

INTRODUCCIÓN A LA EVALUACIÓN DE IMPACTO

Por Patricia J. Rogers
Universidad RMIT (Australia) y BetterEvaluation



Esta es la primera guía de orientación de una serie de cuatro partes de guías relacionadas con la evaluación de impacto desarrolladas por InterAction y con el apoyo financiero de la Fundación Rockefeller. Las otras guías de esta serie son: *Relación entre el Monitoreo y la Evaluación con la Evaluación de Impacto*; *Introducción a los Métodos Mixtos de la Evaluación de Impacto*; y *Uso de los Resultados de la Evaluación de Impacto*. Puede encontrar la serie completa en el sitio web de InterAction: <http://www.interaction.org/impact-evaluation-notes>.

AGRADECIMIENTOS

Han beneficiado a esta guía de orientación los comentarios proporcionados por los miembros del Grupo de Trabajo de Efectividad del Programa y de la Evaluación de InterAction. La información sobre los métodos de evaluación que se encuentra en la Guía de orientación 1 ha sido extraída de BetterEvaluation, una colaboración internacional para la puesta en común de información con el fin de mejorar la evaluación. Puede obtener más información acerca de cada uno de los métodos en la página web www.betterevaluation.org.

Foto: David Darg

Contenidos

Introducción	1
1. ¿Qué se entiende por “evaluación de impacto”?	2
2. ¿Por qué debemos realizar una evaluación de impacto?	3
3. ¿Qué preguntas intenta contestar la evaluación de impacto?	3
4. ¿Quién debe conducir la evaluación de impacto?	3
5. ¿Cómo deberíamos elegir los métodos para la evaluación de impacto?	5
6. Aclaración de los valores para una evaluación de impacto	6
7. Desarrollo de una teoría o modelo de cómo se supone que funcione la intervención	7
8. Medición o descripción de los impactos (y otras variables importantes)	8
9. Explicación del alcance que obtuvo la intervención al producir los resultados observados	10
10. Síntesis de la evidencia	14
11. Informe de los hallazgos y respaldo del uso	14
12. ¿Cuándo se debe realizar una evaluación de impacto?	15
13. ¿Qué se necesita para la evaluación de impacto de calidad?	15
14. Desafíos comunes en la evaluación de impacto en desarrollo	17
Resumen	19
Referencias y otros recursos útiles	19

INTRODUCCIÓN

Una evaluación creíble y adecuada de impacto puede mejorar considerablemente la eficacia del desarrollo. El énfasis cada vez mayor que se pone en la evaluación de impacto en desarrollo ha generado muchas preguntas. ¿Qué constituye una evaluación creíble y adecuada de impacto? ¿Cómo se deben manejar las evaluaciones de impacto? ¿Qué medidas y fuentes de datos son adecuadas? ¿Cómo se pueden combinar eficazmente los datos cualitativos y cuantitativos en la evaluación de impacto? ¿Qué se debe hacer para respaldar el uso adecuado de las evaluaciones de impacto? ¿Cuáles son las repercusiones del foco cada vez mayor sobre la evaluación de impacto para otros tipos de monitoreo y evaluación (M&E)?

InterAction ha creado una serie de guías de orientación que abordan estas preguntas para respaldar al personal de administración, del programa y de M&E en ONG internacionales para planificar, diseñar, administrar, conducir y utilizar evaluaciones de impacto. Estas guías también pueden proporcionar información sobre sus debates con evaluadores externos, socios y donantes.

La primera guía de orientación, *Introducción a la Evaluación de Impacto*, brinda una visión general de la evaluación de impacto, explica cómo se diferencia de (y complementa a) otros tipos de evaluaciones, por qué se debe realizar la evaluación de impacto, cuándo y por quién. En ella se describen

diferentes métodos, enfoques y diseños que se pueden utilizar para los distintos aspectos de la evaluación de impacto: aclaración de los valores para la evaluación, desarrollo de una teoría sobre cómo se entiende que funciona la intervención, 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. La guía desarrolla qué se considera una buena evaluación de impacto (la evaluación que alcanza un equilibrio entre los imperativos competitivos de ser útil, riguroso, ético y práctico) y cómo lograrla. Las notas a pie de página que se encuentran a lo largo del documento contienen referencias para obtener material de lectura adicional sobre áreas específicas.

1. ¿Qué se entiende por “evaluación de impacto”?

La evaluación de impacto investiga los cambios que provoca una intervención. Se puede emprender una evaluación de impacto en intervenciones a cualquier escala: un proyecto pequeño y local sobre VIH/SIDA; un programa completo de fortalecimiento de sociedad civil de una ONG; 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.

Los resultados esperados de una intervención son una parte importante de una evaluación de impacto, pero también es importante investigar los resultados inesperados. En esta guía de orientación, los impactos se definen como:

los efectos positivos y negativos, deseados y no deseados, directos e indirectos, primarios y secundarios, que produce una intervención. (Definición del Comité de Asistencia para el Desarrollo de la OECD)¹

Normalmente, se entiende que los impactos ocurren después, y como consecuencia, de resultados intermedios. Por ejemplo, lograr los resultados intermedios de un acceso mejorado a la tierra y mayores niveles de participación en la toma de decisiones de la comunidad puede ocurrir antes

¹ A veces, los impactos se definen de manera muy diferente. Por ejemplo, la guía de la Fundación W.K.Kellogg “Logic Model Development Guide” (www.wkkf.org/knowledge-center/resources/2006/02/WK-Kellogg-Foundation-Logic-Model-Development-Guide.aspx) distingue al impacto en lo que respecta a su difusión más allá de aquellos involucrados inmediatamente en el programa. Los cambios específicos en el comportamiento, el conocimiento, las habilidades, el estado y el nivel de funcionamiento de los participantes del programa se denominan «resultados», y solo se describen como «impactos» los cambios en las organizaciones, comunidades o sistemas como resultado de las actividades del programa dentro de siete a 10 años.

del, y contribuir al, impacto final deseado de un mejoramiento en la salud y el bienestar de las mujeres. La diferencia entre resultados e impactos puede ser relativa, y depende de los objetivos establecidos de una intervención.

En la práctica, a veces es útil que una evaluación incluya tanto los resultados como los impactos. Esto permite indicar de manera pronta si una intervención está funcionando o no; y si no está funcionando, ayuda a identificar dónde y, quizás, por qué.

En esta guía de orientación, una evaluación de impacto incluye cualquier evaluación que investigue de manera sistemática y empírica los impactos que una intervención produce. Algunas personas y organizaciones utilizan una definición más reducida de la evaluación de impacto, y solo incluyen evaluaciones que contienen un contrafactual de algún tipo (un cálculo de lo que habría pasado si la intervención no hubiese ocurrido) o una clase particular de contrafactual (por ejemplo, comparaciones con un grupo que no recibió la intervención). USAID, por ejemplo, utiliza la siguiente definición: “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”. Estas diferentes definiciones son importantes cuando se deciden los métodos o diseños de investigación que los usuarios potenciales de la evaluación o los socios o donantes considerarán creíbles.

La evaluación de impacto no es, por supuesto, el único tipo de evaluación que respalda el desarrollo eficaz. Es importante asegurarse de que

las inversiones en la evaluación de impacto (en términos de tiempo y dinero) no se realicen a costa del monitoreo o de otros tipos de evaluación (como la evaluación de necesidades, la evaluación del proceso y la evaluación de costo-beneficio) que también se precisan para informar las decisiones acerca de la práctica y la política. La Guía de orientación 2 desarrolla cómo se pueden realizar la evaluación de impacto y estos otros tipos de monitoreo y evaluación de manera que se respalden mutuamente. Por ejemplo, datos de monitoreo pueden proporcionar una buena base para la evaluación de impacto y una evaluación de impacto puede guiar el desarrollo de los sistemas de monitoreo. La evaluación de impacto proporciona la información necesaria para las evaluaciones de costo-beneficio y costo-eficacia.

2. ¿Por qué debemos realizar una evaluación de impacto?

La mejor manera de emprender una evaluación de impacto en particular depende, en parte, de su objetivo y de quiénes son sus usuarios potenciales primarios. Algunas razones comunes para realizar una evaluación de impacto incluyen:

- **Decidir si financiar una intervención o no:** una “evaluación ex-ante” se lleva a cabo *antes de que se implemente una intervención*, para calcular sus posibles impactos e informar decisiones de financiación.
- **Decidir si continuar o ampliar una intervención o no.**
- **Aprender cómo replicar o aumentar un piloto.**
- **Aprender cómo adaptar satisfactoriamente una intervención exitosa para otro contexto.**
- **Asegurarles a los financiadores, incluyendo los donantes y los contribuyentes (responsabilidad ascendente), que el dinero se está invirtiendo con prudencia:** lo que incluye

que la organización está aprendiendo lo que funciona y lo que no, y está utilizando esta información para mejorar las futuras decisiones de inversión e implementación.

- **Informarles a las comunidades y beneficiarios propuestos (responsabilidad descendente) sobre si un programa está beneficiando a la comunidad o no y en caso de así serlo, en qué maneras lo está haciendo.**

La Guía de orientación 4 desarrolla con más detalle cómo respaldar estas diferentes maneras de utilizar la evaluación de impacto.

3. ¿Qué preguntas intenta contestar la evaluación de impacto?

Una evaluación de impacto se debe concentrar en una pequeña cantidad (cinco a siete) de preguntas clave específicas de evaluación. Estas son las preguntas de alto nivel que una evaluación aborda; no son preguntas específicas que se pueden realizar en una entrevista o en un cuestionario. Es mejor concentrarse en un pequeño número de preguntas que estén directamente relacionadas al objetivo en lugar de distribuir los recursos de la evaluación, y la atención de los usuarios, entre una gran cantidad de preguntas. (vea los ejemplos de preguntas clave de evaluación para la evaluación de impacto en la página siguiente).

4. ¿Quién debe conducir la evaluación de impacto?

La evaluación de impacto puede ser llevada a cabo por: un evaluador o un equipo de evaluación externo; una unidad interna pero separada de la organización de implementación; aquellos involucrados en una intervención (incluyendo los miembros de la comunidad); o un equipo combinado de evaluadores internos y externos.

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?

Un evaluador externo puede proporcionar un rango de especialización y experiencia que posiblemente no esté disponible dentro de la organización, y puede tener mayor independencia y credibilidad que un evaluador interno. Por ejemplo, la Política de Evaluación de USAID exige que la mayoría de las evaluaciones sean realizadas por un evaluador externo.

Sin embargo, para algunas partes interesadas, los evaluadores externos no siempre son considerados imparciales, ya que su recolección de datos e interpretaciones pueden ser afectadas por la falta de familiaridad con el contexto. En algunos casos, involucrar a las partes interesadas del programa y/o los miembros de la comunidad en la realización de una evaluación puede agregar rigor

y credibilidad al respaldar un mejor acceso a los datos (especialmente informantes clave) y una interpretación más adecuada de los datos.

Tres prácticas en particular pueden, a menudo, producir la evaluación de mejor calidad: establecer un equipo de evaluadores con perspectivas internas y externas, asegurar la transparencia en respecto al tipo de datos que se utiliza en la evaluación y cómo se utiliza, y aplicar la triangulación: el uso de diferentes fuentes de evidencia (que tengan fuerzas complementarias) y varias perspectivas en el análisis y en la interpretación. Es particularmente útil incluir en el equipo expertos locales en evaluación que conozcan el contexto, la historia y las intervenciones comparativas de otras agencias.

Una evaluación puede ser administrada por un grupo interno (quizás un comité directivo interno, orientado por un grupo asesor con membresía externa) o por un grupo combinado. Los enfoques participativos para administrar evaluaciones normalmente incluyen al personal del programa, miembros de la comunidad y socios de desarrollo. Participan no solo en la recopilación de datos, sino también en la negociación del objetivo de la evaluación de impacto, desarrollando las preguntas clave de la evaluación, diseñando una evaluación para contestarlas y llevando a cabo los resultados.²

5. ¿Cómo deberíamos elegir los métodos para la evaluación de impacto?

Ha habido un debate importante sobre la evaluación de desarrollo y, más ampliamente, acerca de los métodos que son mejores para la evaluación de impacto. Estos debates reflejan diferentes vistas sobre qué constituye evidencia creíble, rigurosa y útil, y quién debe estar involucrado en la realización y control de las evaluaciones.

Algunas organizaciones y evaluadores han sostenido que se deben utilizar métodos o diseños de investigaciones particulares cuando sea posible; por ejemplo, pruebas aleatorias controladas o métodos participativos. Otros han defendido la adecuación situacional. Esto significa elegir métodos que se adapten al objetivo de la evaluación,

² Algunas fuentes adicionales sobre métodos participativos pueden ser: Marisol Estrella et al. (eds), *Learning from Change: Issues and Experiences in Participatory Monitoring and Evaluation* (Brighton: Instituto de Estudios de Desarrollo, 2000), <http://www.idrc.ca/EN/Resources/Publications/Pages/IDRCBookDetails.aspx?PublicationID=348>; Andrew Catley et al., "Participatory Impact Assessment," Centro Internacional Feinstein, Universidad Tufts: octubre de 2008, <http://sites.tufts.edu/feinstein/2008/participatory-impact-assessment>; Robert Chambers, "Who Counts? The Quiet Revolution of Participation and Numbers," Documento de trabajo n.º 296 (diciembre de 2007), Brighton: Instituto de Estudios de Desarrollo, <http://www.ids.ac.uk/files/Wp296.pdf>.

los tipos de preguntas de evaluación que se realizan, la disponibilidad de recursos y la naturaleza de la intervención; particularmente si es estandarizada o adaptable, y si las intervenciones funcionan de la misma manera en todos lados y para todos o si se ven afectadas en gran medida por el contexto.

Cuando se eligen los métodos, es importante abordar cada uno de los seis diferentes aspectos de una evaluación de impacto:

- **Aclarar los valores** que respaldarán la evaluación: ¿qué se considerarán procesos, impactos y distribución de costos y beneficios deseables e indeseables?
- **Desarrollar y/o probar una teoría** sobre cómo se supone que funciona la intervención: a veces, se las denomina teorías de cambio, modelos lógicos o teorías de programa.
- **Medir o describir** estos impactos y otras variables importantes, incluyendo los procesos y el contexto.
- **Explicar** si la intervención fue la causa de los impactos observados.
- **Sintetizar** evidencia en un juicio evaluativo general.
- **Informar hallazgos y respaldar su uso.**

Esta guía de orientación desarrolla cada uno de estos aspectos y brinda información acerca de un rango de métodos que se pueden utilizar para ellos. Se proporcionan vínculos a fuentes adicionales de información. La Guía de orientación 3 desarrolla la manera en que el enfoque de métodos mixtos, que combina datos cuantitativos y cualitativos de maneras complementarias, puede ser tanto la medida/descripción como la explicación.

6. Aclaración de los valores para una evaluación de impacto

El primer paso es aclarar los valores que respaldarán la evaluación. 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 en respecto a:

- **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? ¿Qué nivel no sería un intercambio aceptable?
- **Lograr una distribución deseable de los beneficios.** Por ejemplo, ¿deberíamos juzgar el éxito en términos del resultado educacional promedio, mejoras para los más desfavorecidos o colocar un grupo vulnerable o desfavorecido (como niñas) en el mismo nivel que sus contrapartes más favorecidas?

Los objetivos formales establecidos (incluyendo los Objetivos de Desarrollo del Milenio) y las políticas organizacionales son un inicio importante para aclarar los valores, pero normalmente no son suficientes por sí mismos. Diferentes partes interesadas pueden tener diferentes puntos de vista acerca de los valores que se deben utilizar en una evaluación.

Algunos métodos para aclarar los valores para una evaluación de impacto son:

Métodos que ayudan a que las personas expresen valores tácitos

Indagación apreciativa: las partes interesadas clave (incluyendo el personal del programa) recuerdan momentos en los que el programa funcionó particularmente bien, luego identifican los valores que se demostraban durante dichos momentos.³

Encuestas en la comunidad: Los individuos de la comunidad designan o califican los problemas que consideran más importantes para tratar.

Cambio más significativo: Un proceso estructurado para generar y seleccionar historias de cambio que identifiquen qué ven las diferentes personas y grupos como los resultados o impactos más importantes.⁴

Métodos que ayudan a negociar entre diferentes conjuntos de valores

Delphi: proceso que funciona a través de una serie de interacciones por escrito sin contexto cara a cara, donde las partes interesadas clave proporcionan sus opiniones acerca de lo que consideran importante y luego responden a los resultados agregados.⁵

Votación de punto adhesivo: en una reunión cara a cara, los individuos

³ Material de lectura adicional: Hallie Preskill, "Using Appreciative Inquiry in Evaluation Practice," <http://www.eval.org/summerinstitute07/handouts/sio7.preskillf.pdf>.

⁴ Material de lectura adicional: Most Significant Change (MSC), <http://mande.co.uk/special-issues/most-significant-change-msc/>.

⁵ Material de lectura adicional: Delphi Method, http://www.unido.org/fileadmin/import/16959_DelphiMethod.pdf; The Delphi Method: Techniques and Applications, <http://is.njit.edu/pubs/delphibook/>; Delphi Survey – Europa, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/downloads/sb2_delphi_survey.doc.

distribuyen sus distintos “votos” (en forma de puntos adhesivos) en todas las opciones.⁶

7. Desarrollo de una teoría o modelo de cómo se supone que funcione la intervención

A menudo, es útil basar una evaluación de impacto en una teoría o en un modelo de cómo se entiende que la intervención produce sus impactos deseados. Esto se podría llamar una teoría de programa, una teoría de cambio (ToC), una cadena de resultados o un modelo lógico. Es mejor desarrollar la teoría de cambio como parte de la planificación de una intervención, y luego examinarla y revisarla según sea necesario mientras se planifica una evaluación de impacto. Si esto no se ha realizado antes del inicio de la intervención, es posible desarrollar retroactivamente una teoría de cambio acordada.

Según cuándo se desarrolla la teoría de cambio, puede basarse en una combinación de fuentes: documentos oficiales y objetivos establecidos; investigación de intervenciones similares; observaciones de la intervención o intervenciones similares; o preguntarle a las diferentes partes interesadas (incluyendo los planificadores, el personal y los beneficiarios deseados) cómo piensan que funciona (o debería funcionar).

Puede haber diversas teorías de cambio: diferentes teorías que muestran cómo funciona la intervención en las diferentes etapas, en diferentes contextos (reconociendo efectos de influencias externas) y para diferentes impactos, y diferentes teorías que se desarrollan con el tiempo mientras se desarrolla un mejor conocimiento.

⁶ Material de lectura adicional: NRCOI Quick Tip, <http://centerforchildwelfare.fmhi.usf.edu/kb/bpam/NRCOI-StickyDotTechnique.pdf>.

Las teorías de cambio pueden mejorar la evaluación de impacto al ayudar a:

- 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.
- Identificar, si una intervención no fue exitosa, en qué partes del proceso dejó de funcionar o fracasó.
- 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).
- Identificar los aspectos de la intervención que la hizo funcionar y son, por lo tanto, críticos y necesitan ser continuados cuando se adapta una intervención para otros entornos.
- 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.
- Proporcionar un marco conceptual para reunir diferentes evidencias sobre un programa que involucra un gran número de diversas intervenciones.

Algunos métodos para representar una teoría de cambio son:

Enfoque de marco lógico (marco lógico): el formato clásico utilizado en diversas organizaciones de desarrollo, que utiliza una matriz de 4x4. Las cuatro filas son *actividades*, *productos*, *propósito* y *objetivo general*, y las cuatro columnas son *una descripción narrativa*,

indicadores objetivamente identificables (OVI), medios de verificación (MoV) y suposiciones.⁷

Cadena de resultados: la intervención se representa como una serie de cuadros en una secuencia de insumos, actividades, productos, resultados a corto plazo, resultados a largo plazo e impactos.⁸

Cadena de resultados/jerarquía de resultados/teoría de cambio: la teoría está representada como una serie de resultados intermedios que conducen a los impactos deseados finales. Este formato centra la atención en cómo se produce el cambio, y es útil para representar programas donde ocurren diferentes actividades a lo largo de la cadena causal, no solo por anticipado.⁹

Mapeo de resultados: se centra en identificar los “socios de límites” (organizaciones o grupos cuyas acciones están más allá del control de la intervención, pero que son esenciales para que se logre el impacto) y, luego, en enunciar aquello que necesitan realizar estos socios y cómo la intervención puede buscar influenciarlos.¹⁰

Puede encontrar otros recursos útiles para el desarrollo de una teoría de cambio en *Desarrollar un Modelo Lógico o Teoría de Cambio*.¹¹

También se puede utilizar una teoría de cambio para administrar posibles impactos negativos, o para planificar una evaluación de impacto que los mida.

Por ejemplo, un programa destinado a mejorar la productividad agrícola que aliente a los agricultores a que apliquen fertilizante en sus campos puede llevar a un mayor escurrimiento de fosfato y a un daño medioambiental de las vías fluviales. Una evaluación de impacto equilibrada investigará este posible impacto además del impacto deseado de la productividad mejorada. Se puede desarrollar una teoría de cambio para examinar cómo una intervención podría producir impactos negativos. Esto se puede utilizar para adaptar la intervención con el fin de minimizar o evitar tales impactos negativos, para desarrollar indicadores de alerta temprana para propósitos de monitoreo, y para asegurarse de que se incluyan en el plan de la evaluación de impacto.

8. Medición o descripción de los impactos (y otras variables importantes)

Una evaluación de impacto necesita evidencia creíble, y no solo de los impactos. Asimismo, se necesita buena información acerca de cuán bien se ha implementado una intervención para distinguir entre una falla de implementación y una falla de teoría. También es necesaria la información acerca del contexto para comprender si una intervención funciona solo en situaciones particulares.

⁷ Material de lectura adicional: The Logical Framework Approach, <http://www.ausaid.gov.au/ausguide/pdf/ausguideline3.3.pdf>; <http://www.gdrc.org/ngo/logical-fa.pdf>; Logical Framework Analysis, <http://www.gdrc.org/ngo/logical-fa.pdf>; Beyond Logframe: Critique, Variations and Alternatives: <http://www.fasid.or.jp/shuppan/hokokusho/pdf/h21-3.pdf>.

⁸ Material de lectura adicional: Results Chain: Enhancing Program Performance with Logic Models Guide, <http://www.uwex.edu/ces/lmcourse/>; W.K. Kellogg Foundation Logic Model Development Guide, <http://www.evaluationtoolsforracialequity.org/evaluation/tool/doc/pub3669.pdf>.

⁹ Material de lectura adicional: Theories of Change and Logic Models: Telling Them Apart,

¹⁰ Material de lectura adicional: Outcome Mapping, <http://publicwebsite.idrc.ca/EN/Programs/Evaluation/Pages/IDRCBook-Details.aspx?PublicationID=121>; Outcome Mapping: ILAC Brief 7, http://www.outcomemapping.ca/download.php?file=/resource/files/csette_en_ILAC_Brief07_mapping.pdf.

¹¹ The Community Toolbox, “Developing a Logic Model or Theory of Change,” http://ctb.ku.edu/en/tablecontents/sub_section_main_1877.aspx.

Es útil identificar cualquier dato que ya esté disponible en fuentes existentes (como estadísticas oficiales, documentación del programa, Sistemas de Información Geográfica (GIS) y proyectos anteriores de investigación y de evaluación) acerca de los impactos, la implementación y el contexto. Se pueden reunir datos adicionales para completar los vacíos o mejorar la calidad de datos existentes mediante el uso de métodos como entrevistas (individuales y grupales; estructuradas, semi estructuradas o desestructuradas), cuestionarios (pueden ser cuestionarios basados en la web y recopilación de datos por medio del teléfono celular), observación (estructurada, semi estructurada o desestructurada) y medidas directas (por ejemplo, calidad del agua frente a un estándar internacional).

Las descripciones de los impactos no deben informar solo acerca del promedio, sino también acerca de la variedad de los resultados, y en particular informar sobre los patrones. Howard White habla de la importancia de mirar la heterogeneidad en su artículo del 2010:

“Es probable que un estudio que presenta una estimación de impacto único (el efecto promedio de tratamiento¹²) sea de menor uso para los formuladores de políticas que un estudio que examina en qué contexto las intervenciones son más eficaces, qué grupos meta se benefician más y qué entornos medioambientales son útiles o perjudiciales para lograr el impacto. Por lo tanto, se podría mostrar que una intervención educativa, como los rotafolios, funciona pero solo si los maestros mismos tienen un cierto nivel de educación, o solo si la escuela ya está bien equipada con materiales de lectura,

¹² El efecto promedio de tratamiento es un cálculo de la diferencia promedio que realiza una intervención. Por ejemplo, los estudiantes del programa permanecieron en la escuela un promedio de 2,5 años más (comparados con el grupo de control).

*o si los padres mismos de los estudiantes tienen educación”.*¹³

Algunas fuentes para medidas e indicadores en sectores particulares pueden ser:

Catálogo de cuestionarios de encuesta: Red Internacional de Encuestas en Hogares. Más de 2.000 cuestionarios que se pueden buscar por país, fecha y tipo de encuesta.¹⁴

Gobernanza democrática: Centro de Oslo para la Gobernanza del PNUD.¹⁵

Índice de pobreza humana: Tres indicadores que se relacionan con la supervivencia, el conocimiento y la provisión económica del PNUD.¹⁶

Objetivos de desarrollo del milenio: 48 indicadores técnicos y 18 metas para los 8 objetivos.¹⁷

Desarrollo sostenible: 130 indicadores de aspectos sociales, económicos, medioambientales e institucionales del desarrollo sostenible, propiedad de la Comisión sobre el Desarrollo Sostenible de la ONU.¹⁸

Indicadores del desarrollo mundial: El Banco mundial posee datos sobre más de 200 países en lo que respecta a más de 1000 indicadores.¹⁹

¹³ Howard White, “A Contribution to Current Debates in Impact Evaluation,” *Evaluation* (abril de 2010, vol. 16 n.º 2): 160.

¹⁴ http://www.ihsn.org/home/index.php?q=country_questionnaires

¹⁵ http://www.undp.org/oslocentre/flagship/democratic_governance_assessments.html

¹⁶ <http://hdr.undp.org/en/statistics/indices/hpi/>

¹⁷ Millennium Project: <http://www.unmillenniumproject.org/goals/gti.htm>

¹⁸ http://www.un.org/esa/sustdev/natlinfo/indicators/methodology_sheets.pdf

¹⁹ <http://data.worldbank.org/data-catalog/world-development-indicators>

La Guía de orientación 3 proporciona más detalles sobre métodos específicos para medir y describir impactos, el uso de métodos mixtos (datos cuantitativos y cualitativos utilizados de maneras complementarias) y formas de abordar desafíos en la medida y descripción de impactos.

9. Explicación del alcance que obtuvo la intervención al producir los resultados observados

Una de las características importantes de una evaluación de impacto es que no solo reúne evidencia de que han ocurrido impactos, sino que intenta comprender la función que tuvo la intervención al producir dichos impactos. Rara vez una intervención es la única causa de cambios. Normalmente, una intervención funciona en combinación con otros programas, con un contexto favorable o con otros factores. A menudo, un grupo colabora para producir un impacto conjunto, como cuando las ONG internacionales se asocian con gobiernos o comunidades locales. Por lo tanto, la “atribución causal” no se refiere normalmente a la atribución total (es decir, la intervención fue la única causa), sino a la atribución parcial o al análisis de la contribución de la intervención. A veces se las llama “posibles contribuciones”.

Por ejemplo, en la investigación agrícola, los impactos en términos de productividad mejorada pueden ser causa de una larga cadena de investigación básica y aplicada, y desarrollo del producto y comunicación. Una inversión en cualquiera de estos puede reclamar razonablemente que fue esencial en la producción de impactos, pero no podría haberlo hecho sin las otras intervenciones. En otras palabras, pudo haber sido una intervención necesaria pero no suficiente para generar el impacto en sí.

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.

Métodos posibles para examinar lo factual (en qué medida los resultados reales coincidieron con lo que se esperaba):

Estudios de caso comparativos: ¿Produjo la intervención resultados solo en los casos en que los demás elementos necesarios ocupaban su lugar?

Dosis-respuesta: ¿Hubo mejores resultados para los participantes que recibieron más de la intervención (por ejemplo, asistieron a más talleres o recibieron mayor respaldo)?

Atribución del beneficiario/experto: ¿Creyeron los participantes/informantes clave que la intervención hizo una diferencia y pudieron proporcionar una posible explicación de por qué esto fue así?

Predicciones: ¿Lograron estos participantes o sitios previstos para alcanzar los

mejores impactos (debido a la calidad de la implementación y/o contexto favorable realizarlo)? ¿Cómo se pueden explicar las anomalías?

Temporalidad: ¿Ocurrieron estos impactos en un momento coherente con la teoría de cambio; no antes de que se haya implementado la intervención?

Métodos posibles para examinar el contrafactual (un cálculo de lo que habría sucedido en ausencia de la intervención) pueden ser:

Diferencia en diferencia: La diferencia antes y después para el grupo que recibe la intervención (donde no hayan sido asignados aleatoriamente) se compara con la diferencia antes y después para aquellos que no la recibieron.²⁰

Contrafactual construido lógicamente: En algunos casos es creíble utilizar la línea de base como un cálculo del contrafactual. Por ejemplo, donde se instaló una bomba de agua, puede ser razonable medir el impacto comparando el tiempo invertido en obtener agua desde una bomba lejana, antes y después de la intervención, ya que no hay una razón creíble que indique que el tiempo invertido habría disminuido sin la intervención (White, 2007). El seguimiento de procesos puede respaldar este análisis en cada paso de la teoría de cambio.²¹

²⁰ Material de lectura adicional: Difference-in-differences, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/counterfactual_impact_evaluation/difference-in-differences/difference-in-differences_en.htm.

²¹ Material de lectura adicional: Process Tracing in Case Study Research, <http://users.polisci.wisc.edu/kritzer/teaching/ps816/ProcessTracing.htm>.

Comparaciones pareadas: Cada uno de los participantes (individuos, organizaciones o comunidades) se pareo con alguien que no es participante en variables que se consideran relevantes. Puede ser difícil coincidir adecuadamente con todos los criterios importantes.²²

Diversas líneas de base o líneas de base cambiantes: La implementación de una intervención se presenta de manera escalonada a través del tiempo y en las poblaciones de la intervención. El análisis busca en cada comunidad un patrón repetido de un cambio en el resultado medido luego de haber implementado la intervención, junto con una ausencia de fluctuaciones sustanciales en los datos en otros puntos de tiempo. Tiene un uso cada vez mayor en las intervenciones de salud a nivel de la población.²³

Puntuaciones de la propensión: Esta técnica crea estadísticamente grupos comparables según un análisis de los factores que influenciaron la propensión de las personas para participar en el programa; es particularmente útil cuando la participación es voluntaria (por ejemplo, mirar un programa de televisión con mensajes de promoción de la salud).²⁴

²² Material de lectura adicional: Techniques for Improving Constructed Matched Comparison Group Impact/Outcome Evaluation Designs, <http://knol.google.com/k/techniques-for-improving-constructed-matched-comparison-group-impact-outcome>.

²³ Material de lectura adicional: The Multiple Baseline Design for Evaluating Population-Based Research, <http://rds.epi-ucsf.org/ticr/syllabus/courses/23/2011/06/02/Lecture/readings/Hawkins.pdf>.

²⁴ Material de lectura adicional: Propensity Scores: What Do They Do, How Should I Use Them, and Why Should I Care?, <http://www.chrp.org/love/ASACleveland2003Propensity.pdf>; A Practical Guide to Propensity Score Models, <http://www.academyhealth.org/files/2009/monday/HebertP.pdf>.

Prueba aleatoria controlada (RCT): Los participantes (o comunidades u hogares) potenciales se asignan aleatoriamente para recibir la intervención o para pertenecer a un grupo de control (sin intervención o con la intervención normal) y se comparan los resultados promedio de los distintos grupos.²⁵

Regresión discontinua: En los casos en que una intervención está disponible solo para los participantes que se encuentran por encima o por debajo de un punto límite (por ejemplo, ingresos), este enfoque compara los resultados de los individuos por debajo del punto límite con aquellos por encima del punto límite.²⁶

Contrafactual creado estadísticamente: Un modelo estadístico, como un análisis de regresión, se utiliza para desarrollar un cálculo de lo que podría haber sucedido en ausencia de una intervención. Esto se puede utilizar cuando la intervención ya está en escala; por ejemplo, una evaluación del impacto de la privatización de los servicios nacionales de suministro de agua.

Desarrollar un contrafactual creíble puede ser difícil en la práctica. A menudo es difícil parear individuos o comunidades en las variables que realmente marcan una diferencia. Las pruebas aleatorias controladas pueden crear grupos que no resulten equivalentes de manera aleatoria. Otros métodos dependen de diversas suposiciones que podrían no ser alcanzadas. En situaciones de

²⁵ Material de lectura adicional: Using Randomization in Development Economics Research: A Toolkit, <http://www.povertyactionlab.org/sites/default/files/documents/Using%20Randomization%20in%20Development%20Economics.pdf>.

²⁶ Material de lectura adicional: Impact Evaluation: Regression Discontinuity, <http://www.urban.org/toolkit/data-methods/regression.cfm>.

cambios rápidos e impredecibles, desarrollar un contrafactual creíble podría ser imposible. Se podría desarrollar un caso sólido y empírico en el que una intervención produjo ciertos impactos, pero sin tener la seguridad de lo que habría sucedido si no se hubiese implementado la intervención. Por ejemplo, puede ser posible mostrar que el desarrollo de la infraestructura de una comunidad destinada a la cría de pescado para su consumo o venta fue directamente debido al proyecto local, sin poder declarar con seguridad que esto no habría sucedido en ausencia del proyecto (quizás a través de un proyecto alternativo que otra organización haya implementado). La evaluación de impacto se puede centrar en los otros dos elementos del análisis causal: lo *factual* y *descartar* las alternativas.

El tercer componente para comprender los vínculos causales es investigar y descartar las explicaciones alternativas. Los impactos aparentes (o la falta de estos) puede reflejar problemas metodológicos, como el sesgo de selección (donde los participantes son sistemáticamente diferentes de aquellos que no participan), y los efectos de contaminación (donde quienes no participan también se benefician de la intervención, lo que reduce la diferencia entre ellos y los que participan en respecto a los impactos). Pueden reflejar la influencia de otros factores, incluyendo otras intervenciones o movimientos de la población entre las áreas asignadas para recibir una intervención y las que no lo han sido.

Métodos posibles para identificar y descartar explicaciones alternativas posibles incluyen:

Metodología de eliminación general: Se identifican las explicaciones alternativas posibles y luego se investigan para ver si se pueden descartar.²⁷

²⁷ Material de lectura adicional: Can We Infer Causation from Cross-Sectional Data?, http://www7.national-academies.org/BOTA/School-Level%20Data_Michael%20Scriven-Paper.pdf

Búsqueda de evidencia refutatoria/ Seguimiento de las excepciones²⁸

Una evaluación de impacto de la legislación para utilizar cascos de bicicletas obligatorios descubrió que hubo una disminución significativa en el número de lesiones en la cabeza en los ciclistas. Aunque esto era coherente con la teoría de cambio, una explicación alternativa era que el nivel general de las heridas disminuyó debido a la mayor cantidad de construcciones de carriles para bicicletas durante el mismo periodo. El examen de heridas graves mostró que, mientras que el nivel de las lesiones en la cabeza disminuyó en este periodo, el número de otros tipos de heridas se mantuvo estable, lo que respalda la teoría que fueron los cascos los que produjeron el cambio. (Walter et al., 2011)

Algunos enfoques que combinan estos diferentes elementos de explicación pueden ser:

Diversas líneas y niveles de evidencia

(MLLE): Un panel de expertos creíbles que abarcan un rango de disciplinas relevantes examinan un amplio rango de evidencia de diferentes fuentes. El panel identifica una coherencia con la teoría de cambio mientras que también identifica y explica las excepciones. Las MLLE examinan

la evidencia para una relación causal entre una intervención y los impactos observados en lo referente a su fortaleza, consistencia, especificidad, temporalidad, coherencia con otra evidencia aceptada, viabilidad y analogía con intervenciones similares.²⁹

Análisis de contribución: Un enfoque sistemático que implica desarrollar una teoría de cambio, mapear datos existentes, identificar desafíos a la teoría (incluyendo los vacíos en la evidencia y los vínculos causales impugnados) y reunir evidencia adicional de manera iterativa para abordarlos. La Guía de orientación 2 proporciona información adicional sobre el análisis de contribución.³⁰

Informe de resultados colaborativos: Este nuevo enfoque combina el análisis de contribución con las MLLE. Traza datos existentes contra la teoría de cambio y completa vacíos importantes en la evidencia a través de la recopilación enfocada de datos adicionales. Luego, se utiliza una combinación de examinación experta y de consulta comunitaria para verificar la credibilidad de la evidencia respecto a los impactos que han ocurrido y en qué medida se pueden atribuir de manera realista a la intervención.³¹

²⁸ Material de lectura adicional: Matthew B. Miles and A. Michael Huberman, *Qualitative Data Analysis: An Expanded Sourcebook, 2nd Edition* (Thousand Oaks, California: Sage Publications, 1994).

²⁹ Material de lectura adicional: Patricia Rogers, "Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation", *Journal of Development Effectiveness*, 1 (2009): 217-226. La versión del documento de trabajo está disponible en: http://www.3ieimpact.org/admin/pdfs_papers/50.pdf.

³⁰ Material de lectura adicional: Contribution Analysis: ILAC Guide Brief 16, http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf; Contribution Analysis, <http://www.scotland.gov.uk/Resource/Doc/175356/0116687.pdf>.

³¹ Material de lectura adicional: Collaborative Outcomes Reporting Technique, <http://www.aes.asn.au/regions/vic/2010/COR%20AES%202010.pdf>.

Una evaluación del impacto de un programa de desarrollo de liderazgo del gobierno no pudo utilizar un grupo de control aleatorio porque fue imposible asignar personas de manera aleatoria para estar en un grupo de control (o incluso para participar en el programa). La evaluación tampoco pudo utilizar un grupo de comparación porque la naturaleza del programa era tal que aquellos aceptados eran sistemáticamente diferentes de aquellos que no fueron aceptados. En cambio, la evaluación utilizó otras estrategias para la explicación causal, incluyendo la atribución por parte de los beneficiarios, la temporalidad y la especificidad (los cambios se produjeron en las áreas específicas abordadas por el programa). (Davidson, 2006)

10. Síntesis de la evidencia

Es poco común basar el juicio evaluativo general de una intervención en una sola medida de rendimiento. Normalmente, se requiere sintetizar evidencia sobre el rendimiento en diferentes dimensiones.

Una manera común de hacerlo es desarrollar una escala ponderada, donde un porcentaje de la calificación del rendimiento general se basa en cada dimensión. Sin embargo, una escala ponderada numérica a veces tiene problemas, como los pesos arbitrarios y la falta de atención a elementos esenciales, entre otros.³²

³² Material de lectura adicional: The Synthesis Problem: Issues and Methods in the Combination of Evaluation Results Into Overall Evaluative Conclusions, <http://realevaluation.com/pres/SynthHNL.pdf>.

Una alternativa es desarrollar una escala (o rúbrica) de evaluación global con usuarios deseados que puede ser utilizada para sintetizar la evidencia de manera transparente. La escala incluye una etiqueta para cada punto (por ejemplo, “sin éxito”, “con un poco de éxito”, “muy exitoso”) y una descripción de cómo se presenta cada uno.³³

11. Informe de los hallazgos y respaldo del uso

El formato del informe de la evaluación se debe acordar cuando se planea la evaluación de impacto. Algunas organizaciones tienen formatos de informe estándar, incluyendo los requisitos de extensión, que se deben cumplir. En otros casos, es importante acordar un “informe básico” de los títulos y subtítulos antes de que se escriba el informe.

Los informes de la evaluación de impacto son más accesibles cuando están organizados en torno a las preguntas clave de la evaluación, en lugar de informar de manera separada los datos de diferentes componentes de la recopilación de datos.³⁴

Se puede mejorar la calidad de los informes de evaluación involucrando las partes interesadas adecuadas. Aún en los casos en que un evaluador externo independiente esté realizando una evaluación, se pueden involucrar las partes interesadas al proporcionar respuestas formales sobre los hallazgos y comentar sobre los datos y cómo se han interpretado.

³³ Material de lectura adicional: The Rubric Revolution, <http://kinnect.co.nz/wp-content/uploads/2011/09/AES-2011-Rubric-Revolution-Davidson-Wehipeihana-McKegg-xx.pdf>.

³⁴ E. Jane Davidson, “Improving Evaluation Questions and Answers: Getting Actionable Answers for Real-World Decision Makers” (sesión de demostración en la Conferencia de la Asociación Americana de Evaluación, Orlando, FL, 18 de noviembre de 2009), <http://comm.eval.org/resources/viewdocument/?DocumentKey=e5bac388-f1e6-45ab-9e78-10e60cea0666>.

En los casos en que las recomendaciones se incluyen en los informes de evaluación, estas deben ser respaldadas con evidencia de los hallazgos de la evaluación y sobre la viabilidad y adecuación de las recomendaciones. Involucrar a las partes interesadas importantes en el desarrollo de las recomendaciones no solo mejora la viabilidad de las recomendaciones, sino que también puede lograr que las partes interesadas se apropien de ellas y se comprometan a implementarlas.

El uso de los informes de evaluación de impacto se puede mejorar mediante formatos de informes creativos, presentaciones verbales, oportunidades para participar con otros en el debate sobre las implicaciones de las evaluaciones de impacto, y al asegurarse de que los informes permanezcan accesibles para los usuarios potenciales.

La Guía de orientación 4 aborda la necesidad de comunicar bien los hallazgos a las audiencias deseadas.

12. ¿Cuándo se debe realizar una evaluación de impacto?

Las evaluaciones de impacto se deben realizar cuando haya necesidad e intenciones claras de utilizar los hallazgos. Si todas las intervenciones precisaran tener una evaluación de impacto, los evaluadores correrían el riesgo de necesitar un exceso de recursos o de dispersar tanto los recursos que las evaluaciones resultarían superficiales. Una estrategia más eficaz es centrar los recursos de evaluación de impacto en las intervenciones en las que pueden ser más útiles:

- Las intervenciones y los programas piloto innovadores que, si se prueban de manera exitosa, se pueden ampliar o reproducir.
- Intervenciones en donde no hay un buen

entendimiento de sus impactos y es necesaria una mejor evidencia para informar las decisiones acerca de su financiación; si se debe continuar o dirigirla a otras intervenciones.

- Las evaluaciones periódicas del impacto de una cartera de intervenciones en un sector o una región para orientar políticas, el diseño de intervención futura y las decisiones de financiación.
- Intervenciones con un mayor perfil de riesgos, como una gran inversión (actual o futura), un alto potencial para impactos negativos importantes o problemas delicados de políticas.
- Intervenciones en donde las partes interesadas necesitan comprender de mejor manera las contribuciones y perspectivas de cada uno.

El momento de una evaluación de impacto es importante. Si se realiza demasiado pronto, la evidencia de los impactos que han ocurrido o están en curso pueden resultar insuficientes. Si se realiza demasiado tarde, puede ser difícil realizar un seguimiento de los participantes y demasiado tarde para influenciar las decisiones acerca de la futura dirección de la intervención. En cualquier caso, es mejor planificar la evaluación de impacto, en lo posible, desde el inicio de la intervención. Esto permite que se reúna la evidencia a lo largo de toda la intervención, incluyendo los datos de base, y permite utilizar métodos como las pruebas aleatorias controladas, que requieren la creación de un grupo de control asignado aleatoriamente desde el inicio de la implementación.

13. ¿Qué se necesita para la evaluación de impacto de calidad?

Puede ser útil pensar acerca de la evaluación de calidad en términos de cinco imperativos competitivos: *utilidad*, *precisión*, *ética*, *funcionalidad* y

responsabilidad.³⁵ Estos cinco estándares están, a menudo, en conflicto. Por ejemplo, una evaluación más exhaustiva de impacto que sería más precisa, podría no ser práctica en respecto a los recursos disponibles, podría ser intrusiva en los datos reunidos, o podría llevar demasiado tiempo para completarse para que pueda informar decisiones clave acerca del futuro de la intervención.

Utilidad: Una buena evaluación de impacto es útil. La posible utilidad de una evaluación se puede mejorar planificando cómo sería utilizada desde el comienzo, lo que incluye relacionarla con el tiempo y los procesos de toma de decisiones organizativas, ser claros acerca del motivo por el que se está realizando y acerca de quién la utilizará, contratar partes interesadas clave en el proceso y luego elegir diseños y métodos para lograr este objetivo.

Precisión: Una buena evaluación de impacto es minuciosa. Le presta atención a todos los impactos importantes y observa si hay alguno indeseado. Le presta atención a la distribución de impactos y observa si solo algunos se benefician, y quiénes son esas personas. La precisión requiere el uso de evidencia adecuada, que incluye los datos cuantitativos y cualitativos, la interpretación adecuada y la transparencia acerca de las fuentes de datos que se han utilizado y sus limitaciones. Las estrategias para lograr la precisión incluyen sistemas para verificar la calidad de los datos en el punto de la recopilación y durante el procesamiento, y que los hallazgos se hayan informado de manera justa, completa y clara.

³⁵ Joint Committee Standards for Educational Evaluation <http://www.jcsee.org/program-evaluation-standards/program-evaluation-standards-statements>. Estos fueron originalmente desarrollados para evaluaciones educativas pero, a veces, se los usa de manera más amplia.

Propiedad (ética): Los problemas éticos se deben abordar de manera adecuada; es decir, en confidencialidad y anonimato, así como también se deben abordar los posibles efectos nocivos de estar involucrados en la evaluación. Algunos problemas éticos, como la necesidad de cumplir con las promesas que se realizan acerca de la privacidad y de la confidencialidad, son comunes en diferentes tipos de evaluaciones e investigaciones. Hay otros problemas que son particulares de la evaluación de impacto. A veces se formulan inquietudes acerca de la ética de utilizar un diseño RCT, ya que implica retener una intervención de ciertas personas (el grupo de control). Hay una preocupación ética menor cuando el acceso a la intervención se raciona en cualquier caso, y se puede abordar asignando el grupo de control a la cola para que reciban la intervención después de que se haya finalizado la evaluación de la primera fase (si muestra ser eficaz). Sin embargo, esta estrategia es factible solo cuando los impactos (o indicadores creíbles de ellos) son evidentes de manera temprana y cuando la intervención aún es relevante para el grupo de control al momento que finaliza la evaluación.

También hay problemas éticos potenciales en respecto a los intereses de quiénes abordan una evaluación. La Asociación Americana de Evaluación debate esto en sus Principios rectores en términos de “Responsabilidades para el bienestar general y público”:

Los evaluadores expresan y toman en cuenta la diversidad de los intereses y valores generales y públicos, y por lo tanto deben:

1. *Incluir perspectivas e intereses importantes de la gama completa de las partes interesadas.*

2. *Considerar no solo las operaciones y los resultados inmediatos de la evaluación, sino también los supuestos generales, las implicaciones y los posibles efectos secundarios.*

3. *Permitir que las partes interesadas accedan a la información evaluativa y que puedan divulgarla activamente, y presentar los resultados en formas entendibles que respeten a las personas y cumplan las promesas de confidencialidad.*

4. *Mantener un equilibrio entre las necesidades e intereses del cliente y de las demás partes interesadas.*

5. *Tener en cuenta el interés y el bienestar público; ir más allá del análisis de los intereses particulares de la parte interesada para considerar el bienestar de la sociedad como un todo.*

Normalmente se necesita una aprobación formal realizada por la junta de revisión institucional adecuada para realizar una evaluación de impacto. Las solicitudes para la aprobación necesitan seguir el formato obligatorio y abordar problemas de beneficencia, justicia y respeto.³⁶

Funcionalidad: Las evaluaciones de impacto deben ser prácticas. Deben tener en cuenta los recursos disponibles (tiempo, dinero, experiencia y datos existentes) y cuándo se necesitan los resultados de la evaluación para informar las decisiones. Asociarse con uno o más profesionales de evaluación, organizaciones de investigación,

³⁶ Material de lectura adicional: Evaluation Consent and the Institutional Review Board Process, <http://meera.snre.umich.edu/plan-an-evaluation/plonearticlemultipage.2007-10-30.3630902539/evaluation-consent-and-the-institutional-review-board-process>.

universidades y organizaciones de sociedad civil puede impulsar los recursos necesarios.

Responsabilidad: las evaluaciones necesitan aclarar la evidencia y los criterios en los cuales se basaron las conclusiones, y reconocer sus limitaciones. La transparencia sobre las fuentes de datos es importante; por ejemplo, mostrar las fuentes que se utilizaron para cada pregunta de evaluación. Un proceso formal de metaevaluación (hacer que se evalúe su propia evaluación aprobando un plan de evaluación y luego un informe de evaluación) realizado por un revisor experto o un comité de individuos con integridad e independencia respetadas, puede mejorar la responsabilidad de una evaluación de impacto.

14. Desafíos comunes en la evaluación de impacto en desarrollo

A continuación se describe un número de desafíos comunes en la evaluación de desarrollo, junto con algunas sugerencias para abordarlos.

- **Variación en la implementación y en el entorno en diferentes sitios**

Es posible que se haya implementado una intervención de distintas maneras para adaptarla a los diferentes contextos en las diferentes oficinas nacionales en todo el mundo, o en diferentes áreas geográficas dentro de un país. Puede ser útil comparar las teorías de cambio para cada sitio. En particular, identificar si los diferentes sitios utilizan la misma teoría acerca de cómo suceden los cambios (por ejemplo, al aumentar el conocimiento de las personas acerca de sus derechos a los servicios) pero diferentes teorías de acción (por ejemplo, folletos impresos en comparación con teatro comunitario),

o si están utilizando teorías de cambio completamente diferentes (por ejemplo, aumentar el conocimiento de las personas acerca de sus derechos a los servicios en un sitio en comparación con reducir barreras para el acceso a los servicios, como cuotas de usuarios, mediante la promoción en otras).

- **Impactos heterogéneos**

A menudo, las intervenciones de desarrollo solo funcionan para algunas personas, y pueden ser ineficaces o hasta perjudiciales para otras. Además, el éxito de una intervención en cuanto al logro de los impactos deseables, a veces, se ve afectado por la calidad de la implementación. Por lo tanto, es importante no solo calcular e informar el efecto promedio, sino también buscar diferentes efectos. Esto requiere reunir evidencia donde sea posible acerca de la calidad de la implementación y recopilar datos acerca de los factores contextuales que podrían afectar a los impactos, incluyendo las características de los participantes y el entorno de la implementación.

- **Componentes diversos**

Un programa podría abarcar un rango diverso de proyectos, y aún así se necesitaría una evaluación general del impacto del programa entero. Puede ser útil desarrollar una teoría general de cambio para el programa, reuniendo diferentes componentes. A veces, es posible hacer esto en la etapa de planificación, pero, especialmente en los casos en que los proyectos y componentes varían con el tiempo, podría ser necesario realizarlo de manera retroactiva.

- **Plazos largos**

A menudo, los impactos buscados no son evidentes por muchos años, pero se necesita la evidencia para informar las decisiones antes de ese momento (por ejemplo, si se debe o no ejecutar una fase posterior o replicar un modelo en otro lugar). Una teoría de cambio puede identificar resultados intermedios que podrían ser evidentes en la vida de una evaluación. En algunos casos, la evidencia de la investigación se puede utilizar para completar vínculos posteriores y calcular posibles impactos según el logro de los resultados inmediatos. También se debe dar consideración a la trayectoria esperada de cambio; cuando es posible que los impactos sean evidentes.³⁷

- **Influencia de otros programas y factores**

Los impactos de las intervenciones de desarrollo están muy influenciados por las actividades de otros programas y otros factores contextuales que podrían respaldar o prevenir dichos impactos. Por ejemplo, las transferencias de dinero que dependen de la asistencia escolar solo conducen a un logro estudiantil mejorado en situaciones en donde las escuelas les enseñan a los estudiantes de manera adecuada. Es posible identificar estos otros programas y factores contextuales como parte del desarrollo de una teoría de cambio, para reunir evidencia acerca de ellos y para buscar patrones en los datos.

37 Material de lectura adicional: Michael Woolcock sobre la Importancia del Tiempo y las Trayectorias en la Comprensión de la Eficacia del Proyecto, <http://blogs.worldbank.org/impactevaluations/guest-post-michael-woolcock-on-the-importance-of-time-and-trajectories-in-understanding-project-effe>.

- **Limitaciones de los recursos**

La evidencia existente (en la forma de documentación del programa, datos de base y estadísticas oficiales) puede tener vacíos y es posible que hayan pocos recursos (en términos de la financiación, el tiempo del personal o el acceso a los conocimientos técnicos especializados) para reunir los tipos de evidencia necesarios para una evaluación de impacto de calidad. Para una evaluación específica, cuando la evidencia existente es escasa y hay pocos recursos para reunir evidencia adicional, las entrevistas de informantes clave de diversos informantes pueden proporcionar suficientes datos, inclusive reconstruir datos de base. Planificar con anticipación para la evaluación de impacto puede reducir las limitaciones de los recursos al desarrollar suficientes recursos en la etapa de diseño y presupuesto, y/o asignar estratégicamente recursos de evaluación en todas las intervenciones para que estén concentradas en mayor medida en un número más pequeño de evaluaciones más exhaustivas de intervenciones estratégicamente importantes.

Resumen

Una evaluación de impacto debe comenzar con un plan que aclare los objetivos que pretende alcanzar, lo que incluye identificar a los usuarios previstos, las preguntas de evaluación clave que intenta responder y cómo abordará los seis componentes de la evaluación de impacto: clarificar valores, desarrollar una teoría de cambio, medir o describir variables importantes, explicar qué ha producido los impactos, sintetizar evidencia e informar y respaldar el uso. Que se revise este plan (puede ser por los usuarios previstos) aumentará la

posibilidad de producir una evaluación de impacto de alta calidad que se utilice realmente.

Referencias y otros recursos útiles

Alton-Lee, A. (2003) "Quality Teaching for Diverse Students in Schooling: Best Evidence Synthesis." Wellington, New Zealand: Ministry of Education. <http://www.educationcounts.govt.nz/publications/series/2515/5959> *Un ejemplo de síntesis de evidencia de diversas fuentes para comprender qué funciona para quién.*

Catley A., Burns, J., Abebe, D., Sufi, O. Participatory Impact Assessment: A Guide for Practitioners. Boston: Tufts University. <http://www.prevention-web.net/english/professional/publications/v.php?id=9679>

Chambers R. (2007) "Who Counts? The Quiet Revolution of Participation and Numbers" Working Paper No. 296, Brighton: Institute of Development Studies. <http://www.ids.ac.uk/files/Wp296.pdf>

Davidson, E. J. (2006) "Causal Inference Nuts and Bolts." Sesión de demostración en la Conferencia de la Asociación Americana de Evaluación de 2006, Portland, Ore., Nov. 2006 <http://realevaluation.com/pres/causation-an-zea09.pdf>.

Davidson, E. J. (2009) "Improving Evaluation Questions and Answers: Getting Actionable Answers for Real-World Decision Makers." Sesión de demostración en la Conferencia de la Asociación Americana de Evaluación de 2009, Orlando, Fla., Nov. 2009. <http://comm.eval.org/resources/viewdocument/?DocumentKey=e5bac388-f1e6-45ab-9e78-10e60ceao666>.

- Funnell S. and Rogers, P. (2011) *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco: Jossey-Bass/Wiley.
- Guijt, I. (1999) Participatory Monitoring and Evaluation for Natural Resource Management and Research. Socio-economic Methodologies for Natural Resources Research. Chatham, UK: Natural Resources Institute. <http://www.nri.org/publications/bpg/bpgo4.pdf>
- Miles, M. and Huberman, M. (1994) *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.) Thousand Oaks California; Sage Publications. *Bosquejo de estrategias para verificar explicaciones causales, como la búsqueda de evidencia refutatoria, el seguimiento de excepciones y la realización y prueba de las predicciones.*
- Patton, MQ (2008) "State of the Art in Measuring Development Assistance." *Debate sobre la importancia de la interpretación y el manejo de la incertidumbre en una administración eficaz.*
- Paz R., Dorward A., Douthwaite B. (2006). "Methodological Guide for Evaluation of Pro-Poor Impact of Small-Scale Agricultural Projects." Centre for Development and Poverty Reduction. Imperial College, London. <http://boru.pbworks.com/f/modulosjano7.pdf> *Descripción de 22 métodos y herramientas para la evaluación de los impactos directos e indirectos de la adopción de innovación.*
- Roche, C. (1999) *Impact Assessment for Development Agencies: Learning to Value Change* Oxford: OXFAM, Novib
- Rogers, Patricia J. (2009) "Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation", *Journal of Development Effectiveness*, 1(3): 217-226. La hoja de trabajo está disponible en: http://www.3ieimpact.org/admin/pdfs_papers/50.pdf.
- Walter, S., Olivier, J., Churches, T. and Grzebieta, R. (2011). "The impact of compulsory cycle helmet legislation on cyclist helmet head injuries in New South Wales, Australia", *Accident Analysis and Prevention*, 43 : 2064-2071.
- White, S. and J. Petit (2004) *Participatory Methods and the Measurement of Wellbeing Participatory Learning and Action 50*, London: IIED
- www.betterevaluation.org – información sobre métodos de evaluación para el desarrollo, que incluye ejemplos y comentarios que el usuario aporta
- www.mymande.org – información, videos y vínculos a información sobre los métodos de evaluación