GOBIERNO REGIONAL CUSCO
GERENCIA REGIONAL DE PLANEAMIENTO, PRESUPUESTO Y MODERNIZACION
Sub Gerencia de Planeamiento, Ordenamiento y Demarcación Territorial
Servicio de Información Estadística Regional

unicef
Office of Research-Innocenti

Síntesis metodológica

Sinopsis de la evaluación de impacto n° 10

Sinopsis: Métodos de Recolección y Análisis de Datos en la Evaluación de Impacto

Greet Peersman

RMIT UNIVERSITY
3ie International Initiative for Impact Evaluation
BetterEvaluation
unicef Office of Research-Innocenti

CENTRO DE INVESTIGACIONES INNOCENTI DE UNICEF

El Centro de Investigaciones Innocenti es la oficina de UNICEF especializada en investigación científica. El objetivo principal del Centro de Investigaciones es mejorar la comprensión internacional de una serie de cuestiones relacionadas con los derechos de la infancia, a fin de facilitar la plena aplicación de la Convención sobre los Derechos del Niño mediante su promoción en todo el mundo. El centro tiene el cometido de establecer un marco integral de investigación y conocimiento dentro de la organización para brindar apoyo a sus programas y políticas mundiales. Las publicaciones producidas por el centro contribuyen al debate global sobre la infancia y los derechos del niño e incluyen una amplia gama de opiniones.

Las opiniones expresadas corresponden a los autores o editores y se publican para estimular un mayor diálogo sobre métodos de análisis de impacto. Esta publicación no refleja necesariamente las políticas o perspectivas de UNICEF sobre ciertos temas.

SINTESIS METODOLOGICAS DEL CENTRO DE INVESTIGACIONES

Las síntesis metodológicas del Centro de Investigaciones de UNICEF pretender compartir prácticas de investigación, métodos, diseños y recomendaciones de reconocidos investigadores y analistas. La audiencia a la que van dirigidas es principalmente personal de UNICEF que lleve a cabo, encargue o interprete los resultados de investigación y análisis para la toma de decisiones sobre programas y políticas de sensibilización.

Esta síntesis metodológica ha seguido un proceso de revisión por pares interna.

El testo no ha sido editado de acuerdo con los estándares de publicación oficiales y UNICEF declina toda responsabilidad por posibles errores

Se permite la reproducción de cualquier parte de la presente publicación siempre que se incluya referencia a la presente. Si se desea utilizar una parte sustancial de la publicación dirijan su solicitud al Departamento de Comunicación en la dirección de correo electrónico: florence@unicef.org

Para consultas o descargas, pueden encontrar estas síntesis metodológicas en http://www.unicef-irc.org/KM/IE/

Recomendamos la siguiente cita para cualquier referencia al presente documento:

Peersman, G. (2014).Sinopsis: Métodos de recolección y análisis de datos en la evaluación de Impacto, *Síntesis metodológica n.°10*, Centro de Investigaciones de UNICEF, Florencia.

Agradecimientos: Varios autores han proporcionado orientación en la preparación de esta síntesis. El autor y el Centro de Investigaciones de UNICEF desean agradecer a todos aquellos que han participado en la preparación de la presente publicación, especialmente a:
Por su contribución: Simon Hearn, Jessica Sinclair Taylor
Por su revisión: Nikola Balvin, Claudia Cappa, Yan Mu

© Fondo de las Naciones Unidas para la Infancia (UNICEF), septiembre de 2014
Centro de Investigaciones Innocenti de UNICEF

Piazza SS. Annunziata, 12
50122 Florencia (Italia)
Tel: (+39) 055 20 330
Fax: (+39) 055 2033 220
florence@unicef.org
www.unicef-irc.org

1. RECOLECCIÓN Y ANÁLISIS DE DATOS: BREVE DESCRIPCIÓN

Uno Elegir y aplicar bien los métodos de recolección y análisis de datos es esencial en todos los tipos de evaluaciones. Esta síntesis presenta una visión general de las cuestiones relacionadas con la elección y uso de métodos para las evaluaciones de impacto, es decir, las evaluaciones que proporcionan información sobre los efectos a largo plazo previstos y no previstos producidos por los programas o políticas.

Las evaluaciones de impacto deben ir más allá de la simple evaluación de la magnitud de los efectos (el impacto medio) para determinar con quién y de qué forma ha tenido éxito un programa o política. Se deben examinar por anticipado los factores que contribuyen al éxito y la forma de analizar y sintetizar los datos para responder a las preguntas clave de evaluación específicas, puesto que la recolección de datos debe orientarse a la combinación de pruebas empíricas necesarias para tomar decisiones adecuadas sobre el programa o política. En otras palabras, durante la fase de planificación de la evaluación se debe especificar el marco analítico: la metodología para analizar el «significado» de los datos examinando patrones de manera sistemática y transparente. El marco contiene la forma en que el análisis de datos abordará los supuestos formulados en la teoría del cambio del programa sobre cómo se cree que el programa producirá los resultados previstos (véase la Síntesis n.º 2 (Teoría del cambio)). En una verdadera evaluación con métodos mixtos, esto implica utilizar métodos de análisis numéricos y textuales adecuados y triangular múltiples fuentes de datos y perspectivas para maximizar la fiabilidad de las constataciones de la evaluación.

Puntos principales

1. Los métodos de recolección y análisis de datos deben elegirse en función de las preguntas clave de evaluación y los recursos disponibles de la evaluación.
2. Las evaluaciones de impacto deben aprovechar al máximo los datos existentes y subsanar las carencias con nuevos datos.
3. Los métodos de recolección y análisis de datos deben elegirse de forma que se complementen entre sí los puntos fuertes y débiles.

Empezar por la planificación general de la evaluación

Antes de tomar decisiones sobre qué datos recopilar y cómo analizarlos, debe decidirse la finalidad de la evaluación (es decir, los usuarios y usos previstos) y las preguntas clave de evaluación (véase la Síntesis n.º 1 (Sinopsis de la evaluación de impacto)). Puede encargarse una evaluación de impacto para que sirva de base a las decisiones sobre la realización de cambios en un programa o política (evaluación formativa) o sobre la continuación, finalización, reproducción o ampliación de un programa o política (evaluación sumativa). Una vez que está clara la finalidad de la evaluación, debe acordarse un pequeño número de preguntas clave de evaluación de alto nivel (no más de 10),

idealmente con la aportación de las principales partes interesadas; a veces, las preguntas clave de evaluación ya están prescritas por un sistema de evaluación o un marco de evaluación anteriormente elaborado. Las respuestas a las preguntas clave de evaluación —independientemente de cómo se llegue a ellas— deben asegurar que se cumpla la finalidad de la evaluación. Contar con un conjunto acordado de preguntas clave de evaluación ofrece orientación sobre qué datos recopilar, cómo analizar los datos y cómo informar sobre las constataciones de la evaluación.

Una herramienta esencial en la evaluación de impacto es una teoría del cambio bien desarrollada. Esta describe cómo se entiende el funcionamiento del programa o política: representa un modelo causal que relaciona los insumos y las actividades con los productos y los resultados e impactos deseados (véase la Síntesis n.º 2 (Teoría del cambio)). La teoría del cambio debe tener en cuenta también los resultados no previstos (positivos y negativos). Esta herramienta no solo resulta útil en la fase de diseño del programa, sino que también ayuda a centrar la evaluación de impacto en aquello que necesitan saber las partes interesadas sobre el programa o política para apoyar la toma de decisiones; en otras palabras: las preguntas clave de evaluación. Unas buenas preguntas de evaluación no solo plantean «¿cuáles fueron los resultados?» (preguntas descriptivas), sino también «*¿cómo de buenos* fueron los resultados?» (juzgando el valor del programa o política). Las evaluaciones de impacto deben recabar pruebas empíricas de los impactos (por ejemplo, cambios positivos en las tasas de mortalidad de los niños menores de 5 años) y también examinar cómo se lograron los impactos previstos o por qué no se alcanzaron. Esto requiere datos sobre el contexto (por ejemplo, el marco normativo y jurídico que afecta a la protección infantil en un país determinado), la idoneidad y calidad de las actividades del programa o la aplicación de la política, y una serie de resultados intermedios (por ejemplo, asimilación de la inmunización), como variables explicativas en la cadena causal[1].

Aprovechar al máximo los datos existentes

La planificación de la recolección de datos debe comenzar por revisar en qué medida pueden utilizarse los datos existentes. En términos de indicadores, la evaluación debe procurar basarse en distintos tipos de indicadores (insumos, productos, resultados, impactos) para reflejar los principales resultados en la teoría del cambio del programa. Las evaluaciones de impacto deben utilizar idealmente los indicadores que se seleccionaron para supervisar el desempeño durante el periodo de aplicación del programa, es decir, los indicadores clave del desempeño. En muchos casos, también es posible basarse en los datos recopilados mediante encuestas de población normalizadas como la Encuesta a base de indicadores múltiples de UNICEF, la Encuesta Demográfica y de Salud y el Estudio sobre la Medición de Niveles de Vida.

Es especialmente importante comprobar si existen datos de referencia para los indicadores seleccionados, así como para las características sociales y demográficas y otras características pertinentes de la población objeto de estudio. Cuando el diseño de evaluación implica comparar cambios a lo largo del tiempo en distintos grupos, los datos de referencia pueden utilizarse para determinar la equivalencia de los grupos antes de que empiece el programa o «emparejar» distintos grupos (como en el caso de los diseños cuasi experimentales; véase la Síntesis n.º 8 (Diseño y métodos cuasi experimentales)). También son importantes para determinar si se ha producido un cambio a lo largo del tiempo y la magnitud de este cambio (es decir, la magnitud del efecto). Si no se dispone de datos de referencia, será necesario recopilar datos adicionales para reconstruir las bases de referencia, por ejemplo, utilizando la «memoria» (pedir a las personas que recuerden información específica sobre un acontecimiento o experiencia que se produjo en el pasado). Aunque la memoria puede estar abierta a sesgos, pueden reducirse sustancialmente siendo realistas sobre lo que

[1] La Síntesis n.º 1 (Sinopsis de la evaluación de impacto) cubre la necesidad de distintos enfoques de evaluación de las políticas en vez de programas.

pueden recordar las personas y qué es menos probable que recuerden, y utilizando herramientas de encuesta establecidas[2].

Otras fuentes comunes de datos existentes son las estadísticas oficiales, los datos de supervisión del programa, los registros del programa (que pueden incluir una descripción del programa, una teoría del cambio, actas de reuniones importantes, etc.), documentos de política formales y planes e informes sobre los avances en la aplicación del programa. Si bien es importante aprovechar al máximo los datos existentes en aras de la eficiencia, los datos deben tener una calidad suficiente para no comprometer la validez de las constataciones de la evaluación (véase más adelante).

Detectar y corregir las carencias importantes de datos

Tras examinar la información actualmente disponible, conviene crear una matriz de evaluación (véase el cuadro 1) que muestre qué métodos de recolección y análisis de datos se utilizarán para responder a cada una de las preguntas clave de evaluación y a continuación detectar y priorizar las carencias de datos que deben corregirse recopilando nuevos datos. Esto ayudará a confirmar que la recolección de datos planificada (y el cotejo de los datos existentes) va a abarcar todas las preguntas clave de evaluación, determinar si hay suficiente triangulación entre las distintas fuentes de datos y ayudar al diseño de las herramientas de recolección de datos (como cuestionarios, preguntas de entrevista, herramientas de extracción de datos para la revisión de documentos y herramientas de observación) a fin de garantizar que recaban la información necesaria.

Cuadro 1. Matriz de evaluación: emparejamiento de la recolección de datos con las preguntas clave de evaluación

Ejemplos de preguntas clave de evaluación (PCE)	Encuesta de participantes del programa	Entrevistas a informantes clave	Registros de proyectos	Observación de la ejecución del programa
PCE 1 ¿Cuál fue la calidad de la ejecución?		✔	✔	✔
PCE 2 ¿En qué medida se cumplieron los objetivos del programa?	✔	✔	✔	
PCE 3 ¿Qué otros impactos produjo el programa?	✔	✔		
PCE 4 ¿Cómo podría mejorarse el programa?		✔		✔

2 White, Howard, «A contribution to current debates in impact evaluation», *Evaluation*, 16(2), 2010, págs. 153–164.

Existen muchos métodos diferentes para recopilar datos. El cuadro 2 proporciona ejemplos de posibles fuentes de datos (existentes y nuevas)[3].

Cuadro 2. Opciones de recolección (datos primarios) y cotejo (datos segundarios)

Opción	¿Qué podría incluir?	Ejemplos
Recuperación de documentos y datos existentes	• Documentos de política formales, planes e informes de ejecución • Estadísticas oficiales • Datos de supervisión del programa • Registros del programa	• Revisar los documentos de planificación del programa, las actas de las reuniones, los informes sobre los avances • El perfil político, socioeconómico o sanitario del país o el lugar específico donde se ejecutó el programa
Recolección de datos de personas o grupos	• Entrevistas[4] a informantes clave, personas, grupos, grupos dirigidos de discusión, técnicas de proyección • Cuestionarios o encuestas a través de correo electrónico, sitios web, cara a cara, datos móviles • Métodos especializados (por ejemplo, votación por puntos, clasificación de tarjetas, calendarios estacionales, técnicas de proyección, vivencias[5]).	• Entrevistas con informantes clave entre los representantes de departamentos gubernamentales competentes, organizaciones no gubernamentales o la comunidad de desarrollo en general • Entrevistas con directores del programa, ejecutores del programa y los responsables de la supervisión rutinaria del programa • Entrevistas, debates en grupo (como grupos de discusión) o cuestionarios con participantes del programa

[3] Puede encontrarse más información sobre cada una de ellos y una lista más exhaustiva de opciones de recolección/cotejo de datos en «Collect and/or Retrieve Data» en el sitio web BetterEvaluation, en http://betterevaluation.org/plan/describe/collect_retrieve_data.

[4] Véase la Síntesis n.º 12 (Entrevistas).

[5] Votación por puntos: recabar niveles de acuerdo sobre declaraciones escritas entre un gran número de personas. Clasificación de tarjetas: da una idea de cómo las personas categorizan y clasifican distintos fenómenos. Calendarios estacionales: visualizar patrones de variaciones a lo largo de determinados periodos de tiempo. Técnicas de proyección: dar pie a entrevistas (por ejemplo, utilizando fotolenguaje; los participantes seleccionan una o dos imágenes de un conjunto y las utilizan para ilustrar sus comentarios sobre algo). Vivencias: por ejemplo vivencias personales para dar una idea de cómo viven la vida las personas.

Observación	• Estructurada o no estructurada • De participantes o no participantes • Participativa o no participativa • Registrada a través de notas, fotos o vídeos	• Observaciones de las actividades del programa y las interacciones con los participantes
Medición física	• Mediciones biofísicas • Información geográfica	• Peso de los lactantes • Lugares con alta prevalencia de infección por VIH

Utilizar una serie de métodos de recolección y análisis de datos

Aunque muchas evaluaciones de impacto utilizan varios métodos, lo que distingue a una «evaluación de métodos mixtos» es la *integración sistemática* de las metodologías y métodos cuantitativos y cualitativos en todas las fases de una evaluación[6]. Una de las razones principales de la combinación de métodos es que ayuda a corregir los puntos débiles inherentes a cada método cuando se utiliza individualmente. También aumenta la fiabilidad de las constataciones de la evaluación cuando la información de distintas fuentes de datos converge (es decir, coincide en la dirección de las constataciones) y puede profundizar el entendimiento del programa o política, sus efectos y su contexto[7].

Las decisiones en torno al uso de un enfoque de métodos mixtos suponen determinar:

- en qué fase de la evaluación se combinan los métodos (el diseño se considera mucho más sólido si los métodos mixtos se integran en varias o en todas las fases de la evaluación);
- si los métodos se utilizarán de forma secuencial (los datos de una fuente sirven de base para la recolección de datos de otra fuente) o simultánea (se utiliza la triangulación para comparar información de distintas fuentes independientes);
- si se dará relativamente la misma ponderación a los métodos cualitativos y cuantitativos o no;
- si el diseño será de un solo nivel (por ejemplo, los hogares) o de múltiples niveles (por ejemplo, un programa nacional que requiere la descripción y el análisis de los vínculos entre los distintos niveles).

El marco analítico particular y la elección de los métodos de análisis de datos específicos dependerán de la finalidad de la evaluación de impacto y del tipo de preguntas clave de evaluación intrínsecamente relacionadas:

- Las **preguntas descriptivas** requieren métodos de análisis de datos que incluyan datos cuantitativos y cualitativos.
- Las **preguntas causales** requieren un diseño de investigación que aborde la atribución (si los cambios observados se deben o no a la intervención o a factores externos) y la contribución

[6] Bamberger, Michael, «Introduction to Mixed Methods in Impact Evaluation», *Guidance Note* No. 3, InterAction, Washington, D.C., agosto de 2012. Véase http://www.interaction.org/impact-evaluation-notes.

[7] *Ibíd.*

(en qué medida causó la intervención los cambios observados, véase la Síntesis n.º 6 (Estrategias de atribución causal)).

- Las **preguntas evaluativas** requieren estrategias de síntesis que apliquen los criterios evaluativos a los datos para responder a las preguntas clave de evaluación (véase la Síntesis n.º 3 (Criterios de evaluación)). Definir por anticipado los factores de éxito construyendo rúbricas evaluativas específicas (estándares o niveles de desempeño del programa o política) sienta la base sobre la cual puede combinarse sistemáticamente la información recabada para tomar decisiones con base empírica transparentes sobre el valor del programa o política (también llamado «razonamiento evaluativo», véase la Síntesis n.º 4 (Razonamiento evaluativo)).

Si bien la evaluación de impacto tiene por objeto analizar los resultados a largo plazo de un programa o política, los responsables de tomar decisiones suelen necesitar información más puntual y, por lo tanto, deben recopilar también datos sobre los resultados a corto plazo. Por ejemplo, es bien sabido que los resultados de las intervenciones en materia de educación solo aparecen después de un periodo de tiempo prolongado. En el caso de la iniciativa de escuelas amigas de la infancia en Moldava, su evaluación solo captó los resultados a corto plazo (como «la mayor participación de los estudiantes en el aprendizaje a través de métodos de enseñanza interactivos y participativos»[8]), que fueron medidos durante la intervención o poco después de su finalización, y supuso que estos predecían los efectos a largo plazo.

Sin embargo, determinar simplemente que se ha producido un cambio —midiendo indicadores clave— no indica por qué ha ocurrido. También se necesita información sobre las actividades específicas que se llevaron a cabo y sobre el contexto en el que se llevaron a cabo. Como se ha señalado anteriormente, es esencial una teoría del cambio explícita para el programa o política a fin de determinar qué mediciones deben recabarse, además de que orienta sobre los aspectos de la ejecución del programa —y su contexto— en los que debe centrarse la recolección de datos. Si se especifica el marco de análisis de datos al inicio, se incorporan claramente las necesidades concretas de recolección de datos (datos primarios o nuevos que deben recopilarse) y cotejo de datos (datos secundarios o existentes), reflejando también cómo se analizarán los datos para responder a las preguntas clave de evaluación y realizar ajustes evaluativos. Las necesidades de datos y los métodos de recolección y análisis de datos vinculados a cada una de las preguntas clave de evaluación deben describirse en el plan de evaluación junto con las especificaciones de cómo, dónde, cuándo y de quién se recolectarán datos, con referencia a la estrategia de muestreo de la población objeto de estudio, los lugares o los periodos de tiempo.

Asegurar que los métodos de recolección y análisis de datos seleccionados son viables

Una vez que se completa la planificación, es importante comprobar la viabilidad de los métodos de recolección y análisis de datos para asegurar que lo que se ha propuesto puede lograrse realmente dentro de los límites del marco temporal y los recursos de la evaluación. Por ejemplo, puede ocurrir que los informantes clave no puedan reunirse en el momento en que se necesitan los datos. También es importante analizar los equipos y cualificaciones que se necesitarán para utilizar estos métodos y evaluar si están disponibles o pueden conseguirse o desarrollarse. Por ejemplo, la recolección de datos mediante un cuestionario a través del teléfono móvil exigirá que todos los responsables de la

[8] Velea, Simona y CReDO (Centro de Recursos de Derechos Humanos), *Child-Friendly Schools, External Evaluation Report of the Child-Friendly School Initiative (2007–2011), Republic of Moldova*, Ministerio de Educación de la República de Moldova/UNICEF, 2012. Véase http://www.unicef.org/moldova/CFS_EN_PRINT.pdf.

recolección dispongan de un teléfono móvil o que exista un sistema fiable para compartir móviles entre los responsables de la recolección. Las diferencias importantes entre lo que hay disponible y lo que se necesita deben corregirse adquiriendo recursos adicionales o, de forma más realista, adaptando los métodos a los recursos disponibles.

Teniendo en cuenta que no todo puede anticiparse y que determinadas condiciones pueden cambiar durante el transcurso de la evaluación, quizá resulte necesario revisar las opciones y también el plan de evaluación. En estos casos es importante documentar lo que ha cambiado y por qué, y tener en cuenta y documentar todas las implicaciones que pueden tener estos cambios en el producto de evaluación y su uso.

3. GARANTIZAR UNA BUENA GESTIÓN DE DATOS

Una buena gestión de datos implica desarrollar procesos eficaces para la recolección y el registro sistemáticos de datos, el almacenamiento seguro de los datos, la depuración de datos, la transferencia de los datos (por ejemplo, entre distintos tipos de programas informáticos utilizados para el análisis), la presentación eficaz de los datos y la accesibilidad de los datos para su verificación y utilización por terceros.

Los aspectos de calidad de los datos que se mencionan comúnmente son:

- **Validez:** los datos miden lo que se pretende que midan.
- **Fiabilidad:** los datos se miden y recopilan sistemáticamente según las definiciones y metodologías estándar; los resultados son los mismos cuando se repiten las mediciones.
- **Exhaustividad:** se incluyen todos los elementos de información (según las definiciones y metodologías especificadas).
- **Precisión:** los datos están lo suficientemente detallados[9].
- **Integridad:** los datos están protegidos de sesgos o manipulaciones deliberados por motivos políticos o personales.
- **Puntualidad**: los datos están actualizados y la información está disponible a tiempo.

Es recomendable utilizar herramientas normalizadas de recolección de datos, que ya se hayan probado en situaciones reales, y mejorarlas en caso necesario para maximizar la calidad de los datos. Cuando sean necesarias adaptaciones al contexto local, o cuando deban desarrollarse herramientas de recolección de datos, es importante llevar a cabo primero una prueba piloto (y mejorar la herramienta) antes de utilizarla de manera generalizada. Recurrir a responsables de recolección de datos con experiencia, impartir capacitación a los responsables de la recolección sobre una tarea o herramienta específica, o supervisar la recolección de datos entre múltiples responsables de recolección también puede ayudar a reducir los sesgos (por ejemplo, dar pie de forma inadecuada a las respuestas durante las entrevistas) o errores (por ejemplo, no entender bien qué elementos del programa deben observarse) en los datos obtenidos.

[9] Una definición más específica de las mediciones cuantitativas sería: una medición se considera válida si es exacta y precisa. La exactitud se define como la desviación del valor «verdadero» y la precisión como la «dispersión». En otras palabras, la exactitud se refiere a cuánto se acerca la medición tomada al valor (verdadero) real; la precisión es el grado en el que las mediciones repetidas en condiciones invariables muestran los mismos resultados.

La recolección de datos no es necesariamente responsabilidad exclusiva de los evaluadores. Las ventajas de la «evaluación participativa» están bien documentadas y es un planteamiento que puede ir más allá de la recolección de datos para incluir al personal del programa, los participantes u otras partes interesadas en la configuración del programa de evaluación; la identificación de los principales resultados y la determinación de los factores de éxito; la contribución a la recolección de los datos; y el análisis e interpretación de los resultados (véase la Síntesis n.º 5 (Enfoques participativos)).

Aun cuando los datos se hayan recopilado utilizando procedimientos bien definidos y herramientas normalizadas, estos deben comprobarse para detectar posibles datos inexactos u omitidos. Esto es lo que se conoce como «depuración de datos», y también implica encontrar y corregir los errores que se producen durante la escritura, lectura, almacenamiento, transmisión o procesamiento de los datos informatizados.

La garantía de la calidad de los datos también se extiende a la garantía de un análisis de datos adecuado y a su presentación en el informe de evaluación para que las constataciones estén claras y las conclusiones puedan fundamentarse. Esto a menudo implica hacer accesibles los datos para que puedan ser verificados por otros o utilizados para fines adicionales como la sintetización de resultados de distintas evaluaciones (revisión sistemática, metaanálisis, revisión realista u otra metaevaluación).

4. CUESTIONES ESPECÍFICAS DE LA GARANTÍA DE CALIDAD EN LA RECOLECCIÓN DE DATOS

Como se ha señalado anteriormente, garantizar la calidad en la recolección de datos forma *parte* de una buena gestión de datos. En esta sección se examinan con más detalle algunas cuestiones especialmente pertinentes.

Explicitar la estrategia de muestreo

Una cuestión especialmente importante que debe atenderse y describirse íntegramente en el informe y en el plan de evaluación de impacto es la estrategia de muestreo utilizada. Se debe empezar por definir la «población de interés» de la que se extraerá la muestra. Este será el grupo o «unidades» de interés (por ejemplo, todas las parejas casadas) en la zona geográfica de interés (por ejemplo, un determinado pueblo) durante el tiempo de interés (por ejemplo, las parejas registradas desde 1995). El muestreo es el proceso de selección de unidades (por ejemplo, personas o grupos) de la población de interés para estudiarlas en profundidad con el objetivo de extraer conclusiones sobre el conjunto de la población (obsérvese que la extrapolabilidad de las constataciones de la evaluación dependerá de cómo se haga el muestreo). En cualquier muestra se da un error de muestreo. En otras palabras, ninguna muestra generará exactamente la misma información que si se incluyese a todas las personas de la población de interés en la recolección de datos. Por lo tanto, la información recabada a partir de la muestra se utilizará para realizar *estimaciones* sobre la población de interés. En el cuadro 3 se presenta un resumen de los tres grupos básicos de opciones de muestreo.

Cuadro 3. Resumen de las opciones de muestreo con métodos ilustrativos

Grupo de opciones de muestreo	Algunos métodos específicos	Riesgo de introducir sesgos
Probabilístico: Utilizar métodos aleatorios o cuasi aleatorios para seleccionar la muestra y utilizar posteriormente la extrapolabilidad estadística para extraer conclusiones sobre esa población.	Muestreo aleatorio simple Muestreo estratificado Muestreo por etapas múltiples Muestreo secuencial	Este grupo presenta normas específicas sobre la selección del marco de muestreo, el tamaño de la muestra y la gestión de la variación en la muestra
Intencional: Estudiar casos con abundante información de una población determinada para extraer conclusiones analíticas sobre la población. Las unidades se seleccionan en función de una o más características predeterminadas y el tamaño de la muestra puede ser tan solo de uno.	Confirmar y rechazar Caso crítico Muestreo de variación máxima Valor atípico Muestreo de bola de nieve Basado en teorías Caso típico	Este grupo fomenta la transparencia en la selección de casos y la triangulación, y busca rechazar las pruebas empíricas
De conveniencia: Estas opciones de muestreo utilizan a individuos que están disponibles o casos a medida que ocurren.	Fácilmente disponible Voluntarios	Este tipo presenta la menor fiabilidad, pero requiere menos tiempo, fondos y esfuerzos

Las distintas formas de muestreo introducirán distintos tipos de sesgo al evaluar los resultados de un programa o política. Como tal, es importante describir claramente los puntos fuertes y débiles de la estrategia de muestreo utilizada y tenerlos en cuenta, si es posible, en el análisis de datos o describir cómo afectan a las conclusiones extraídas sobre el programa o política sometidos a prueba.

Aunque se han mencionado tres grupos de opciones de muestreo, es mejor evitar el muestreo «de conveniencia», puesto que no permite una extrapolabilidad adecuada de las constataciones (es decir, realizar estimaciones sobre el total de la población a partir de las constataciones de la muestra estudiada). En el muestreo probabilístico, todas las unidades tienen la misma posibilidad o una posibilidad conocida de ser seleccionadas, lo que aumenta la probabilidad de que la información recabada sobre la muestra sea representativa de toda la población de interés, especialmente si se cumplen las condiciones sobre el tamaño de la muestra (véase la Síntesis n.º 7 (Ensayos controlados aleatorios)). El muestreo probabilístico permite una «extrapolabilidad estadística» en la que se da un cierto margen de error (normalmente, entre un 5% y un 10%) que define el nivel de riesgo aceptable de equivocación sobre los resultados del programa. El muestreo intencional no se basa en la

representación o la aleatoriedad, y se utiliza para seleccionar aquellas unidades de las que se puede aprender mucho sobre las cuestiones que son importantes para la evaluación (como los «informantes

clave»)[10]. Permite una «extrapolabilidad analítica» (realizar previsiones sobre la transferibilidad probable de las constataciones sobre la base de un análisis teórico de los factores que producen los resultados y el efecto del contexto).

Hacer frente a los abandonos y las negativas

Aunque un muestreo adecuado puede ayudar a reducir el «sesgo de selección», también puede aparecer un sesgo cuando haya tasas de abandono elevadas (o una «pérdida de seguimiento» sustancial): muchos sujetos abandonan el programa y ya no están disponibles para la recolección de datos. Del mismo modo, una baja tasa de respuesta cuando se pide a los sujetos que rellenen un cuestionario o un elevado número de sujetos que se niega a participar en la recolección de datos pueden generar un «sesgo de medición». Algunas formas de hacer frente a estos problemas son: realizar un esfuerzo especial por aumentar la participación; recabar información de varias fuentes, que posteriormente puede triangularse; y determinar en qué son diferentes los sujetos que participan de los que abandonan o se niegan a participar, y cómo se comparan los participantes con la población de interés (y, a su vez, cómo afecta esto a la interpretación y extrapolabilidad de los resultados). Una vez más, la transparencia y el debate en profundidad son una garantía.

5. CUESTIONES ESPECÍFICAS DE LA GARANTÍA DE CALIDAD EN EL ANÁLISIS DE DATOS

Al igual que con la recolección de datos, garantizar la calidad en el análisis de datos forma *parte* de una buena gestión de datos. En esta sección se examinan con más detalle algunas cuestiones especialmente pertinentes.

Cada fuente de datos tiene sus puntos fuertes y sus limitaciones, todos los cuales deben describirse explícitamente (en términos de calidad de los datos, véase más adelante) y entenderse claramente a fin de asegurar que las fuentes se utilicen e interpreten de manera adecuada (véase la sección 7). Por este motivo, si se integran intencionadamente distintos métodos de recolección de datos o se triangulan distintas fuentes de datos para responder a las preguntas clave de evaluación se conseguirá superar las carencias inherentes a cada fuente de datos de manera individual. Un enfoque de métodos mixtos también mejora la fiabilidad de las constataciones cuando la información de distintas fuentes de datos converge (coincide en la dirección de las constataciones) y pueden profundizar el entendimiento de un programa o política, sus efectos y su contexto. Aunque la combinación de métodos puede realizarse *post hoc* (es decir, después de haber recopilado los datos), es recomendable, y con frecuencia mucho más fácil, hacerlo al inicio (es decir, durante la planificación de la evaluación), para que la recolección de datos pueda orientarse específicamente a garantizar la disponibilidad de una combinación adecuada de pruebas empíricas justo en el momento de responder a las preguntas clave de evaluación. En el cuadro 4 se recogen los propósitos principales de la combinación de fuentes de datos.

[10] Taylor-Powell, Ellen, *Sampling*, University of Wisconsin, Milwaukee, 1998.

Cuadro 4. El valor de mezclar métodos: principales propósitos de la combinación de fuentes de datos

Enriquecer	Utilizar datos cualitativos para detectar problemas u obtener información sobre variables que no pueden obtenerse mediante enfoques cuantitativos
Examinar	Generar hipótesis a partir de los datos cualitativos que deben ponerse a prueba a través de datos cuantitativos (como identificar subgrupos que deben analizarse por separado en los datos cuantitativos, por ejemplo, para investigar el impacto diferencial)
Explicar	Utilizar datos cualitativos para entender los resultados no previstos de los datos cuantitativos
Triangular (confirmar o rechazar)	Verificar o rechazar los resultados de los datos cuantitativos utilizando datos cualitativos (o viceversa)

Analizar los datos para resumirlos y buscar patrones es una parte importante de toda evaluación. Las estrategias de análisis de datos y la forma de sintetizar los datos deben definirse en la fase de diseño de la evaluación, adecuarse al tipo de preguntas clave de evaluación y describirse íntegramente en el plan de evaluación.

En el caso de las preguntas clave de evaluación descriptivas, existe toda una serie de opciones de análisis (véase el cuadro 5). Las opciones pueden agruparse principalmente en dos categorías: opciones para datos cuantitativos (números) y opciones para datos cualitativos (por ejemplo, texto).

Cuadro 5. Opciones de análisis de datos numéricos y textuales

Análisis numérico Analizar datos numéricos como el costo, la frecuencia o las características físicas. Algunas opciones son:
• Correlación: técnica estadística para determinar el grado en que se relacionan dos o más variables. • Tabulaciones cruzadas: obtener una indicación de la frecuencia de dos variables (por ejemplo, género y frecuencia de la asistencia escolar) que ocurren al mismo tiempo. • Manejo de datos y texto: técnicas informáticas automatizadas que repasan grandes cantidades de texto o datos para encontrar nuevos patrones e información. • Técnicas exploratorias: echar un «primer vistazo» a un conjunto datos resumiendo sus principales características, a menudo utilizando métodos visuales. • Tablas de frecuencia: colocar valores de datos recopilados por orden ascendente de magnitud, junto con sus correspondientes frecuencias, para asegurar una imagen más clara del conjunto de datos. • Medidas de tendencia central: medida sintética que intenta describir un conjunto de datos con un único valor que representa el medio o centro de su distribución. • Medidas de dispersión: medida sintética que describe cómo se distribuyen los valores en torno al centro.

- Descriptivo multivariable: proporcionar resúmenes simples de (grandes cantidades de) información (o datos) con dos o más variables relacionadas.
- Inferencial no paramétrico: datos que son flexibles y no siguen una distribución normal.
- Inferencial paramétrico: llevado a cabo sobre datos que siguen determinados parámetros. Los datos serán normales (es decir, la distribución es paralela a la curva de campana); los números pueden sumarse,
- Resumen estadístico: proporcionar un resumen rápido de los datos, que resulta especialmente útil para comparar un proyecto con otro, antes y después.
- Análisis de series temporales: observar datos bien definidos obtenidos mediante mediciones repetidas en el tiempo

Análisis textual

Analizar palabras, orales o escritas, incluidas respuestas a cuestionarios, entrevistas y documentos. Algunas opciones son:

- Análisis de contenido: reducir grandes cantidades de contenido textual no estructurado a datos manejables relevantes para las preguntas de investigación (evaluación).
- Codificación temática: registrar o identificar pasajes de texto o imágenes relacionados por un tema o idea común, permitiendo la indexación del texto en categorías.
- Narraciones: construcción de narraciones coherentes de los cambios ocurridos respecto a una persona, una comunidad, un lugar o un programa o política.
- Cronogramas: una lista de acontecimientos importantes ordenados por orden cronológico.

En el caso de las preguntas clave de evaluación causales, básicamente hay tres enfoques de análisis de la atribución causal: 1) enfoques contrafácticos; 2) coherencia de las pruebas empíricas con la relación causal; y 3) descarte de alternativas (véase el cuadro 6). Lo ideal es combinar estos enfoques para determinar la causalidad (véase la Síntesis n.º 6 (Sinopsis: Estrategias de atribución causal)).

En el caso de las preguntas clave de evaluación, deben aplicarse rúbricas evaluativas específicas a los criterios de evaluación empleados (como los criterios del CAD de la OCDE) para sintetizar las pruebas empíricas y emitir juicios sobre el valor del programa o política (véase la Síntesis n.º 4 (Razonamiento evaluativo)).

Cuadro 6. Enfoques de análisis de datos para la atribución causal

Enfoques contrafácticos	Elaborar una estimación de lo que habría pasado en ausencia del programa o política; esto implica utilizar un grupo de control o un grupo de comparación.
Coherencia de las pruebas empíricas con la relación causal	Identificar patrones que serían coherentes con una relación causal y después tratar de confirmar o rechazar las pruebas.
Descarte de alternativas explicativas	Definir posibles explicaciones causales alternativas y después buscar información para determinar si pueden descartarse.

Una poderosa herramienta para el análisis y la comunicación de las constataciones de la evaluación es la visualización de datos. Se trata del proceso de representar los datos gráficamente para identificar tendencias y patrones que de otro modo no estarían claros o serían difíciles de discernir. Existen muchas opciones que se deben tener en cuenta[11], y la selección del tipo de visualización dependerá de la naturaleza de los datos (por ejemplo, relacionales, comparativos, cronológicos). Además, el tipo de visualización utilizada para facilitar el análisis de datos (por ejemplo, picos en el movimiento del sitio web relacionados con una campaña en las redes sociales) no tiene que ser necesariamente el mejor tipo a efectos de comunicación. A la hora de comunicar datos, puede resultar necesario simplificar la visualización y hacer hincapié en las esferas principales para atraer la atención de los lectores y las partes interesadas. Las conclusiones de la evaluación de impacto deben anclarse firmemente en los datos y presentarse de una manera transparente y verificable por los lectores del informe de evaluación.

6. CUESTIONES ÉTICAS Y LIMITACIONES PRÁCTICAS

Cuestiones éticas

Todas las evaluaciones deben seguir las d*irectrices éticas* de *evaluación*[12] *del* Grupo de *Evaluación* de las Naciones Unidas (UNEG), a las que se adhiere UNICEF. Estas directrices establecen tanto un código de conducta para evaluadores como obligaciones éticas para los participantes en la evaluación en relación con la recolección y el almacenamiento de datos (véase asimismo la Síntesis n.º 1 (Sinopsis de la evaluación de impacto)). Estas incluyen: respeto de la dignidad y la diversidad; derechos; confidencialidad; y prevención del daño. Estas obligaciones deben cumplirse explícitamente en el diseño de la evaluación y durante todo el proceso de evaluación (véase el sitio web *Ethical Research Involving Children*)[13].

En la mayoría de los casos, la aprobación oficial de la forma de recolección, gestión y uso de los datos en la evaluación requerirá una junta de examen institucional. El objetivo es asegurar que estas prácticas cumplan las obligaciones éticas para todos aquellos de quienes se recaban datos.

La evaluación de la iniciativa de escuelas amigas de la infancia en Moldava[14] brindó algunos buenos ejemplos de estrategias para abordar las cuestiones éticas durante la recolección y análisis de datos:

- Durante la fase preparatoria, se consultó a las partes interesadas (UNICEF y los asociados en la ejecución) y posteriormente se pusieron a prueba las herramientas de recolección de datos en un pequeño grupo de profesores, niños y padres que participaron voluntariamente tras recibir información sobre el proceso de prueba. Esto se hizo para detectar cualquier cosa que pudiese interpretarse que podría causar molestia, crear situaciones raras o afectar a los derechos de los participantes.
- Durante la fase de recolección de datos, se aseguró a los encuestados que todos los datos se tratarían de forma confidencial y que se utilizarían exclusivamente para fines de evaluación.

[11] Para obtener más información sobre los métodos específicos de visualización de datos, véase «Visualise Data», sitio web de BetterEvaluation, http://betterevaluation.org/plan/describe/visualize_data.

[12] Grupo de Evaluación de las Naciones Unidas, *Ethical Guidelines for Evaluation*, UNEG, Nueva York, 2007.

[13] El sitio web www.childethics.com es un depósito de información con base empírica, recursos y enlaces a artículos de revistas compilado por el proyecto Ethical Research Involving Children para orientar y mejorar la investigación con niños y proporcionar una plataforma de reflexión crítica y diálogo.

[14] Velea, Simona y CReDO (Centro de Recursos de Derechos Humanos), *Child-Friendly Schools, External Evaluation Report of the Child-Friendly School Initiative (2007–2011), Republic of Moldova*, Ministerio de Educación de la República de Moldova/UNICEF, 2012.

Esto también se dejó patente en el texto introductorio de los cuestionarios. Los cuestionarios no contenían ningún campo de datos personales que pudiese dar lugar a la identificación de los encuestados.

- Durante los debates en grupo, no se registraron los nombres de los participantes ni se grabaron las sesiones. La participación se basó en el consentimiento informado: los participantes recibieron información sobre el enfoque de evaluación, su papel en la evaluación y las ventajas para ellos a nivel personal, tanto directas (como experiencia de aprendizaje) como indirectas (al contribuir a entender mejor el sistema educativo).
- Durante las entrevistas, las interacciones entre los propios participantes y entre los participantes y los evaluadores se basaron en el respeto mutuo y la confianza.

La misma evaluación presenta también buenos ejemplos de salvaguardias para garantizar la confidencialidad de los datos:

- Durante el procesamiento de los datos, no se hizo referencia precisa a escuelas o declaraciones realizadas por determinados alumnos, profesores o padres.
- Los datos de los cuestionarios se sometieron a tratamiento estadístico y los datos cualitativos se revisaron en su conjunto, sin ninguna información personal. (Sin embargo, se realizaron observaciones específicas sobre determinadas escuelas identificadas).

Limitaciones prácticas

Algunas limitaciones prácticas comunes en la recolección de datos pueden ser:

- Los informantes clave, los participantes y las ubicaciones pertinentes pueden no estar disponibles/localizables si:
  - no hay información completa sobre las ubicaciones del programa o los participantes;
  - no se les contacta con suficiente antelación antes o durante el periodo de recolección de datos;
  - la evaluación tiene lugar después de que se haya completado el programa.
- Las fuentes de datos pertinentes existentes (como los indicadores de los informes de avances de los asociados en la ejecución; los datos de encuestas) son limitadas en cuanto a tiempo (por ejemplo, no hay base de referencia o series temporales disponibles), o contenido (por ejemplo, faltan indicadores importantes) o calidad (hay seria preocupación respecto a la calidad de los datos).
- A menudo existe información escasa sobre la calidad de los servicios prestados por los distintos ejecutores en el caso de un programa a gran escala.

Si los recursos y el marco temporal de la evaluación lo permiten, algunas de estas limitaciones de la recolección de datos pueden corregirse (por ejemplo, ampliando el periodo o alcance de la recolección de datos), pero es probable que las conclusiones tengan que extraerse sobre la base de datos imperfectos. Por lo tanto, es importante ser transparente respecto a las limitaciones y describir cómo pueden haber afectado a las constataciones, las conclusiones y las recomendaciones de la evaluación.

7. EJEMPLO DE BUENAS PRÁCTICAS

La evaluación de la iniciativa de escuelas amigas de la infancia[15] mencionada anteriormente brindaba algunos buenos ejemplos de una reflexión y ejecución bien desarrolladas respecto a las opciones de recolección y análisis de datos.

Determinar las necesidades de datos desentrañando adecuadamente los principales conceptos:

El *paidocentrismo* se definió como enfoque metodológico que podría «desentrañarse» en los siguientes componentes específicos:

- conocer los estilos de aprendizaje, problemas y aspiraciones de los alumnos;
- incluir y valorar a todos los alumnos, con independencia de su etnia, estatus social, discapacidad, religión o género;
- cambiar el foco de atención de la enseñanza al aprendizaje, y adaptar todo el proceso de enseñanza a los niños;
- hacer partícipes a los alumnos en la fijación de objetivos, el aprendizaje de nuevos métodos de estudio y la evaluación del rendimiento (establecimiento y comunicación de criterios) y la autoevaluación;
- garantizar que se enseñe a los alumnos cómo aprender y que lo que aprendan sea relevante para su vida cotidiana, integrando en la clase las experiencias de aprendizaje positivas que estos adquieran fuera de la escuela;
- diversificar los casos de situaciones de aprendizaje y funciones de los docentes: transmitir información, facilitar su propia búsqueda de información, orientación, moderación;
- cultivar una buena comunicación con los padres y fomentar su participación en la educación del niño.

El desentrañamiento de este concepto fundamental llevó a los evaluadores a constatar la necesidad de una serie de datos y, con ello, una serie de fuentes de datos. Los cuestionarios eran una de las formas de recopilar datos pertinentes. La complejidad del paidocentrismo no permitía utilizar únicamente unos indicadores simples para evaluar los logros; por el contrario, esta complejidad requería aunar varias mediciones (obtenidas a través de una serie de preguntas específicas en el instrumento de recolección de datos; véanse las notas al pie en el cuadro 7, extraído del informe de evaluación) relacionadas con las principales esferas de resultados (como condiciones escolares favorables para la igualdad de trato; véanse los elementos enumerados en el cuadro 7).

[15] Velea, Simona y CReDO (Centro de Recursos de Derechos Humanos), *Child-Friendly Schools, External Evaluation Report of the Child-Friendly School Initiative (2007–2011), Republic of Moldova*, Ministerio de Educación de la República de Moldova/UNICEF, 2012, págs. 47–48.

Cuadro 7. Opiniones de los alumnos sobre varios indicadores relacionados con la enseñanza. (Medias; comparación en columnas, puntuación más baja = mejor situación del indicador determinado). Extraído del informe de evaluación sobre escuelas amigas de la infancia en Moldava, cuadro 8, pág. 47

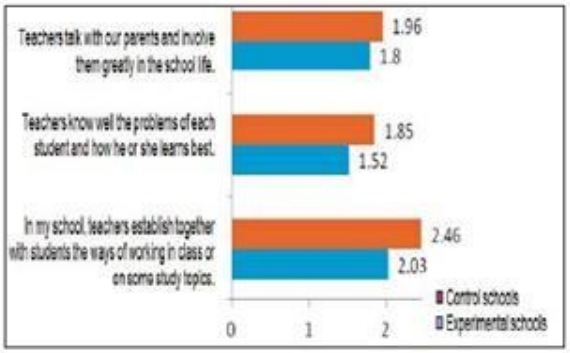
Items	Experimental schools	Control schools
1. Encouraging thinking in students, and self-assessment of their own learning styles[31]	3.77	4.33
2. Setting hours and homework assignments together with the students[32]	15.48	16.89
3. Working with children who have disabilities, and those with learning difficulties [33]	7.74	8.51
4. Working with children of ethnic minorities [34]	6.7	7.33
5. Treating other teachers and all students with equal respect [35]	2.79	3.05
6. Creating school conditions favourable to equal treatment [36]	6.62	7.3
7. Encouraging communication and dialogue with other teachers and with parents and students [37]	25.02	26.24
8. Management of knowledge (correct learning, use of information technologies, etc.). [38]	5.1	5.31

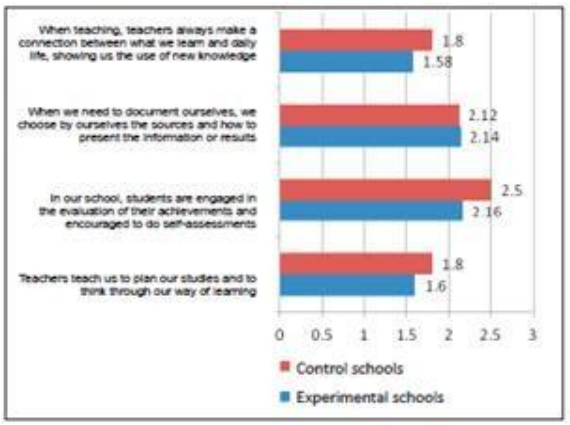
31 Student questionnaire, Item 1, questions 10.4, 11.4, 11.12.
32 Ibid., Item 2, questions 9.1, 9.4, 9.3, 9.7, 9.8, 11.1, 11.7, 12.3, 12.4.
33 Ibid., Item 3, questions 11.3, 10.6, 11.12, 11.17.
34 Ibid., Item 4, questions 10.4, 11.4, 11.12.
35 Ibid., Item 5, questions 10.7, 11.7.
36 Ibid., Item 6, questions: 11.9, 11.11, 11.12, 11.17.
37 Ibid., Item 7, questions: 9.1–9.9, 11.1, 11.2, 11.11.
38 Ibid., Item 8, questions: 12.5, 12.1, 11.8.

Uso de distintos métodos de recolección de datos con la intención de triangular diferentes perspectivas y ofrecer una presentación clara de los resultados:

La evaluación no solo recopiló datos sobre el paidocentrismo de los docentes, sino también datos sobre las prácticas de enseñanza de los beneficiarios finales: los alumnos. Si bien esta puede parecer una fuente de datos «obvia» para algunos evaluadores, «preguntar a los participantes del programa» no se incluye necesariamente de manera sistemática en todas las evaluaciones de impacto, y cuando se incluye, a menudo se trata de forma «simbólica» (véase también la Síntesis n.º 5 (Enfoques participativos)). El estudio sobre las escuelas amigas de la infancia presentaba datos complejos de una manera muy accesible mediante una buena visualización de datos (véanse los gráficos 15 y 16, una vez más extraídos del informe de evaluación).

Gráfico 1. Opiniones de los alumnos respecto a las prácticas de enseñanza relevantes para el paidocentrismo (gráfico izquierdo) y respecto a las prácticas de enseñanza (gráfico derecho). (Clasificación de promedios; 0 = mucho,3 = poco). Extraído del informe de evaluación sobre escuelas amigas de la infancia en Moldava, gráfico 15 (izquierda) y gráfico 16 (derecha), pág. 48.





Valorar los impactos diferenciales del programa

Además de informar sobre los «promedios», los evaluadores también analizaron los impactos diferenciales del programa. Estos impactos son esenciales para entender mejor lo que funcionó, con quién y en qué condiciones.

Así por ejemplo, los evaluadores señalaron resultados positivos en el comportamiento de los alumnos, lo que se reflejó en una mayor participación y más frecuente en las actividades escolares, especialmente de los niños con discapacidad o de grupos vulnerables (definidos en el estudio). Asimismo, un mayor número de niños obtuvo mejores resultados académicos, con un progreso notable entre los alumnos con bajo rendimiento escolar (definidos en el estudio).

8. EJEMPLOS DE POSIBLES PROBLEMAS

Un posible problema en la recolección y análisis de datos puede referirse a una *mala elección de los métodos* así como a una *mala aplicación de los métodos*. A continuación se presentan algunos ejemplos.

Mala elección de los métodos:

1. Elegir los mismos métodos antiguos independientemente de su adecuación a la evaluación específica.

   Como se desprende claramente de esta síntesis, existe toda una serie de métodos de recolección y análisis de datos para su uso en evaluaciones de impacto. La combinación más adecuada de métodos depende de la naturaleza de lo que se está evaluando y el contexto de la evaluación. En otras palabras, cuáles son las preguntas de evaluación que deben responderse; cuándo deben estar disponibles las respuestas para apoyar la toma de decisiones de distintas partes interesadas; de qué recursos se dispone para la evaluación; y qué otras condiciones pueden limitar la evaluación (por ejemplo, el entorno político).

2. Elegir métodos que no están específicamente relacionados con las respuestas a las preguntas clave de evaluación.

   Construir una matriz de evaluación (véase el cuadro 1) puede ayudar a asegurar que se engloben todas las preguntas clave de evaluación.

3. Elegir solo un método para responder a una pregunta clave de evaluación.

   No se trata únicamente de emplear varios métodos, sino de elegirlos de manera que se complementen, reduciendo así los sesgos y garantizando la triangulación. Es decir, el punto fuerte de cualquier método individual seleccionado corregirá (al menos algunos de) los puntos débiles de otro método seleccionado, además de incluirse distintas perspectivas sobre la intervención en términos de las principales partes interesadas.

Mala aplicación de los métodos:

1. Mala ejecución del muestreo.

   Por ejemplo, utilizar una muestra de conveniencia o tener una tasa de respuesta invariablemente baja sin comprobar sistemáticamente y mostrar claramente en qué se pueden haber diferenciado los encuestados de los no encuestados.

   El principal objetivo del muestreo es que los datos de la muestra puedan extrapolarse adecuadamente, desde un punto de vista estadístico o analítico. Una infrarrepresentación importante en la muestra seleccionada puede dar lugar a conclusiones falsas o incorrectas sobre el programa o política objeto de investigación, por ejemplo cuando hay una infrarrepresentación importante de determinadas edades, paridad de género u otras variables demográficas pertinentes, o de determinadas características que se cree que están relacionadas con los resultados o impactos de interés (como la exclusión de los alumnos con un bajo rendimiento en el caso del análisis del impacto en los porcentajes de aprobación de exámenes).

2. Mala calidad de la recolección de datos.

   Por ejemplo, cuestionarios mal elaborados; no dar a las personas la oportunidad de hablar realmente durante las entrevistas; falta de buenos traductores; y no registrar los datos con precisión. Algunas formas de evitar o minimizar estos errores frecuentes y comunes son: contar con una opinión experta a la hora de diseñar o adaptar los instrumentos de recolección de datos; garantizar una recolección de datos de buena calidad mediante la selección, capacitación y supervisión de responsables de la recolección de datos (entre otras cosas asegurando la paridad de género en los equipos de entrevista); proporcionar espacios privados para las entrevistas, a fin de que las personas se sientan más cómodas para expresar sus opiniones, o hacer uso de la tecnología para proporcionar más «privacidad» en las respuestas y evitar obtener respuestas «socialmente deseables»; y recurrir a traductores certificados y con experiencia.

3. Mala calidad del análisis de datos.

   Por ejemplo, extraer conclusiones que no están respaldadas por los datos o centrarse solo en el efecto medio del programa o política. Algunas formas de apoyar un análisis de datos de buena calidad son: asegurar que se cuestionan las conclusiones «iniciales» manteniendo sesiones de comentarios y sugerencias con las principales partes interesadas o buscar intencionadamente rechazar las pruebas empíricas; buscar patrones; y asegurar que se aborden cuestiones «transversales» como la equidad.

4. Mala calidad de la presentación de informes y resultados.

   Por ejemplo, plantear la presentación de informes según los distintos instrumentos de recolección de datos (ir repasando los elementos del cuestionario uno por uno) en vez de responder a las preguntas clave de evaluación, o utilizar gráficos complicados que requieren una extensa explicación para entenderlos. Una buena visualización de datos es una herramienta esencial para comunicación los mensajes principales; por ejemplo, crear un «mapa» visual de las ideas principales para los datos textuales y añadir citas específicas en recuadros para ilustrar constataciones resumidas, o presentar las principales constataciones a través de hojas de cálculo cuidadosamente seleccionadas (es decir, hay muchas opciones entre las que elegir una vez que se decide ir más allá de las técnicas habituales).

9. LECTURAS Y ENLACES CLAVE

Bamberger, Michael, «Introduction to Mixed Methods in Impact Evaluation», *Guidance Note* No. 3, InterAction, Washington, D.C., agosto de 2012. Véase http://www.interaction.org/impact-evaluation-notes.

BetterEvaluation, «Analyze Data», sitio web de BetterEvaluation, http://betterevaluation.org/plan/describe/look_for_patterns.

BetterEvaluation, «Collect and/or Retrieve Data», sitio web de BetterEvaluation, http://betterevaluation.org/plan/describe/collect_retrieve_data.

BetterEvaluation, «Combine Qualitative and Quantitative Data», sitio web de BetterEvaluation, http://betterevaluation.org/plan/describe/combining_qualitative_and_quantitative_data.

BetterEvaluation, «Manage Data», sitio web de BetterEvaluation, http://betterevaluation.org/plan/describe/manage_data.

Evergreen, Stephanie, D.H., *Presenting Data Effectively, Communicating Your Findings for Maximum Impact*, Sage, Thousand Oaks, 2013.

Measure Evaluation, «Data Quality Assurance», sitio web de Measure Evaluation, http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools/data-quality-assurance-tools.

Patton, Michael Quinn, *Qualitative Research & Evaluation Methods, tercera* edición, Sage, Thousand Oaks, 2001.

Perrin, Burt, «Linking Monitoring and Evaluation to Impact Evaluation», *Guidance Note* No. 2, InterAction, Washington, D.C., abril de 2012. Véase http://www.interaction.org/document/guidance-note-2-linking-monitoring-and-evaluation-impact-evaluation.

GLOSARIO

Actividad	*Acciones emprendidas o procesos mediante los que se movilizan insumos para generar productos determinados. Por ejemplo, sesiones de asesoramiento que se adhieren a unas normas de calidad.*
Datos de referencia	*Datos recopilados inicialmente que sirven de base para la comparación con los datos obtenidos en una fase posterior. Por ejemplo, los datos recopilados antes de una intervención se comparan posteriormente con los datos recopilados después de ejecutar la intervención.*
Depuración de datos	*Proceso de comprobación de los datos para ver si se ajustan a los estándares, coherencia interna, integridad referencial y ámbito válido, y de sustitución/reparación de los datos incorrectos por datos correctos. «Depurar» un archivo de datos consiste en comprobar si hay códigos rebeldes y respuestas incongruentes; verificar que el archivo tiene el número correcto y esperado de registros, casos y tarjetas o registros por caso; y corregir los errores detectados.*
Diseño cuasi experimental	*Diseño de investigación/evaluación en el que los participantes no se asignan de forma aleatoria a las condiciones del tratamiento, sino que los grupos de comparación se construyen por medios estadísticos. Difiere del experimento controlado (clásico) en que no hay una asignación aleatoria del tratamiento/intervención.*
Error de muestreo	*Grado en el que los resultados de la muestra se desvían de los que se habrían obtenido con toda la población debido al error aleatorio en la selección de participantes y la correspondiente reducción de la fiabilidad.*
Evaluación participativa	*Enfoque de evaluación que involucra a las partes interesadas, especialmente a los participantes en un programa o a los afectados por una determinada política, en algunos aspectos del proceso de evaluación, entre ellos, el diseño, la aplicación de la recolección y el análisis de datos o la supervisión del trabajo de los expertos técnicos.*
Impacto	*Efectos de largo plazo positivos y negativos, primarios y secundarios, producidos directa o indirectamente por la intervención, intencionalmente o no.* [16]
Indicador	*Medida verificable que ha sido seleccionada por la dirección de un programa o política para tomar decisiones sobre dicho programa o política. Por ejemplo, la proporción de estudiantes que aprueban un examen estandarizado.*
Informante clave	*Participante en la investigación/evaluación que es probable que esté bien informado sobre una cuestión y dispuesto a responder sin sesgos.*
Insumo	*Recursos financieros, humanos y materiales empleados en un programa o una política. Por ejemplo, los materiales de capacitación elaborados.*
Magnitud del efecto	*Término estadístico de la magnitud de la relación entre dos variables.*

[16] CAD-OCDE, *Glosario de los principales términos sobre evaluación y gestión basada en resultados*, OCDE, París, 2010. Véase http://www.oecd.org/development/peer-reviews/2754804.pdf.

Metaanálisis	*Proceso o técnica de sintetización de los resultados de la investigación de dos o más estudios utilizando varios métodos estadísticos para recuperar, seleccionar y combinar los resultados de anteriores estudios que abordan la misma cuestión.*
Metaevaluación	*Revisión/valoración de la calidad de una evaluación. El término se utiliza también para describir la actividad de sintetizar constataciones de una serie de evaluaciones con el fin de valorar su calidad o medir el desempeño de los evaluadores. También se conoce como «síntesis de evaluación».*
Muestreo aleatorio simple	*Tipo específico de muestreo probabilístico en el que los sujetos se seleccionan de una lista de población utilizando una tabla de números aleatorios o un generador de números aleatorios. (Una muestra aleatoria requiere una lista de miembros de la población en la que pueda asignarse a cada miembro un número diferente.) La asignación al azar de los sujetos a distintos tratamientos, intervenciones o condiciones en vez de forma sistemática. La asignación aleatoria de los sujetos aumenta la probabilidad de que las diferencias observadas entre los grupos de sujetos sean fruto de la intervención experimental.*
Muestreo de bola de nieve	*Muestreo no probabilístico que se crea recurriendo a los miembros del grupo de interés para que señalen a otros miembros del grupo (por ejemplo, pidiendo sugerencias a un sujeto al término de una entrevista sobre a quién más se podría entrevistar).*
Muestreo de variación máxima	*Utilizado en la investigación cualitativa para seleccionar deliberadamente una amplia variedad de casos.*
Muestreo estratificado	*Muestreo probabilístico que se determina dividiendo la población en grupos o estratos definidos por la presencia de determinadas características y posteriormente realizando un muestreo aleatorio de cada estrato. Es una buena forma de asegurar que una muestra de estudiantes sea racialmente diversa (por ejemplo).*
Muestreo por etapas múltiples	*Muestreo probabilístico que conlleva varias etapas (y frecuentemente una etapa de muestreo por conglomerados), por ejemplo, seleccionar conglomerados de una población y posteriormente seleccionar personas de cada conglomerado.*
Muestreo secuencial	*Método de muestreo que puede ser probabilístico y no probabilístico. Como método de muestreo probabilístico, el muestreo secuencial requiere un marco ordenado de unidades de muestreo desde el que se selecciona la muestra utilizando una probabilidad o previsión explícitamente indicada, por ejemplo, seleccionando uno de cada diez casos de una lista.* *Como técnica de muestreo no probabilístico, el muestreo secuencial empieza por el muestreo de una única observación o un grupo de observaciones que posteriormente se ponen a prueba con respecto a la hipótesis. Si no se rechaza la hipótesis nula, se somete a muestreo otra observación o grupo de observaciones y la hipótesis se pone a prueba otra vez, etc.*
Producto	*Efectos inmediatos de las actividades de un programa o política, o los productos o entregables directos de estas actividades. Por ejemplo, el número de vacunas administradas.*

Prueba piloto	*Estudio a pequeña escala llevado a cabo para comprobar la fiabilidad de una herramienta de recolección de datos.*
Resultado	*«Algo que ocurre» a las personas, las comunidades, el medio ambiente o la economía, que al menos parcialmente se debe al programa o a la política que se está evaluando. Los resultados pueden ser cambios positivos o negativos, previstos o no previstos, que se producen, o cambios que se evita que se produzcan. En el contexto de evaluación, el uso del término «resultado» implica que el programa/política contribuyó al cambio, ayudó a provocarlo o lo catalizó (lo aceleró).*
Revisión realista	*Tipo de síntesis de investigación explicativa e iterativa que pone a prueba y crea teorías de alcance intermedio. Facilita un análisis exploratorio para responder qué funciona, con quién, en qué circunstancias, por qué y cómo.*
Revisión sistemática	*Método de investigación que ofrece un marco para extraer y evaluar de forma crítica los resultados derivados de las pruebas empíricas existentes sobre una cuestión o tema específico. Las pruebas empíricas se seleccionan de acuerdo con criterios claros y la revisión se lleva a cabo a través de un protocolo estandarizado. Una vez que termina la compilación de toda la literatura pertinente, se valora la calidad de los estudios e informes utilizando criterios claramente definidos, y las conclusiones se extraen únicamente a partir de los datos que cumplen las normas de calidad.*
Teoría del cambio	*Explica cómo se entiende que las actividades produzcan una serie de resultados que contribuyen a lograr los impactos finales previstos. Puede elaborarse para cualquier nivel de intervención: un acontecimiento, un proyecto, un programa, una política, una estrategia o una organización.*
Valor atípico	*Observación en un conjunto de datos cuyo valor está alejado de los otros en el conjunto de datos. Es un valor inusualmente grande o pequeño en comparación con los demás.*