartar explicaciones alternativas: metodología de eliminación general, búsqueda de evidencia refutatoria, análisis de contribución, informes de resultados colaborativos.

Para una presentación breve de estos métodos, ver [Rogers, 2012:10-13](#).

- 5) **Sintetizar evidencia:** No es habitual que el juicio evaluativo general de una intervención se base en una sola medida de rendimiento. El procedimiento habitual implica por lo general el sintetizar evidencia sobre el rendimiento en diferentes dimensiones o criterios, desarrollando por ejemplo una escala ponderada (opción no exenta de inconvenientes¹³). Una alternativa es hacer uso de las *rúbricas de evaluación*, las cuales brindan un cuadro de cómo debería verse la evidencia en los diferentes niveles de desempeño. Un aspecto muy valioso de las rúbricas es que las mismas exigen el uso sistemático de la inferencia evaluativa para arribar a conclusiones. Por ejemplo, cuando se califica determinada dimensión de un programa como "buena", el evaluador debe demostrar de manera concreta qué evidencia lo llevó a creer que dicho desempeño fue "generalmente bueno" -en lugar de "claramente ejemplar" (excelente) o "inconsistente" (adecuado). Si el desarrollo de la *rúbrica* de evaluación global es realizado participativamente, se convierte en una forma muy valiosa de sintetizar evidencia con gran transparencia. Un desarrollo mayor del tema *rúbricas* puede consultarse en el libro [Principios Básicos de la Evaluación para la Acción](#) (Davidson, 2013).
- 6) **Informar hallazgos y respaldar su uso.** Una evaluación de impacto se plasma en un informe o reporte final, el cual debería contar con ciertas características comunes a cualquiera evaluación. Por un lado, procurar que sus contenidos sean rápidamente accesibles para los usuarios finales o clientes que demandaron la evaluación. Una forma pertinente para ello es organizar sus secciones como respuesta a las preguntas clave que guiaron la evaluación. Al mismo tiempo, debe contar con cierta orientación hacia el uso de su contenido, destacando

¹³ Ver la presentación de Michael Scriven y Jane E. Davidson, [The Synthesis Problem: Issues and Methods in the Combination of Evaluation Results Into Overall Evaluative Conclusions](#).

con claridad y concreción los hallazgos más relevantes. Su calidad y capacidad de uso posterior puede beneficiarse mucho si se suma a las partes interesadas pertinentes. En caso que se plasmen recomendaciones, las mismas deben estar articuladas con los hallazgos producidos en la evaluación y sortear de algún modo la barrera psicológica que implica una “recomendación externa”. En tal sentido algunos evaluadores consideran que su rol no comprende el de dar recomendaciones concretas, optando por plasmar líneas generales de acción, mientras que el curso concreto a seguir debe ser sentido y abordado por los actores pertinentes y relevantes de la intervención.

Para ahondar en este ítem, consultar la guía [Uso de los resultados de la evaluación de impacto, de David Bonbright \(2013\)](#).

7. UNA MIRADA A LOS ENFOQUES DE DISEÑO DE EVALUACIÓN DE IMPACTO¹⁴

Si reconocemos que existen métodos alternativos para llevar a cabo una evaluación de impacto, es posible considerar que los diseños más adecuados deben ser deliberadamente elegidos en cada caso. El contexto actual de desarrollo de la evaluación de impacto da cuenta tanto de nuevos paradigmas de análisis causal que rechazan la simple distinción entre los métodos cuantitativos y cualitativos, reconociendo la posibilidad de combinar diferentes métodos -cuantitativos y cualitativos, así como diferentes métodos cualitativos. De acuerdo con una perspectiva de métodos mixtos no se sugiere que los diseños sean meramente alternativos: la mayoría de las evaluaciones del mundo real implicarán diseños híbridos.

En las ciencias sociales la noción de "diseño" refiere a la lógica global respecto a cómo se realiza la investigación. En el caso de la evaluación de impacto adoptamos un foco más acotado, centrándonos en *las formas de la teoría y el uso de los datos* que apoyan la inferencia causal. Desde este punto de vista, los “experimentos”, las evaluaciones basadas en “teoría”, o los “estudios de casos” serían distintos ejemplos de enfoques de diseño. Como se explica a continuación, los diseños particulares dentro de estos amplios enfoques podrían incluir a los RCTs en la categoría general de los experimentos, el acercamiento “realista” en la categoría general de las evaluaciones basadas en “teoría”, y la etnografía en la categoría general de los estudios basados en casos.

Vale la pena señalar que diseños diferentes pueden compartir métodos y técnicas similares: tanto los enfoques experimentales como los estudios de caso pueden usar datos de entrevistas, basarse en grupos de discusión, y hacer uso del análisis de datos estadísticos. Finalmente, lo que mantiene coherente a un diseño es su referencia a la lógica de base del enfoque, antes que el uso de métodos particulares.

Ahora bien, esta clasificación tan clara y ordenada de los diseños no resulta así en la práctica. Los diseños de evaluación en el mundo real son rara vez tan “puros”, siendo común la presencia de diseños “híbridos” que



¹⁴ Esta sección se basa principalmente en el trabajo liderado por Elliot Stern (2012), “*Ampliando el rango de diseños y métodos para la evaluación de impacto*” (material en inglés). Los restantes autores involucrados en dicho estudio fueron Nicoletta Stame, John Mayne, Kim Forss, Rick Davies y Barbara Befani.

combinan, por ejemplo, los modelos estadísticos con estudios de caso en profundidad, o diseños participativos con encuestas estadísticas.

El análisis realizado por Stern y colegas sobre diversas *guías de evaluación de impacto* ha permitido detectar los énfasis y visiones predominantes en ellas, los cuales, a su vez, funcionan como parámetros para la realización de las evaluaciones. Algunos de sus hallazgos en relación a lo que dichas guías señalan son los siguientes:

- En la mayoría de ellas, los métodos experimentales suelen ser presentados como el "estándar de oro" para ser aplicado siempre que sea posible;
- Los métodos cualitativos son generalmente considerados como complemento de los métodos cuantitativos;
- Las evaluaciones basadas en teoría son vistas principalmente como una forma de interpretar los resultados cuantitativos, o bien una manera de profundizar la comprensión de los mecanismos causales derivados de experimentos. Muy ocasionalmente han sido consideradas como un enfoque central.
- Se tiende a asumir que sólo los métodos experimentales y estadísticos podrían ser *rigurosos*. En general este término se asocia con métodos contrafactuales, más que a la lógica subyacente de un enfoque o al modo que la evaluación de impacto es llevada a cabo.
- Los métodos participativos reciben relativamente poca atención. En ocasiones se sugiere que no resultan apropiados para la evaluación de impacto, mientras que en otros casos se reconoce su capacidad de producir visiones e ideas (*insights*), pero no llegan a calificar como rigurosos.

A partir de la revisión de la literatura metodológica (plasmada en las guías para la realización de evaluación de impacto) así como de casos reales de evaluación donde se utilizaron una "amplia gama" de métodos, el trabajo de Stern y colegas presenta una lista acotada de diseños y métodos de evaluación de impacto. La misma da cuenta en su primera columna de "enfoques de diseño", seguida de algunas "variantes específicas" dentro de cada enfoque, y señalando finalmente algunos elementos clave que sirven en cada caso como base para la inferencia causal (ver página siguiente).

Como se señaló al principio de esta guía, los enfoques estadísticos y experimentales cuentan con abundante bibliografía, la cual permite acceder y profundizar temas de estimaciones de contrafactuales, métodos de selección aleatoria, diseño de regresión discontinua, diferencias en diferencia, pareamiento, etc. Valga como recomendación inicial la valiosa [sección 3. Métodos para la Evaluación de Impacto, de la Guía práctica 5: Evaluación de Impacto](#) (Ivalúa, Barcelona:2009). Un tratamiento más amplio y detallado, fruto de diversas instancias de formación y refinamiento didáctico es el trabajo del Banco Mundial en [La evaluación de impacto en la práctica](#) (particularmente la "*Segunda parte: cómo evaluar*").

Además de estos enfoques generalmente más desarrollados, el cuadro anterior presenta tres enfoques de diseño que no están actualmente ampliamente implementados en la evaluación de impacto, pero que a la vez ofrecen un potencial considerable para vincular las intervenciones de desarrollo con resultados e impactos. Dada su usual ausencia en la formación sobre evaluación de impacto, y reconociendo sus méritos, capacidades y promesa de reforzar y fortalecer la práctica actual de evaluación de impacto, les dedicaremos a continuación una breve referencia.

Enfoques de Diseño, Variantes e Inferencia Causal		
Enfoques de Diseño	Variantes Específicas	Bases para la inferencia causal
Experimental	RCTs Cuasi-experimentos, Experimentos naturales	Contrafactuales; la co-presencia de causa y efectos
Estadístico	Modelización estadística Estudios longitudinales Econometría	Correlación entre causa y efecto o entre variables, influencia de causas múltiples (generalmente) aislables en un solo efecto Control de factores de confusión (<i>cofounders</i>)
Basado en Teoría	<i>Diseños causales de procesos</i> : Teoría del Cambio, Seguimiento de procesos (<i>Process tracing</i>), análisis de contribución, análisis de vías de impacto (<i>impact pathways</i>) <i>Diseños causales</i> : Evaluación Realística (<i>realist evaluation</i>), análisis de congruencia	Identificación / confirmación de procesos causales o 'cadenas' Factores y mecanismos de apoyo que actúan en el contexto
Enfoque basado en casos	<i>Interpretativo</i> : Naturalístico, Teoría Fundamentada (<i>Grounded Theory</i>), Etnografía <i>Estructurado</i> : Configuraciones, Análisis Cualitativo Comparado, Análisis al interior del caso, Simulación	Comparación de combinaciones de factores causales entre casos y al interior de los mismos Generalización analítica basada en la teoría
Participativo	<i>Diseños Normativos</i> : Evaluación participativa o democrática, Evaluación de empoderamiento <i>Diseños de mediación</i> : Aprender haciendo Diálogo político, Investigación-acción colaborativa	Validación por los participantes que sus acciones y los efectos experimentados fueron "causados" por el programa La adopción, adaptación y compromiso con un objetivo
Estudios Síntesis	Meta-análisis, Síntesis Narrativa, " <i>Realist based synthesis</i> "	Acumulación y agregación al interior de un número de perspectivas (estadísticas, basadas en teoría, etnográfico, etc.)

Enfoques basados en la teoría

Es interesante notar cómo Howard White (2009),¹⁵ a cargo de la iniciativa 3Ie -generalmente asociada con los acercamientos más experimentales hacia la evaluación de impacto-, aboga por una mayor aplicación de un enfoque basado en la teoría como un medio para mejorar la relevancia política de las evaluaciones de impacto.

En la investigación social se arriba a explicaciones contando con teoría que permita salvar la brecha entre los datos y su interpretación. Aplicando esa idea a la evaluación de impacto, la teoría tiene un rol en cerrar la brecha entre "causas" y "efectos". El término "teoría" se utiliza en evaluación de diversas maneras e inconsistentemente, sin embargo, podemos identificar dos tendencias principales en los enfoques basados en la teoría que tienen relevancia para los fines de la evaluación de impacto

¹⁵ [White, H. \(2009b\) Theory-based impact evaluation: Principles and practice, Working Paper 3, International Initiative for Impact Evaluation, New Delhi](#)

(si bien en la práctica están inextricablemente entretejadas y la mayoría de las practicantes las combinan).

La primera tendencia en los enfoques basados en teoría es la que enfatiza la orientación hacia los procesos. El programa o proyecto es entendido como un "conjunto" de causas que sigue una secuencia. Sigue la trayectoria de un programa desde su inicio a través de varios vínculos causales encadenados, hasta que los resultados deseados se alcanzan. El proceso se basa en una "teoría" -un conjunto de supuestos acerca de cómo una intervención alcanza sus objetivos y en qué condiciones. Ésta fue la lógica fundacional con que la *Aspen Round Table* y la *Fundación Kellogg* comenzaron usando originalmente las "Teorías del Cambio", y continúa dando forma a diversas representaciones de 'causalidad' y trayectorias" de impacto".

Hay representaciones débiles y fuertes de la teoría de un programa. En las formas débiles llegan a ser poco más que un modelo lógico que expresa las intenciones de los responsables políticos, ignorando las acciones e intenciones de otros actores o partes interesadas; y donde no se hacen explícitos los supuestos que se tienen acerca de las condiciones en que se espera el éxito del programa. Las representaciones más fuertes tienen en cuenta estos elementos, asumiendo que la evaluación -y la evaluación de impacto en particular- es una oportunidad para poner a prueba la teoría de un programa a través de los vínculos de la cadena causal. En términos de método, esta tendencia está cerca del llamado "*process tracing*"¹⁶, el que es definido como: "narrativas teóricamente explícitas que cuidadosamente rastrean y comparan las secuencias de acontecimientos que constituyen el proceso ...".¹⁷ Estas cadenas causales se representan gráficamente como mapas causales o redes neuronales.

La segunda tendencia en los enfoques basados en teoría enfatiza los entendimientos centrados en el mecanismo. Se basan en la generalizada suposición de la filosofía de la ciencia que con el fin de poder hacer afirmaciones causales plausibles es necesario que haya una identificación de un mecanismo que hace que las cosas sucedan. Los mecanismos de un programa "dan el paso desde preguntar si un programa funciona hacia el entender qué hay en el programa que hace que funcione".¹⁸ Una intervención en educación asegura que los niños sean capaces de permanecer en la escuela después de los once años. El mecanismo que asegura que esto conduzca a resultados positivos en materia de educación es la interacción entre un niño y un maestro en la escuela. Sin embargo, hay otras condiciones entendidas como "contextuales" o "condiciones favorecedoras" que deben estar presentes. Estos incluyen: a) la disponibilidad de un profesor, b) la formación que el profesor ha recibido, y c) la construcción previa y equipamiento de un aula para albergar tanto al alumno y al maestro. Por lo tanto, los mecanismos no son "libres de contexto" para asegurar su funcionamiento -son contingentes y dependen de diferentes condiciones de partida, factores de apoyo, y predisposiciones. Los mecanismos causales operan en contextos específicos, por lo que es importante analizar los mismo. Esto a menudo se hace sólo de manera *ad hoc*, pero las tipologías de contexto son un paso intermedio útil



¹⁶ George AL and McKeown TJ (1985) "Case Studies and Theories of Organizational Decision Making." *Advances in Information Processing in Organizations* 2: 21-58

Collier D (2011) *Understanding Process Tracing*. *Political Science and Politics* 44 No.4 (823-830)

¹⁷ Aminzade, R (1993) *Class Analysis, Politics, and French Labor History in Rethinking Labor History*, Berlanstein L (ed). Urbana and Chicago: University of Illinois Press, 90-113.

¹⁸ Programme mechanisms 'take the step from asking whether a programme works to understanding what it is about the programme that makes it work'. Pawson R and Tilley N (1997). *Realistic Evaluation*. Thousand Oaks and London: Sage. Pag.6

para la generalización de las evaluaciones basadas en mecanismos. Los contextos pueden incluir programas relacionados que afectan a la misma población objetivo, los factores socio-económicos y culturales; y factores históricos, como iniciativas de desarrollo anteriores. Desarrollar tipologías de contexto, si bien no sostienen generalizaciones universales, pueden apoyar generalizaciones del tipo "bajo estas condiciones".

Estudios del 'caso'

La elección de esta denominación está destinada a llevarnos más allá de la comprensión tradicional de los "estudios de caso". A partir de nuevas metodologías para el análisis sistemático causal de "casos" surgidas en el cambio de siglo, los casos pueden ser tanto intervenciones políticas, instituciones, personas, eventos o incluso países durante un determinado período histórico. Desde esta perspectiva, la noción de "casos" conlleva generalmente el entendimiento de los mismos como sistemas "complejos".

Este énfasis representa un cambio desde la tradición de centrar el análisis causal en las variables extraídas de su contexto específico. La localización de las variables en el contexto del "caso", y el llevar a cabo análisis hacia el interior de los casos junto a las comparaciones entre casos, ha abierto nuevas e importantes oportunidades para el análisis causal que en gran parte están siendo todavía ignoradas en la práctica de la evaluación.

Existe una tradición en evaluación de estudios de casos "naturalistas", "constructivistas" e "interpretativos" que por lo general se centran en las características únicas de un solo caso. Estos estudios de casos evitan el análisis causal a pesar de que contribuyen a este tipo de análisis de varias maneras. Por ejemplo, los estudios de casos interpretativos ofrecen una rica comprensión de los contextos en las evaluaciones basadas en teoría; ayudan a definir la validez de la construcción en términos que tengan sentido para los actores sobre el terreno; y dan voz a los beneficiarios del programa tanto en la etapa en que se formulan las preguntas de evaluación como cuando se realizan las interpretaciones de hallazgos.

Los nuevos enfoques para los estudios del "caso" están especialmente interesados en el análisis causal,¹⁹ así como también en la generalización que va más allá del caso individual -aunque no en la 'universalización'.²⁰ Estos enfoques tienden a generalizar "bajo ciertas condiciones" e identificar grupos o subconjuntos de casos sobre los que es posible hacer similares inferencias causales. La inferencia causal basada en el caso puede "probar" (*test*) una teoría (como en la noción de George y Bennett de "congruencia"), así como contribuir a la "construcción de teoría". Sin embargo, el papel de la teoría es menos pronunciado dentro de diseños de estudio del "caso" que tienen como objetivo el análisis causal. Por ejemplo, en el *análisis comparativo cualitativo* (QCA) la inferencia causal depende más de las condiciones que son necesarias y suficientes para lograr un resultado, basando éste en comparaciones de "configuraciones" de casos y sus atributos.

El caso único también se presenta en los enfoques basados en el caso, sobre todo a través de análisis "al interior de los casos" –esto es, el examen de los atributos de un caso a fin de establecer configuraciones. Teóricamente debería ser posible llevar a cabo estudios únicos de "caso" en este marco, ubicando el caso en relación con un conjunto previamente establecido de otros programas o instituciones similares.

¹⁹ George AL and Bennett A (2005) *Case Studies and Theory Development in the Social Sciences*. Cambridge Mass. And London: MIT Press.

Ragin CC (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago and London: University of Chicago Press.

Byrne D and Ragin C (eds) (2009). *The SAGE Handbook of Case-Based Methods*, London: Sage.

²⁰ Byrne D (2009) *Case-Based Methods; Why We Need Them; What They Are; How To Do Them*. In Byrne D and Ragin C (eds) *The SAGE handbook of Case-Based Methods*, London: Sage.

Enfoques participativos

Las políticas de desarrollo se basan en gran medida en ideas tales como participación, apropiación (*ownership*) y la rendición de cuentas democrática, del mismo modo que las principales áreas de los programas se ocupan también de las ideas relacionadas con la "democracia", "transparencia", y "empoderamiento". Los "Principios de la Declaración de París" tales como "apropiación" y "responsabilidad mutua" tienen una resonancia muy fuerte con la idea de participación, al igual que la filosofía general de "partenariado" (*partnership*) o asociación entre los donantes y los "países socios" o receptores. La evaluación en contextos de desarrollo ha expresado valores similares a través de métodos tales como el Diagnóstico Participativo Rápido, la Investigación Acción Participativa, el Cambio Más Significativo, etc. Estos enfoques se relacionan con la evaluación de impacto, aunque sólo sea indirectamente. Por ejemplo, ellos pueden:

- Asegurarse de que los beneficiarios y los pobres tengan voz cuando se planifican los programas, mejorando la focalización y pertinencia.
- Investigar las comunidades locales y sus circunstancias, sus problemas y limitaciones para mejorar así la "validez de constructo".
- Agregar la perspectiva de beneficiarios y partes interesadas a las conclusiones y las lecciones aprendidas de la evaluación de impacto.

Para ser visto como un diseño que contribuye más directamente a la evaluación de impacto, los enfoques participativos deben apoyar la inferencia causal. La lectura de la Declaración de París y los más recientes aportes de la Conferencia de Busan sobre las formas de generar sinergias en la ayuda al desarrollo, sugieren que existe una base relevante para el uso de las lógicas de participación de este modo. Por ejemplo, la Declaración de París señaló que, si los países en desarrollo se apropian de sus planes de desarrollo, es más probable que los mismos se cumplan. Esta perspectiva también es consistente con la noción de que el desarrollo es posible sólo a través de acciones auto-dirigidas, y que el rol de aquellos que apoyan esos procesos es mejorar dicha autonomía. Los enfoques participativos vinculados a la inferencia causal no ven a los destinatarios de la ayuda como receptores pasivos, sino como "agentes" activos. En este entendimiento, los beneficiarios tienen "agencia" y pueden ayudar a "causar" resultados exitosos por sus propias acciones y decisiones. Como se sugirió anteriormente, éste sería el caso si los tomadores de decisiones nacionales participaran activamente en los programas y su evaluación.

Debería ser posible interpretar los impactos en términos del contenido participativo de una intervención: por ejemplo, el grado en que la participación, el apropiamiento y el compromiso mejoran los resultados de desarrollo. Estos procesos se relacionan tanto con el diseño del programa, así como con el diseño de evaluación –lo que es también el caso de las inferencias causales que utilizan otros diseños. Las "evaluaciones basadas en la Teoría", por ejemplo, miran a los aspectos causales de los programas: los diseños participativos se centran en la "agencia" de las partes interesadas (*stakeholders*) y en los métodos participativos para demostrar que el mismo ha marcado una diferencia.

8. A MODO DE CIERRE

La evaluación de impacto precisa de ciertos pasos generales para su desarrollo. En su diseño debe contemplar principalmente los objetivos que procura alcanzar, de los cuales se desprenden las preguntas de evaluación que espera responder, la identificación de los usuarios, y el abordaje de ciertos componentes clave de la evaluación de impacto. El diálogo y revisión de dicho diseño (contemplando incluso a los usuarios identificados anteriormente) acrecienta las posibilidades tanto de lograr una evaluación de impacto de calidad, como también su subsiguiente utilización.

Uno de los elementos que han ido ganando terreno en las discusiones sobre la evaluación de impacto apunta a dar cuenta cómo muchas de las intervenciones de desarrollo se caracterizan por ser complejas, multi-dimensionales, involucrando a una pluralidad de actores, efectivizadas por

diversas vías, y con propósitos de largo plazo y sustentabilidad. Esto tiene implicancias en cuanto a la limitada capacidad de establecer vínculos entre ellos como "causas" y los resultados de desarrollo como "impactos". Hay una serie de iniciativas más recientes en las cuales el trabajo sobre su evaluación es incipiente y exploratorio: programas sobre gobernanza, empoderamiento, transparencia y rendición de cuentas, cambio climático, etc. Ante estas realidades (novedosas y complejas) es conveniente moderar las expectativas y asumir visiones más realistas respecto a la factibilidad de estimaciones precisas de los impactos a largo plazo. En algunas circunstancias, otro tipo de evaluación (antes que la evaluación de impacto) será la factible y pertinente para aportar al aprendizaje y la mejora de las intervenciones de desarrollo.



Diseños y métodos para la evaluación del impacto



EvalParticipativa es la Comunidad de Práctica y Aprendizaje en Evaluación Participativa para América Latina y el Caribe. Es una iniciativa conjunta entre el Programa de Estudios del Trabajo, el Ambiente y la Sociedad (PETAS) de la Universidad Nacional de San Juan (Argentina) y el proyecto Focelac+ para el fomento de una cultura de evaluación y aprendizaje en América Latina con proyección global, del Instituto Alemán de Evaluación para la Cooperación al Desarrollo (DEval).

Esta comunidad nace del interés y crecimiento de la evaluación en el contexto regional y global. En las últimas décadas, se ha incrementado la producción teórica y metodológica, como también el surgimiento de políticas nacionales de evaluación en países de todos los continentes, una creciente institucionalidad de la evaluación y la consolidación de diversas iniciativas orientadas a profesionalizar esta práctica. Junto a ello, los Objetivos de Desarrollo Sostenible dan cuenta de una nueva agenda de prioridades para la evaluación a nivel global, lo que incluye el protagonismo de la sociedad civil en dichos esfuerzos y destaca la dimensión participativa como aspecto central de la práctica evaluativa. Es así como los conceptos de "participación", "acompañamiento", "perspectiva de los actores" y otros similares se convirtieron en recurrentes en el actual quehacer del campo evaluativo.

EvalParticipativa busca potenciar el involucramiento inclusivo de la sociedad civil en procesos evaluativos. Para ello, ha definido tres objetivos estratégicos principales, a saber: (a) fortalecer y consolidar la comunidad de práctica y aprendizaje en torno a la evaluación participativa en América Latina y el Caribe; (b) facilitar la multiplicación e institucionalización de este enfoque de evaluación, mediante el desarrollo de materiales escritos y audiovisuales sobre experiencias y lecciones aprendidas en torno a la evaluación participativa; y (c) impulsar procesos de formación sobre evaluación participativa, a través de cursos para facilitadores/as, seminarios y foros presenciales y virtuales con representantes de organizaciones de la sociedad civil de la región, el sector público y privado, y la academia, así como la generación de instancias de certificación académica de la formación en evaluación participativa.

La comunidad ya cuenta con más de tres mil seguidores a través de sus entornos virtuales. Además, se articulan acciones con la Red de Seguimiento, Evaluación y Sistematización de Latinoamérica y el Caribe (ReLAC), la organización TECHO, la Fundación para la Superación de la Pobreza (Chile) y el programa Servicio País, y el Centro de Apoyo al Desarrollo Local (CEA-DEL) de Argentina.

www.evalparticipativa.net

ISBN 978-631-01-0869-8



9 786310 108698