

Guía para el uso y la actualización del Banco de Evidencia

CONTENIDO

| | |
|---|-----------|
| 1. INTRODUCCIÓN | 2 |
| A. ¿POR QUÉ REVISAR EVIDENCIA? | 2 |
| B. ¿POR QUÉ ELABORAR UN BANCO DE EVIDENCIAS? | 3 |
| 2. ¿QUÉ TIPO DE EVIDENCIA SE INCLUYE EN EL BANCO DE EVIDENCIA? | 4 |
| 3. ¿CÓMO ACTUALIZAR EL BANCO? | 6 |
| A. PASO 1: ¿DÓNDE BUSCAR EVIDENCIA? | 6 |
| B. PASO 2: ¿CÓMO BUSCAR EVIDENCIA? | 7 |
| C. PASO 3: ¿CÓMO IDENTIFICAR LOS ESTUDIOS QUE SON DE UTILIDAD? | 7 |
| D. PASO 4: ¿CÓMO COMPLETAR EL BANCO DE EVIDENCIAS? | 8 |
| 4. ¿CÓMO HACER USO DE LOS ESTUDIOS DEL BANCO DE EVIDENCIA? | 9 |
| REFERENCIAS | 12 |
| ANEXO 2: METODOLOGÍAS DE EVALUACIÓN | 13 |

1. Introducción

En el marco del Convenio Tripartito de Cooperación Interinstitucional entre el Ministerio de la Mujer y Poblaciones Vulnerables (MIMP, Innovations for Poverty Action (IPA) y Abdul Latif Jameel Poverty Action Lab (J-PAL), se elaboró un Banco de Evidencia sobre las temáticas de i) Atención de víctimas de violencia doméstica, ii) Prevención de la violencia doméstica y iii) Fortalecimiento de autonomía de las mujeres. **La presente Guía busca ser una herramienta para facilitar el entendimiento de los contenidos del Banco de Evidencia y orientar a los funcionarios encargados de su actualización.** En la sección introductoria explicamos la importancia de revisar evidencia y cuál es la utilidad de contar con un Banco de Evidencia. En la segunda sección presentamos los tipos de evidencia que se recogen en el banco y se explica la clasificación de evidencias utilizada por el Ministerio de Economía y Finanzas (MEF). En la tercera sección explicamos las tareas que implica actualizar el Banco de Evidencia. Finalmente, en la cuarta sección se presentan algunas consideraciones para el uso de los contenidos del Banco de Evidencia.

a. ¿Por qué revisar evidencia?

En el proceso de diseño de un instrumento, programa o iniciativa es importante tomar en cuenta la evidencia, nacional e internacional, que existe sobre diseños similares al que estamos considerando. Nuestro diseño puede parecer muy bueno, pero podemos estar olvidando algunos aspectos relevantes que podrían hacerlo fracasar, ya sea en términos de rentabilidad, como de un impacto social o económico distinto del pronosticado. En este sentido, al diseñar, es importante incorporar las lecciones que nos ofrece la literatura sobre cuáles han sido los resultados de iniciativas similares desarrolladas en contextos parecidos al nuestro. Esta evidencia puede agruparse en las siguientes categorías:

- **Evidencia entregada por la Experiencia:** basada en la experiencia profesional, el entendimiento del problema, vivencias, habilidades, etc. tanto de uno mismo como de los demás, conocida como conocimiento intuitivo o tácito. Este tipo de evidencia puede ser clave para conocer las necesidades de los beneficiarios de una intervención. También puede ser clave para anticipar aspectos del diseño de la intervención que pueden mejorar la implementación. Sin embargo, no nos dice exactamente qué intervención será mejor.
- **Evidencia Descriptiva:** basada en la información que podemos recolectar sobre nuestros potenciales beneficiarios (sus preferencias, necesidades, limitaciones), su comunidad, el marco regulatorio del país, nuestra competencia, el contexto socio-político, etc. Este tipo de evidencia describe las características de una muestra apuntando a generalizar a la población objetivo. Esta evidencia puede encontrarse en estudios publicados sobre el tema, censos, encuestas o entrevistas. La evidencia descriptiva es útil para conocer las necesidades de los beneficiarios y ayuda a mejorar el diseño de la intervención por lo que es importante incluirla en una evaluación de necesidades o diagnóstico, así como en una evaluación de procesos. Sin embargo, no nos permite tomar una decisión sobre qué intervención sería mejor.

- **Evidencia Causal:** trata de **atribuir causas a efectos** y **medir la magnitud** de dichos efectos, es decir el impacto. Esta evidencia se obtiene de evaluaciones de impacto experimentales, cuasi-experimentales, y de revisiones sistemáticas y meta-análisis.

Así pues, existen varios tipos de evidencia y diferentes formas de recogerla, entre las que se encuentran diversas herramientas de análisis cualitativo y cuantitativo, el primero enfocado en entender en profundidad las razones del comportamiento humano, y el segundo introduce herramientas estadísticas al mismo problema. Todas estas fuentes son igual de útiles, y debieran ser incluidas en nuestro diseño. Sin embargo, esta guía busca profundizar en el **uso y selección de evidencia causal**.

b. ¿Por qué elaborar un Banco de Evidencias?

El Banco de Evidencias es un instrumento que sirve para sistematizar y ordenar información de una serie de estudios sobre temas de interés para sus posibles usuarios. Este documento responde a la misma lógica que el repositorio de evidencias contemplado en la Directiva General n°010-2016-MIMP “Normas para el seguimiento y evaluación en el Ministerio de la Mujer y Poblaciones Vulnerables-MIMP” (MIMP 2016) y busca alimentarlo.

Este Banco sirve un doble propósito con respecto a las prácticas de uso y generación de evidencia en el MIMP. **En primer lugar, el Banco de Evidencia facilita la identificación de estudios que pueden servir para informar el diseño de los programas del MIMP.** El contenido del banco puede inspirar nuevas intervenciones basadas en programas aplicados en otros contextos o advertir a los implementadores sobre los riesgos de transponer un modelo que no ha funcionado. Esta información es igualmente útil para informar programas todavía en fase de diseño, programas que estén siendo implementados, pero a los que se les quiere hacer ajustes o incluso para los que se ha pensado en un rediseño. Por último, las lecciones aprendidas pueden contener información útil para llevar a escala un programa.

En segundo lugar, **el Banco de Evidencia puede servir como insumo para el desarrollo del sustento técnico de las intervenciones del Programa Presupuestal 080 *Lucha contra la Violencia Familiar*** que estén relacionadas a la Atención de víctimas, Prevención de la violencia y Fortalecimiento de la Autonomía. Específicamente, los estudios recogidos en este Banco de Evidencia pueden ser una fuente de evidencia que sirva para completar las siguientes secciones del Anexo 2 del Programa Presupuestal 080:

- 1) “Vinculación del PP con su Resultado Final” (MEF 2016, 108) en la Sección 3.8 del Anexo 2;
- 2) “Análisis de alternativas” (MEF 2016, 108) en la Sección 3.3 del Anexo 2.

En los siguientes apartados vamos a definir de manera precisa i) qué es lo que consideramos evidencia, ii) cómo completar y actualizar el banco, y vamos a dar las pautas sobre iii) cómo hacer uso de esta evidencia.

2. ¿Qué tipo de evidencia se incluye en el Banco de Evidencia?

El principal criterio para incluir estudios en el Banco de Evidencia es la metodología que utilizan. En el presente Banco de evidencia se busca incluir tipos de estudios que utilicen metodologías que permitan argumentar causalidad: **evaluaciones aleatorias y evaluaciones cuasi-experimentales.** Asimismo, incluimos documentos que sistematizan resultados de diversos estudios como las **revisiones sistemáticas y los meta-análisis.**

Las evaluaciones experimentales son aquellas donde se comparan los resultados entre un grupo de tratamiento y un grupo de control que fueron conformados aleatoriamente. Al aleatorizar se controla el sesgo de selección, lo que asegura que el grupo de control y de tratamiento tienen, en promedio, las mismas características observables y no-observables. Por ello, el grupo control representa un contrafactual válido (la situación hipotética en la que el grupo de tratamiento no recibe el tratamiento). Por otro lado, las metodologías cuasi-experimentales no asignan el tratamiento de manera aleatoria; sin embargo, intentan generar grupos de control válidos de distintas maneras. Para una breve explicación de algunas metodologías experimentales y cuasi-experimentales revisar **Anexo 1.**

Tanto los meta-análisis como las revisiones sistemáticas consisten en la revisión de varios estudios sobre un mismo tema para extraer tendencias generales. La revisión sistemática resume y analiza la evidencia respecto de una pregunta específica en forma estructurada, explícita y sistemática. En general se presenta la evidencia en forma descriptiva, sin análisis estadístico. Por su parte el meta-análisis hace uso de técnicas estadísticas para combinar cuantitativamente los resultados de un grupo de estudios. En ambos casos, es importante verificar los criterios de inclusión de los estudios analizados para asegurarse de su rigurosidad pues de esto depende la calidad de las revisiones sistemáticas y los meta-análisis.

Clasificación de evidencias según el Ministerio de Economía y Finanzas (MEF)

El MEF ha establecido una jerarquización de metodologías para clasificar la evidencia que se presenta en el Anexo 2 como sustento de las intervenciones de los Programas Presupuestales. Hemos utilizado esta clasificación para identificar los tipos de estudios incluidos en el Banco de Evidencia.

Cuadro N° 1: Jerarquización de Metodologías según el MEF.

| Nivel | Tipo |
|----------------|---------------------------|
| <u>Nivel A</u> | A.1. Meta análisis |
| | A.2. Revisión sistemática |

| | |
|---|--|
| | A.3. Estudios experimentales |
| Nivel B: Estudios que utilizan métodos cuasi-experimentales y otras metodologías rigurosas de control de sesgo | B.1. Regresión discontinua |
| | B.2. Variables instrumentales |
| | B.3. Estudios con data longitudinal como doble diferencia, estimaciones de panel y estudios de cohorte |
| | B.4. Técnicas de emparejamiento |
| | B.5. Ecuaciones simultáneas |
| Nivel C: Estudios sin control riguroso del sesgo | C.1. Estudios de factores asociados |
| | C.2. Análisis multivariado |
| | C.2. Modelos de elección discreta lineal |
| | C.3. Diferencia simple |
| | C.3. Correlaciones |
| Nivel D: Estudios que utilizan métodos cualitativos | C4. Sistematización de registros administrativos |
| | D.1. Estudios cualitativos de rigor metodológico |
| | D.2. Sistematización de opiniones de expertos |
| | D.3. Análisis de marcos teóricos o conceptuales reconocidos por la comunidad académica internacional |

Fuente: Programas Presupuestales, Diseño, revisión y articulación territorial 2017 (MEF 2016).

Entre estas categorías, el MEF privilegia la evidencia de **Nivel A** que incluye meta-análisis (A.1), revisiones sistemáticas (A.2) y evaluaciones experimentales (A.3) y, en segundo lugar, las evidencias de **Nivel B** que incluye la regresión discontinua (B.1), variables instrumentales (B.2), los estudios que utilizan data longitudinal tales como la “doble diferencia”, “estimaciones de panel” y “estudios de cohorte” (B.3.), las técnicas de emparejamiento (B.4) y las ecuaciones simultáneas (B.5). La mayoría de los estudios incluidos en el Banco de Evidencia pertenecen a estas clasificaciones. El MEF incluye en un rango menor a las metodologías de nivel C y D. El **Nivel C** son estudios sin control riguroso del sesgo e “incluyen estudios no controlados o que utilizan métodos menos estrictos en el control de sesgo, pero hacen una evaluación apropiada de todos los posibles factores contextuales o determinantes” (MEF 2016, 110). Por último, en el **Nivel D** tenemos a los estudios que utilizan métodos cualitativos. El Banco de Evidencia incluye pocos estudios del nivel C y ningún estudio del nivel D.

Es importante indicar los tipos de documentos que según el MEF no constituyen evidencia y que, por lo tanto, **no son incluidos** en el Banco de Evidencia:

- “Normas, leyes o acuerdos nacionales e internacionales.

- Documentos descriptivos de problemas sin rigor metodológico.
- Opiniones de no expertos.
- Puntos de vista u opiniones aisladas, sin una sistematización adecuada” (MEF 2016, 111)

Para una explicación más detallada de la clasificación que realiza el MEF ver el documento *Programas Presupuestales, Diseño, revisión y articulación territorial 2017* (MEF 2016).

3. ¿Cómo actualizar el banco?

a. Paso 1: ¿Dónde buscar evidencia?

Antes de empezar a actualizar el Banco de Evidencia, debemos saber a) dónde y b) cómo buscar evidencia. Como consejo general, recomendamos utilizar el buscador Google Scholar (<https://scholar.google.com.pe/>) para encontrar publicaciones académicas en la web. Sin embargo, numerosas organizaciones han hecho esfuerzos por reunir la evidencia en sitios web fáciles de usar, hemos recopilado estos sitios web en el cuadro N° 2.

Cuadro N° 2: ¿Dónde puedo encontrar evaluaciones de impacto rigurosas?

Abdul Latif Jameel Poverty Action Lab

<http://www.povertyactionlab.org/evaluations>

www.povertyactionlab.org/policy-lessons

<http://www.povertyactionlab.org/publications>

“Revisiones de literatura” que revisan muchos estudios de un mismo tema en un artículo académico:

<https://www.povertyactionlab.org/sectors>

Innovations for Poverty Action

<http://www.poverty-action.org/work/publications>

<http://www.poverty-action.org/search-studies>

International Initiative for Impact Evaluation

El Repositorio de Evaluaciones de Impacto: <http://www.3ieimpact.org/en/evidence/impact-evaluations/impact-evaluation-repository/>

Banco Mundial:

Development Impact Evaluation: <http://www.worldbank.org/en/research/dime>

Independent Evaluation group: <http://ieg.worldbankgroup.org/>

Banco Interamericano de Desarrollo:

Repositorio de evaluaciones de impacto: <http://www.iadb.org/en/topics/development-effectiveness/evaluation-hub/impact-evaluation.17863.html>

Social Science Registry, American Economic Association

<https://www.socialscienceregistry.org/>

The Coalition for Evidence-based policy.

Evaluaciones de programas sociales: <http://evidencebasedprograms.org/>

Annual Reviews

Artículos especializados en revisiones de literatura: <http://www.annualreviews.org/>

The Cochrane Collaboration:

Revisiones de evidencia sobre intervenciones en el ámbito de la salud: <http://www.cochrane.org/>

Campbell Collaboration:

Revisiones de evidencia sobre intervenciones en política social, educación, empleo y crimen y justicia: <http://www.campbellcollaboration.org/>

What Works Clearinghouse

Revisiones de evidencia sobre intervenciones educativas: <http://ies.ed.gov/ncee/wwc/>

Crime Solutions

Revisiones de evidencia sobre crimen: <https://www.crimesolutions.gov/>

Evidencia sobre género y poblaciones vulnerables

REDIN-MIMP

Investigaciones en violencia de género a nivel nacional: <http://redin.gob.pe/>

Rompamos el Círculo-GRADE y Fundación Van Leer

Datos y estudios sobre violencia infantil: <http://rompamoselcirculo.org/>

Africa Gender Innovation Lab del Banco Mundial

Evaluaciones de Impacto sobre temas de género: <http://www.worldbank.org/en/programs/africa-gender-innovation-lab#5>

UNICEF-IRC:

Oficina de investigación de UNICEF-centrada en la niñez: <https://www.unicef-irc.org/research/260/>

b. Paso 2: ¿Cómo buscar evidencia?

Se pueden emplear diversas estrategias, pero de manera general recomendamos el siguiente derrotero para utilizar los buscadores académicos como Google Scholar. Es útil elaborar una lista de palabras clave para hacer la búsqueda más acotada. Por ejemplo, si buscamos estudios sobre atención a víctimas de violencia podemos usar palabras clave tales como “atención a víctimas de violencia doméstica”, “violencia familiar”, “violencia doméstica”, “violencia sexual”, “programas de atención de violencia doméstica”, entre otras. De la misma manera hay que incluir en las palabras de búsqueda términos que nos permitan identificar evaluaciones, como: “impacto”, “efecto”, “evaluación de impacto”, “RCT”, “evaluación experimental”, “Diferencias en Diferencias” etc.

c. Paso 3: ¿Cómo identificar los estudios que son de utilidad?

La primera búsqueda debería arrojar varios resultados por lo que deberemos aplicar una serie de filtros para refinar nuestra selección. En cada nivel de revisión buscamos explorar si los estudios encontrados cumplen con las metodologías o las temáticas que nos interesan.

Revisión de títulos: Primero, podemos leer el título de los estudios para ver si están relacionados con el tema en general y de ahí seleccionar una primera lista.

Revisión de resúmenes: Luego, podemos leer los “abstracts” o resúmenes que consisten en un párrafo corto describiendo el objeto, hipótesis, metodología y principales resultados del estudio. A partir de ahí podemos eliminar de la lista estudios cuyo resumen nos permita saber con claridad que no cumplen con la metodología o la temática que buscamos.

Revisión de estudios completos: Una vez depurada la lista, nos quedan los estudios que leeremos a profundidad para determinar si son relevantes para nuestra búsqueda y para extraer la información que necesitamos para el Banco de Evidencia. La revisión de la sección metodológica es particularmente importante para identificar si el estudio nos ofrece el tipo de evidencia que buscamos. Al leer el texto tengamos siempre estas preguntas en mente:

- ¿Qué tipo de evidencia es? ¿Es descriptiva, de proceso, correlación, o impacto causal?
- ¿La evidencia respalda las afirmaciones o las conclusiones que se hacen? Ej.: ¿La afirmación es causal pero la evidencia muestra sólo correlaciones?
- Si la afirmación es causal (respecto a impacto), ¿cuál es el contrafactual? ¿Se usa una metodología rigurosa para establecer el contrafactual? ¿Hay razones para pensar que sea sesgado?

d. Paso 4: ¿Cómo completar el Banco de evidencias?

Una vez tengamos todos los estudios que necesitamos, ya solo hará falta completar el Banco, siguiendo los campos determinados por el formato de la base de datos. Dichos campos son:

- ***#:** el número de la fuente en el Banco
- ***Título:** el título del estudio
- ***Autor:** el/la o los/las autores/as del estudio
- ***Año:** año de publicación del estudio
- **Temática:** a qué tema de relevancia para los PP corresponde este estudio. En este banco se incluyeron los temas de “Atención a víctimas de violencia y agresores”, “Prevención de la violencia” y “Desarrollo de autonomía”.
- **Lugar:** país, región y ciudad en el que se condujo el estudio.
- **Contexto:** incluir características del lugar y de la población objetivo (rural/urbano, etc.)
- **Propósito del estudio:** problema o asunto que se busca resolver.
- **Tipo de intervención:** qué clase de intervención es (ej. capacitación)
- **Detalles de la intervención:** descripción de la intervención, *no de la evaluación*.
- **Metodología:** nombre de la metodología usada ej. evaluación experimental; doble diferencia, etc.
- ***Nivel de evidencia:** nivel de evidencia según los criterios del MEF ej. A.1; B.3, etc.
- **Detalles metodológicos:** detalles y características importantes usados en el diseño de la evaluación (ej. uso de una variable instrumental). Incluir detalles del análisis.
- **Diseño de aleatorización:** tipo de aleatorización usado si se trata de una evaluación experimental (ej. simple, diseño de estímulo, etc.)
- **Detalles de la evaluación:** tamaño de la muestra.
- ***Fuente de la información:** indicar si la información proviene de una fuente primaria o secundaria.
- ***Relación abordada:** “identifica el factor y efecto, o los cofactores, así como la relación entre ellos.” (MEF 2016, 112)
- **Efectos heterogéneos:** si los hay, indicar si diferentes grupos de población experimentaron diferentes efectos.
- **Advertencias:** temas a considerar

- ***Comentarios sobre los resultados:** indicar los resultados de la intervención. “Identifica los resultados (efectos positivos, negativos, neutros, o de asociación) de los datos analizados en el documento de evidencia” (MEF 2016, 112)
- **Lecciones de política:** qué lecciones se pueden obtener del estudio que podrían ser aplicadas a las intervenciones del PP.
- **Vínculo:** hipervínculo a la página web de la publicación (si la hay).
- **Revista académica:** revista académica donde se publicó (ej. The Lancet, Quarterly Journal of Economics)
- **Factor de impacto de la revista:** es una medida de la importancia de una revista académica. Se calcula como la ratio entre el número de citas de los artículos de una revista y el número de publicaciones anuales (Thomson Reuters Impact Factor 1994). Un número mayor indica un mayor impacto. Sin embargo, es importante tener en cuenta que el Factor de impacto sólo sirve para comparar revistas dentro de un mismo campo de investigación, pues sus posibles valores varían entre áreas de estudio (Reuters 2016).

*Campos exigidos por la Tabla # 1.1. Narrativo de evidencias del Anexo 2: Contenidos mínimos de un programa presupuestal

4. ¿Cómo hacer uso de los estudios del Banco de Evidencia?

Continuamente escuchamos lo importante que es incluir “evidencia científica” en nuestros diseños. Los equipos técnicos encargados de una intervención pueden aprovechar los contenidos del Banco de Evidencia para hacer uso de evidencias en los procesos de diseño o re-diseño. Sin embargo, para hacer uso de la evidencia debemos primero asegurarnos que el estudio que estamos revisando tenga tanto **validez interna** como **validez externa**.

La **validez interna** se refiere a que la evaluación sea metodológicamente consistente y que los resultados, dentro del contexto específico de la evaluación, sean científicamente legítimos. En general, las evaluaciones aleatorias y cuasi-experimentales bien diseñadas tienen mayor validez interna y menos potencial de sesgo. Un buen estudio analiza todos los supuestos y posibles errores de manera detallada. Para el propósito de este documento, vamos a asumir que los estudios publicados en revistas académicas prestigiosas tienen validez interna, dado que tiene que pasar por el proceso de revisión de pares. En el **cuadro 2** se brindan algunas recomendaciones para evaluar la calidad de una fuente.

Cuadro N° 3: Calidad de un estudio a primera vista

Evaluar la calidad de un artículo académico puede ser complicado. No obstante, hay ciertos aspectos que pueden ayudarte antes de empezar a leer:

i) **Calidad de la revista:** El proceso de publicar un artículo en una revista académica es muy competitivo, por regla general las mejores revistas publican los mejores artículos. Para ver las revistas, puedes consultar este link: <http://www.scimagojr.com/journalrank.php>, y para ver artículos de investigaciones, puedes ir a este link: http://ideas.repec.org/top/top_journals.all10.html

ii) **Universidad o Institución a la que pertenecen los autores:** Generalmente, instituciones académicas eligen a sus investigadores por la calidad de su investigación. Si el autor de una publicación es de una institución reconocida, es probable que su evaluación sea rigurosa. Puedes encontrar las mejores instituciones en economía acá: http://ideas.repec.org/top/top_inst.all.html

iii) **Citas:** Si un artículo académico ha tenido una gran influencia en el mundo académico y de políticas públicas, es probable que haya sido citado muchas veces. Se puede averiguar cuantas citas tiene un artículo buscando aquí: <http://www.scholar.google.com>.

A diferencia de la validez interna, que podemos confiársela a los pares de los investigadores, quiénes revisan el estudio antes de que éste sea publicado, la **validez externa**, o **generalización**, es más difícil de determinar. Sin embargo, no estaríamos aprovechando al máximo la evidencia disponible si esperásemos a contar con evidencia sobre un estudio que se ajuste perfectamente a nuestro programa. A continuación, se presentan consideraciones sobre cómo saber cuándo se pueden generalizar los resultados de una evaluación.

El diseño de una evaluación a veces permite fácilmente generalizar los resultados. Por ejemplo, si los participantes de una evaluación fueron elegidos aleatoriamente de una población bien definida (ej. distritos de una región, o hogares de una aldea), se pueden generalizar seguramente los resultados al resto de la población que no participó en el programa pero que hubiera podido ser elegida a participar.

Sin embargo, si se quieren usar los resultados de una evaluación para asumir que estos mismos resultados pueden ocurrir en un contexto distinto al original, hay que estimar qué tan generalizable sería el estudio en el contexto distinto. Para analizar estas diferencias, hay que enfocarse en el diseño de la intervención o en la teoría del cambio del programa que fue evaluado.

Ciertos estudios fueron diseñados para ser más generalizables que otros. Por ejemplo, si un medicamento nuevo ha sido testeado en un lugar y ha resultado eficaz, es muy probable que el medicamento tenga un resultado muy parecido en cualquier lugar del mundo **donde exista la enfermedad**. Independientemente de si el contexto es el mismo o no, no tendría sentido implementar un programa para reducir la malaria en un lugar donde no existe malaria. Así pues, para evaluar cómo los resultados de un estudio se aplicarían a un contexto diferente, hay que revisar cuidadosamente si en el contexto en

el que queremos implementar el programa existen las mismas necesidades que fueron identificadas en el contexto en donde fue realizada la evaluación que tenemos como referencia.

En contraste, otros estudios están diseñados para contextos muy específicos, y cómo funciona la intervención depende mucho del implementador, el contexto institucional, y la población específica que recibe el programa, entre otras variables. Por ejemplo, en una evaluación de un currículo escolar nuevo, influyen muchas variables como la calidad de los profesores y las escuelas, el sistema institucional de la educación del país y otras variables similares, y puede resultar difícil aplicar los resultados de la evaluación a cualquier contexto. Mientras más variables afecten la intervención, más difícil será generalizar una evaluación. En este sentido, podría ser posible usar los resultados de la evaluación de un currículo nuevo en el caso de que donde se desea implementar un currículo similar sea suficientemente parecido al contexto de la evaluación original (sistema de educación del país, tipo de escuela donde se implementará el programa, nivel de educación de los profesores, grado y desempeño de los estudiantes, entre otros).

En estos casos, debemos analizar detalladamente el contexto donde se quiere implementar la intervención. Algunas preguntas que pueden guiar nuestro análisis son:

- Teoría de cambio: El programa fue diseñado para solucionar algún problema. ¿Existe el mismo problema en el nuevo contexto?
- Población: ¿En qué medida la población objetivo se parece a la población del estudio de referencia?
- Implementador: ¿El nuevo implementador tendrá la misma capacidad para que todo funcione igual? Ej.: Las agencias del gobierno no funcionan como las ONGs
- Procesos: ¿Se está en la capacidad de replicar todos los detalles técnicos, operativos, logísticos del programa o adaptarlos al nuevo contexto?

Para hacer el análisis anterior, se puede usar evidencia descriptiva producto de datos cualitativos y cuantitativos. Cuanta más información tengamos sobre el contexto de la evaluación y el contexto en el que queremos usar la evidencia, más podemos confiar en que esos resultados se aplican en nuestro contexto. De esta manera, se combinan teoría, evidencia descriptiva, y resultados de evaluaciones de impacto para responder si los resultados de un contexto se pueden replicar en otro.

Cuándo diseñemos un programa, o cuando ya lo estemos desarrollando, no olvidemos consultar regularmente nueva evidencia. El proceso de producción de evidencia es dinámico, y así como nuestro proyecto va cambiando, la evidencia relevante también, y cada día se terminan nuevos estudios, van cambiando los contextos, van cambiando las necesidades de la gente, etc. No dejemos de recoger información, y de generar nuestra propia evidencia: qué funciona en nuestro proyecto, que no, y por qué. Es fundamental que podamos identificar los factores, los procesos, los componentes, etc. que están generando los resultados, para mejorar lo que no está funcionando bien, y hacer más de lo que si funciona.

Referencias

MEF. *Programas Presupuestales, Diseño, revisión y articulación territorial 2017*. Lima: Neva Studio SAC, 2016.

MIMP. «Directiva General n°010-2016-MIMP “Normas para el seguimiento y evaluación en el Ministerio de la Mujer y Poblaciones Vulnerables-MIMP”.» Lima, 2016.

Reuters, Thomson. *Thomson Reuters website*. 2016.
<http://ipscience.thomsonreuters.com/product/journal-citation-reports/> (último acceso: 2 de Octubre de 2016).

«Thomson Reuters Impact Factor.» *Thomson Reuters*. 20 de junio de 1994.
<http://wokinfo.com/essays/impact-factor/>.

Anexo 2: Metodologías de evaluación

| | Metodología | Descripción | ¿Quién es el grupo de comparación? | Supuestos de la metodología | Datos requeridos |
|------------------------------|-----------------------------------|---|---|--|---|
| Métodos no experimentales | Pre-Post | Mide variación en resultados de participantes en el tiempo. La diferencia entre la situación anterior a un tratamiento y la situación posterior a este. | Los mismos participantes del programa -justo antes de aplicarse el programa- representan el grupo de comparación. | Se asume que el programa es el único factor que influenció cualquier cambio en el resultado obtenido. | Datos de los participantes antes y después del programa. |
| | Diferencia Simple | Mide las diferencias entre aquellos que participaron en el programa y otros que no lo hicieron, después de completado el programa. | Los individuos que no participaron en el programa (por alguna razón), pero para los cuales tenemos datos después de completado el programa. | Los no-participantes en promedio son idénticos a los participantes excepto por la participación en el programa, y tenían la misma probabilidad de participar en el programa antes de que comenzara. | Datos de los participantes y no participantes para después del programa. |
| Métodos Cuasi-Experimentales | Diferencias en Diferencias | Compara la variación en resultados a lo largo del tiempo de los participantes con la de los que no participaron del programa. | Los individuos que no participaron en el programa (por alguna razón), pero de quienes se recolectó información tanto antes como después del programa. Hay que buscar idealmente "similares" a los participantes. | Asume que si el programa no existiera, los dos grupos tendrían idénticas trayectorias a lo largo de este periodo. | Datos de ambos participantes, antes y después del programa. |
| | Regresión Multivariada | Los individuos que recibieron el programa son comparados con los que no recibieron, y otros factores que pueden explicar las diferencias de resultados son tomados en consideración (controlados). | Los individuos que no participan en el programa (por cualquier motivo), pero de quienes se recolectó la información antes y/o después del programa. En este caso la información no está comprometida sólo con indicadores de resultados, sino también por otras variables explicativas. | Los factores que fueron excluidos (porque eran no observables o que no fueron medidos) no sesgan los resultados porque no se correlacionan con los resultados o no difieren entre los que participan y los que no (al menos una de estas condiciones debe cumplirse para que no sesguen la estimación). | Los resultados así como las "variables control" para los que participan y los que no. |
| | "Matching" | Los individuos del grupo tratado son comparados con individuos similares en el grupo de comparación. | Matching exacto: Para cada participante, se escoge al menos un participante que es idéntico en las características seleccionadas. Propensity score matching: los que no participan tienen una serie de características muy parecidas a los participantes, lo que predice que hubiesen podido ser parte del programa. Se busca que el promedio ponderado de las características de los individuos de cada par sean similares. | Los no-participantes en promedio son idénticos a los participantes excepto por la participación en el programa, y tenían la misma probabilidad de participar en el programa antes de que comenzara. | Resultados, así como también "variables de coincidencia o matching" para ambos grupos, los que participan y los que no. |
| | Regresión Discontinua | Los individuos son clasificados en base a criterios medibles específicos. Hay una línea de corte que determina si un individuo es elegible o no. Los participantes que se encuentran en la vecindad del límite después son comparados entre sí y el criterio de elegibilidad es controlado. | Los individuos que están cerca de la línea de corte, pero que caen en el lado "equivocado", y por tanto no pueden participar del programa. | Después de controlar el criterio (y otras medidas de decisión), las diferencias restantes entre los individuos directamente debajo de la línea de corte o por encima de ella no son estadísticamente significativas y no producirán sesgo en los resultados. Un requerimiento necesario y suficiente para mantener este supuesto es que la línea de corte es un criterio estricto de adherencia. | Resultados, así como también todas las medidas para el criterio de selección (y cualquier otra medida de control). |
| | Variables instrumentales | La participación puede ser predicha por un factor accidental (aleatorio o casi aleatorio), o una variable "instrumental", que no está correlacionada con el resultado, de manera distinta al hecho de que predice la participación (y la participación afecta el resultado). | Individuos que, por su cercanía al factor aleatorio, se predice que no participaran y (posiblemente como resultado) no participan. | Si la variable instrumental no tuviese la habilidad de predecir la participación, este "instrumento" no tendría efecto sobre o no estaría correlacionado con el resultado. | Requiere información del resultado, el "instrumento", y otras variables de control. |
| | Método Experimental | Evaluación Aleatorizada | Método experimental para medir relaciones causales entre dos variables comparando participantes y no participantes cuando la participación fue determinada por sorteo. | La participación en el programa es asignada de forma aleatoria a un grupo de potenciales participantes elegibles mayor al número de cupos para el programa | La Aleatorización "funciona", es decir, los dos grupos son estadísticamente idénticos (en factores observable y no observables) |