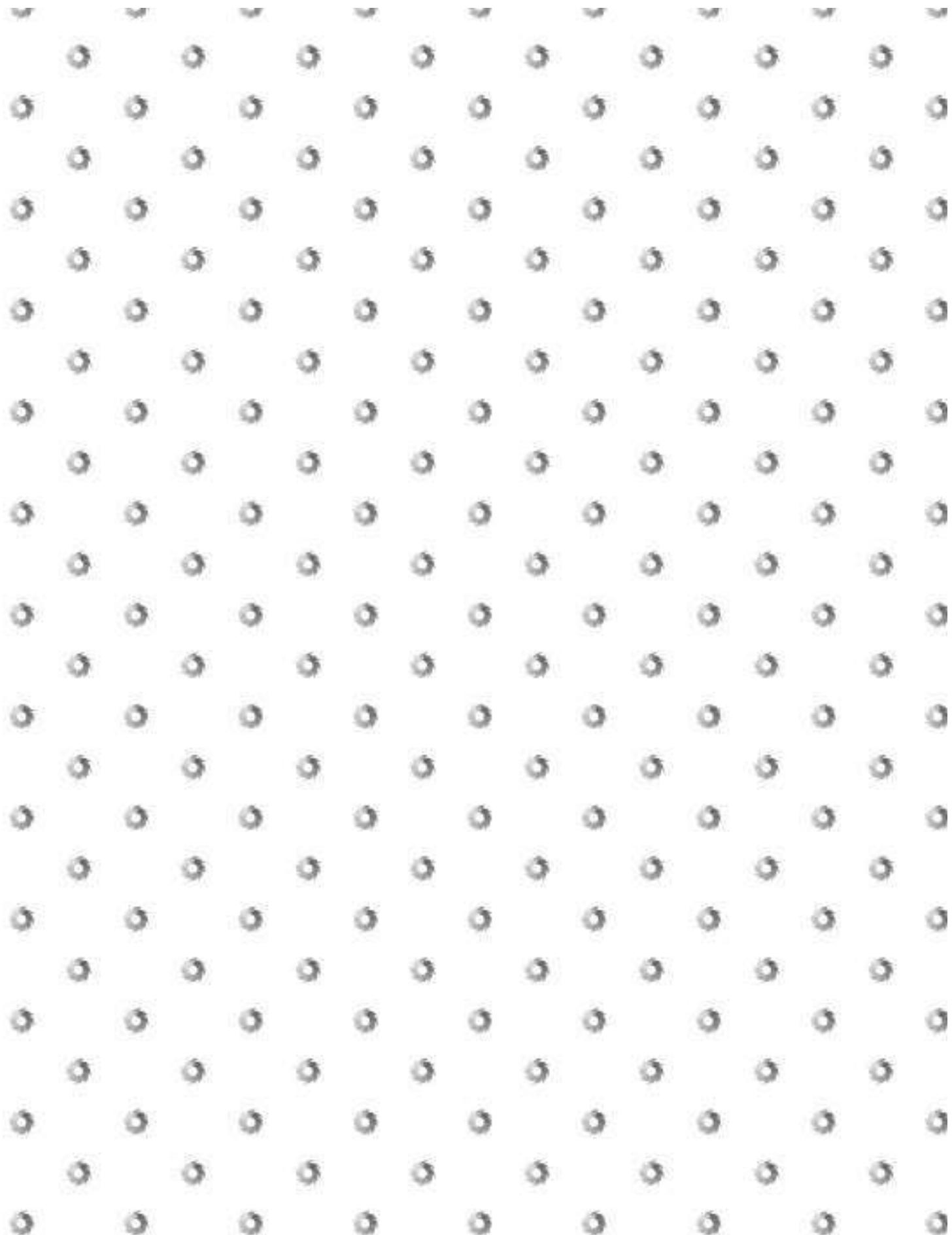


# GUÍA PARA LA EVALUACIÓN DE PROYECTOS SOCIALES

2017





**FUNDACIÓN ANDI**

Guía para la Evaluación de Proyectos Sociales

[www.andi.com.co](http://www.andi.com.co)**AUTORES:**

Arturo Harker, PhD

Marcela Meléndez, PhD

Juliana Ocampo, M.B.A.<sup>2</sup>

Con la colaboración de Fundación ANDI

**DISEÑO Y DIAGRAMACIÓN**

TRiDU Estudio

DERECHOS DE AUTOR DEL BÁNCO DE DESARROLLO PARA AMÉRICA  
LATINA, LA FUNDACIÓN ANDI ECONESTUDIO

2017

PROFESOR DE LA ESCUELA  
DE GOBIERNO DE LA UNIVERSIDAD DE LOS ANDES  
INVESTIGADOR AFILIADO DE ECONESTUDIO

DIRECTOR DE  
CONSULTORÍA ECONÓMICA  
Y SOCIO DE ECONESTUDIO

DIRECTOR DE COMUNICACIÓN  
ESTRÁTICA Y SOCIO  
DE ECONESTUDIO

ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

# *Índice*

GUÍA PARA LA  
EVALUACIÓN DE  
PROYECTOS SOCIALES

<b>1. CARTA DEL PRESIDENTE DE LA ANDI</b>	6
<b>2. ¿PARA QUÉ EVALUAR UN PROGRAMA DE INVERSIÓN SOCIAL EMPRESARIAL?</b>	8
<b>3. ¿QUÉ ES UNA EVALUACIÓN DE PROCESOS?</b>	12
<b>4. ¿QUÉ ES UNA EVALUACIÓN DE RESULTADOS?</b>	14
<b>5. ¿QUÉ ES UNA EVALUACIÓN DE IMPACTO?</b>	16
<b>6. ¿POR QUÉ EVALUAR IMPACTO ES MÁS DIFÍCIL?</b>	20
<b>6.1</b> ¿CUÁL ES LA DIFERENCIA ENTRE CAUSALIDAD Y CORRELACIÓN?	21
<b>6.2</b> ¿DE QUÉ DEPENDE LA POSIBILIDAD DE ESTABLECER CAUSALIDAD?	21
<b>7. ¿CÓMO SE HACE UNA EVALUACIÓN DE IMPACTO?</b>	24
<b>7.1</b> BASES PARA LA EVALUACIÓN	25
<b>7.2</b> EL DISEÑO METODOLÓGICO DE LA EVALUACIÓN	30
<b>7.3</b> ¿CÓMO ESCOGER LA METODOLOGÍA DE EVALUACIÓN?	30
<b>7.4</b> RECOLECCIÓN DE DATOS	43
<b>7.5</b> MANDA A LA OBRA	48
<b>8. NEGOCIOS DE ALTO IMPACTO SOCIAL: NUEVAS TENDENCIAS</b>	50
<b>9. REFERENCIAS BIBLIOGRÁFICAS</b>	54
<b>10. GLOSARIO</b>	56

ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

*Carta del*  
**PRESIDENTE  
DE LA ANDI**

11



# G

arantizar la eficiencia en el uso de los recursos que se invierten hace parte del saber hacer y la lógica del sector privado. Sin embargo, aunque las empresas destinan un monto considerable de recursos a programas sociales, no siempre lo hacen bajo la lógica privada que busca la asignación eficiente de los recursos y que garantiza que los efectos esperados se cumplan.

Según diferentes análisis adelantados por la ANDI, las compañías en Colombia se enfocan en medir y reportar principalmente los insumos de los proyectos sociales [contribuciones financieras, donaciones de producto, e recursos de capital humano invertidos], así como los productos [número de asistentes a una capacitación, instalaciones entregadas, número de personas beneficiadas]. Sin embargo, los impactos generados por la inversión social todavía no están siendo evaluados por las empresas.

Esta es una gran debilidad de las estrategias sociales del sector privado, ya que conocer los impactos que generan estas inversiones es vital para identificar qué tan efectivo está siendo el gasto, qué retornos le está generando a la empresa, y en un plano macro, cómo estamos contribuyendo al desarrollo del país. Los beneficios para una empresa de tener mecanismos de medición de sus intervenciones sociales son múltiples:

- Buena planeación de los proyectos
- Ejecución eficiente de los recursos
- Control del cumplimiento de objetivos
- Decisiones sustentadas sobre expansión, rediseño o eliminación de un proyecto

Desde hace más de dos años nos propusimos en la ANDI fortalecer la gestión social de las empresas, para volverla

más técnica y así promover un verdadero impacto social. Es por esto que esta guía busca ayudar a los gerentes de Responsabilidad Social, Sostenibilidad, Recursos Humanos o aquellos encargados de estructurar proyectos sociales, para que sean conscientes de la importancia de medir y evaluar sus esfuerzos, y conozcan desde un punto de vista técnico qué significa evaluar, qué tipos de evaluación existen y cómo se pueden adelantar.

Las empresas deben aprovechar su conocimiento para implementar procesos eficientes y técnicos y aplicarlos a sus estrategias sociales, porque sólo en la medida en la que los empresarios actúen como empresarios y no como donantes caritativos, estaremos en capacidad de generar grandes transformaciones sociales desde el sector privado.

Bruce Mac Master  
Presidente ANDI



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

¿Para qué evaluar un  
**PROGRAMA  
DE INVERSIÓN  
SOCIAL  
EMPRESARIAL?**

2



# L

os programas de inversión social buscan, en general, cambiar alguna realidad. Por ejemplo, aumentar el ingreso de los hogares, mejorar la calidad de la educación, reducir el crimen o el embarazo adolescente. Esto es común a la inversión social que realiza el gobierno, a la que realiza el sector empresarial, y a aquella que realizan las empresas que tienen como núcleo de negocio la generación de esas transformaciones. En los tres casos es relevante preguntarse por la efectividad de esas inversiones.

¿Se consigue con la inversión lo que se pretende? ¿Es el resultado de la inversión uniforme para los distintos grupos de la población beneficiaria o hay algunos para los que el resultado es más, o menos, contundente? ¿Qué explica las diferencias en el resultado? ¿Qué aspectos del programa de inversión social podrían ajustarse para obtener un mejor resultado? ¿Puede decirse que la inversión ha sido rentable? O, si la inversión no ha conseguido los resultados deseados, ¿Qué ha faltado? ¿Se ha cumplido la cabalidad al plan de actividades propuesto? ¿Qué dificultades se enfrentaron durante la etapa de implementación? ¿Es posible decir, simplemente, que el programa no tiene efecto? Estas son algunas de las preguntas de la evaluación.

En esencia, evaluar permite establecer qué funciona y qué no, y en esa medida es una herramienta poderosa para la rendición de cuentas y para la medición del retorno a cada peso invertido. Así, parecería obvia la necesidad de pesar todo programa de inversión social por el filtro de una evaluación. Esto, sin embargo, no es una práctica común

Las empresas se aproximan a sus inversiones sociales de

máneras diversas. Lo más frecuente es que concentren sus esfuerzos en controlar y medir los insumos y productos inmediatos de los programas –cuánto dinero se gasta, cuántos libros se distribuyen, cuántas personas se benefician–, más que en identificar si los programas consiguen los impactos esperados. Esto es precisamente lo que muestran los hallazgos del ejercicio realizado por la ANDI y KPMG en 2016 (**VER RECUADRO 1**).

## RECUADRO 1

### DE 100 EMPRESAS ANALIZADAS POR KPMG Y FUNDACIÓN ANDI<sup>1</sup>

- El 56% presenta una estrategia de inversión social y la explica de manera detallada, 12% solamente la menciona y 17% desarrolla los programas sociales que no están enmarcados en una estrategia claramente definida. El 15% restante no reporta inversión social en su gestión.
- El 75% cuantifica los insumos de sus inversiones sociales (donaciones en dinero, tiempo de voluntariado de los empleados, y donaciones en especie).
- El 80% cuantifica los productos de sus inversiones sociales (el número de personas, hogares o negocios atendidos).
- El 18% describe los resultados de sus inversiones sociales, en términos cualitativos. Sólo 11% reporta métricas cuantitativas, y las métricas reportadas se refieren exclusivamente a resultados y no a impactos. Además, las métricas utilizadas varían sustancialmente entre empresas y tipos de programas.
- Ninguna de las compañías estudiadas reporta métricas cuantitativas para los impactos.

<sup>1</sup> Estudio realizado a partir de los reportes de sustentabilidad de las 100 empresas con mayores ingresos operativos en Colombia. El estudio completo puede ser consultado en la página de la ANDI ([www.andi.com.co](http://www.andi.com.co)).



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

El monitoreo es el proceso continuo de seguimiento a lo que ocurre al interior de un programa con base en datos que se recogen sistemáticamente, para informar las decisiones que se hacen en el día a día de su ejecución. El uso de un sistema de monitoreo sirve para comparar la

**RECUADRO 2****"MÁS NUTRICIÓN, MÁS EDUCACIÓN"<sup>12</sup>**

Desde 2006 el programa de alimentación escolar "Más nutrición, más educación" ha entregado anualmente refrigerios a niños en 10 escuelas de Cundinamarca, eligiendo cada año 10 escuelas nuevas que serán beneficiarias. Desde su diseño se ha establecido que el programa busca mejorar el desempeño escolar a través de una mejora en la nutrición de los niños del departamento.

Desde el inicio, se diseñó una batería de indicadores para hacer seguimiento al programa, donde cada dos semanas se registra lo siguiente:

- Los recursos invertidos en la compra y distribución de los refrigerios.
  - El número de refrigerios entregados a cada colegio.
  - El número de refrigerios no consumidos en cada colegio.
- Este es un sistema de monitoreo y seguimiento.

mánera en la que un proyecto se va implementando contra el plan inicial, y adoptar correctivos.

En contraste con el monitoreo, la evaluación -que es la revisión periódica de un programa para responder preguntas específicas relacionadas con su diseño, implementación y resultados- se realiza en momentos particulares del tiempo (no en el día a día). Usualmente, para mantener la objetividad, una evaluación incluye la perspectiva de expertos externos. Su diseño, metodología y costo varían según el tipo de preguntas que se quieran responder. En general, las preguntas de la evaluación pueden clasificarse en una de tres categorías, así:

- a. PREGUNTAS DESCRIPTIVAS: cuando la evaluación busca determinar lo que ocurre con el programa y describir los procesos, las relaciones organizacionales y la visión de los distintos grupos de interés.
- b. PREGUNTAS NORMATIVAS: cuando la evaluación busca comparar lo que ocurre con el programa con lo que debería estar ocurriendo, identifica las actividades realizadas y establece si se han alcanzado las metas.
- c. PREGUNTAS DE CAUSA Y EFECTO: cuando la evaluación examina los resultados y busca establecer el cambio que es directamente atribuible al programa. **VER EJEMPLOS DE LOS DISTINTOS TIPOS DE PREGUNTAS EN EL RECUADRO 3.**

**EL MONITOREO**

es el proceso continuo de seguimiento a lo que ocurre al interior de un programa en el día a día de su ejecución.

De acuerdo con las preguntas que plantean e intentan responder, las evaluaciones pueden tomar tres formas: (1) evaluación de procesos, (2) evaluación de resultados o (3) evaluación de impacto. Las tres formas de evaluación pueden pensarse como bloques complementarios incrementales dentro de un mismo proceso de medición y evaluación de una inversión social: cada una aporta lecciones diferentes y es necesaria para posibilitar la siguiente.

**LA EVALUACIÓN**

es la revisión periódica de un programa para responder preguntas específicas relacionadas con su diseño, implementación y resultados.

En esta guía buena parte de la discusión se concentra en la evaluación de impacto, que es la forma de evaluación que establece si en efecto el resultado obtenido (el cambio observado en la población beneficiaria) puede identificarse como un impacto directo y exclusivo del programa. Es la forma de evaluación más exigente en términos de presupuesto, tiempo e información, y la única que sirve para responder preguntas de causa y efecto. Pero esto no quiere

<sup>12</sup>) Programa "Más nutrición, más educación" es un caso hipotético creado por los autores. Datarán la gira en utilizar este caso como herramienta pedagógica.



dicir que el monitoreo y los otros dos tipos de evaluación no sean igualmente importantes. Las distintas formas de evaluación no son incompatibles ni sustitutas. Por el contrario, se complementan; pues responden a preguntas distintas, todas ellas relevantes.

FIGURA 1

### FORMAS DE EVALUACIÓN

De acuerdo con las preguntas que plantean e intentan responder, las evaluaciones pueden tomar tres formas:

(1) evaluación de procesos, (2) evaluación de resultados o (3) evaluación de impacto. Las tres formas de evaluación pueden pensarse como bloques complementarios incrementales dentro de un mismo proceso de medición y evaluación de una inversión social: cada una aporta lecciones diferentes y es necesaria para posibilitar la siguiente.

RECUADRO 3

## "MÁS NUTRICIÓN, MÁS EDUCACIÓN"

Para la evaluación del programa de alimentación escolar "Más nutrición, más educación", un año después de su puesta en marcha se plantearon las siguientes preguntas:

#### a. PREGUNTAS DESCRIPTIVAS

- ¿Qué operador fue responsable de prestar el servicio y qué opinión tienen los rectores de los colegios acerca de la manera en que se prestó este servicio?
- ¿Qué proporción de los colegios atendidos en el último año reportó incumplimientos por parte del operador en la provisión de los refrigerios?
- ¿Con qué frecuencia se reportó que ocurrían esos incumplimientos y cómo se solucionaron?
- ¿Qué impidió en esos casos el cumplimiento en la prestación del servicio?
- ¿Qué opinión tienen los padres de familia y los estudiantes acerca de los refrigerios que se entregaron en los colegios de sus hijos?
- ¿Cambiaron en algo los hábitos de alimentación de los menores desde que el colegio comenzó a entregar los refrigerios? Por ejemplo, ¿Se suscitó el desayuno que antes tomaban en sus casas por el refrigerio?

#### b. PREGUNTAS NORMATIVAS

- ¿Se puede decir que, a pesar de los eventos de incumplimiento, se alcanzó la meta de entregar 5.000 refrigerios diarios en los colegios elegidos?
- Se había previsto un costo promedio por refrigerio entregado de 2.000,00 pesos. ¿En la práctica qué tan bien corresponde este valor con lo efectivamente gastado?
- Se había previsto que cada refrigerio aportaría 20% de los requerimientos de proteína por día y 30% de las calorías necesarias a los menores. ¿Se cumplió con los objetivos de composición calórica de los refrigerios?

- Se había previsto observar cambios en la asistencia escolar de los niños como resultado de la entrega de los refrigerios. ¿Cambiaron las tasas de asistencia desde que se comenzaron a entregar refrigerios?

- Se había previsto observar cambios en el desempeño escolar de los niños como resultado de la entrega de los refrigerios. ¿Cambiaron los resultados de los niños en las pruebas de conocimiento que realiza el colegio desde que se comenzaron a entregar refrigerios?

#### c. PREGUNTAS DE CAUSA Y EFECTO

- ¿Hay un cambio para bien en la nutrición de los menores que sea atribuible al programa? ¿De qué magnitud es?
- ¿Hay un cambio positivo en el desempeño escolar de los menores que sea atribuible al programa? ¿De qué magnitud es?



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

¿Qué es una  
**EVALUACIÓN  
DE PROCESOS?**

3



# L

**a evaluación de procesos es la forma de evaluación que mira la implementación del programa para establecer si se han alcanzado metas de ejecución y si las estrategias se han implementado de acuerdo con lo planeado.** En ese sentido, en una evaluación de procesos se miden, determinan y analizan las acciones adelantadas que apuntan a la implementación o cumplimiento de los programas. Este tipo de evaluación sirve para responder preguntas descriptivas y normativas, pero no preguntas de causa y efecto. Por ejemplo, una evaluación de procesos confirmará que 200 personas han completado un programa de capacitación laboral. Esta es la forma de evaluación más cercana en espíritu al monitoreo, y es útil para establecer si el programa debe continuarse, refinarse

#### RECUADRO 4

## "MÁS NUTRICIÓN, MÁS EDUCACIÓN"

Un año después de puesto en marcha el programa "Más nutrición, más educación", se contrató un equipo de evaluación y seguimiento para evaluar la calidad de la implementación del programa. A través de entrevistas con los rectores de los colegios, con los padres de familia y con los operadores del programa en cada municipio se pudo establecer [1] que para dos de las escuelas beneficiarias hubo un re-

traso en la puesta en marcha del programa por dificultades para asegurar la distribución de los refrigerios; [2] que a uno de los colegios le llegó incompleta sistemáticamente la ración de refrigerios y que algunos alumnos tuvieron que compartir la ración; y [3] que en uno de los municipios, donde habita población muy pobre, los padres sustituyeron el desayuno que daban a sus hijos por el refrigerio escolar.

o eliminarse con base en el éxito de su ejecución.

Parece razonable que todo programa de inversión social cuente al menos con un sistema de monitoreo y una evaluación de procesos. No es posible atribuir un impacto positivo a un programa, si no se ha establecido previamente que este fue imple-

mentado de acuerdo con el plan inicial. Así, la evaluación de procesos es un paso obligado para hacer una evaluación de impacto. No ocurre lo contrario. La evaluación de procesos debe verse como una fase inicial, que es informativa por sí misma, y necesaria para luego avanzar hacia los otros tipos de evaluación.



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

¿Qué es una  
**EVALUACIÓN DE  
RESULTADOS?**

4



**L**a evaluación de resultados es la forma de evaluación más frecuente. Es la comparación del grupo de población beneficiado por el programa antes y después de la aplicación del programa. Sirve para responder preguntas descriptivas y normativas, diferentes a las que responde una evaluación de procesos. Tampoco responde preguntas de causa y efecto. Por ejemplo, una evaluación de resultados dirá cuántas de las 200 personas que completaron el programa de capacitación laboral demostraron sentirse más seguras respecto de sus habilidades, cambiaron algunos de sus comportamientos, y consiguieron trabajo. Mide el cambio observado en la población beneficiaria del programa.

La observación del grupo de población beneficiado por el programa sólo después, es otra forma, más débil, de evaluación de resultados. En este caso, los estimadores de cambio suelen basarse en una combinación de información cualitativa, capturada

por recordación, entrevistas con actores claves, o grupos locales, con los registros administrativos del programa y otros de fuentes oficiales, que se usan para obtener comparaciones globales de los grupos de población beneficiados, con grupos similares.

#### RECUADRO 5

### "MÁS NUTRICIÓN, MÁS EDUCACIÓN"

Al equipo de evaluación y seguimiento se le encargó también evaluar los resultados del programa. Con base en los registros administrativos de los colegios, se hicieron dos hallazgos: [1] desde que se puso en marcha el programa, la asistencia escolar aumentó en promedio un 10%; [2] desde que se puso en marcha el programa, los indicadores de nutrición de los niños (talla y peso para la edad) mejoraron en un 4%.

Si no hay un cambio observable en los indicadores de resultado tras la puesta en marcha del programa que se quiere evaluar, no hay lugar a indagar si el cambio es atribuible al programa. Una evaluación de resultados es una primera aproximación a esos indicadores. Establece lo básico: si hubo algún cambio en los indicadores de interés que pueda justificar una exploración más profunda. No debe confundirse con una evaluación de impacto ni su resultado interpretarse como el hallazgo de un efecto causal del programa.

La elección entre detenerse en una evaluación de resultados o seguir hasta una evaluación de impacto, depende de la magnitud y la recurrencia de la inversión que se realiza. Si la inversión es pequeña y se realiza por una sola vez, probablemente no sea costo-beneficioso invertir tiempo y recursos en una evaluación de impacto.



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

¿Qué es una  
**EVALUACIÓN  
DE IMPACTO?**

50



# U

na evaluación de impacto es aquella que contrasta el cambio en los indicadores de resultado entre el grupo de beneficiarios y un grupo de control, para así poder identificar la proporción de este cambio que es realmente atribuible al programa. Este es el único tipo de evaluación que

permite establecer causalidad entre el programa y el resultado observado. Por esto mismo, también es el único tipo de evaluación que permite dimensionar la rentabilidad de la inversión social e identificar si los recursos invertidos superan los beneficios de un uso alternativa. Es

la única forma de establecer en qué medida se alcanzó el objetivo de transformación propuesto. También es la única forma de aprender sobre los elementos que limitan o potencian la efectividad de la inversión, para dar los pasos necesarios en el proceso de gestión para la calidad.

## Recuadro 6

### "MÁS NUTRICIÓN, MÁS EDUCACIÓN"

Al equipo de evaluación y seguimiento de "Más nutrición, más educación" se le preguntó si los resultados observados podían atribuirse al programa de alimentación escolar. La respuesta fue que para establecer con precisión el impacto del programa era necesario abordar la pregunta con una metodología más sofisticada, que incluyera al menos los siguientes tres elementos:

- Un conjunto de indicadores que definieran y midieran

los insumos del programa, los procesos de implementación, los productos y los impactos deseados.

- Un grupo de control que permitiera establecer si los cambios observados en los indicadores de resultado después de la intervención podían atribuirse al programa. Un grupo de control (también conocido como "grupo de comparación" o "grupo de referencia") es un grupo

idéntico al grupo de población beneficiado por el programa, en todas las variables relevantes antes de la intervención de la empresa, pero que no participa del programa.

- La medición de la contribución del programa al cambio en las variables objetivo –es decir, los impactos deseados sobre la población–, a partir de procedimientos estadísticos universalmente aceptados.



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

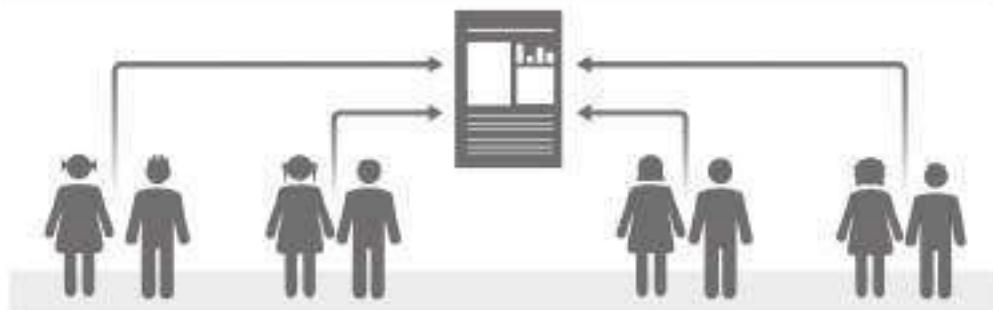
La decisión de evaluar impacto suele darse, sin embargo, en un contexto de restricciones. En particular, es necesario tener en cuenta los siguientes tipos de restricciones:

- a. DE PRESUPUESTO:** Las restricciones de presupuesto determinan la información que es posible recopilar, el tamaño y experiencia profesional del equipo responsable del ejercicio, y la profundidad del análisis.
  - b. DE TIEMPO:** Las restricciones de tiempo determinan cuándo comienza y termina la evaluación, y por ende el horizonte de tiempo para el cual es posible identificar el impacto del programa. En ocasiones, incluso, cuando el programa está aún en curso o está recién implementado, no ha transcurrido tiempo suficiente para evaluar impacto.
  - c. DE INFORMACIÓN:** Las restricciones de información limitan el diseño y alcance de la evaluación.
  - d. DE DISEÑO DEL PROYECTO:** El proceso de selección de la población beneficiaria limita el tipo de evaluación aplicable.
  - e. DE IMPLEMENTACIÓN DEL PROYECTO:** La calidad y uniformidad con la que el programa se pone en práctica limita la forma de evaluación aplicable.
- En efecto, en la práctica lo más frecuente es que no se cumplan las condiciones necesarias para llevar a cabo una evaluación ideal que permita establecer con precisión el impacto de un programa. De esta forma, casi siempre existe el reto de decidir si es posible realizar una evaluación de impacto de buena calidad, bajo las restricciones que se enfrentan. Surge entonces las siguientes tres preguntas:
- a. ¿En qué medida las aproximaciones a la evaluación de impacto en contextos de restricciones de recursos, tiempo e información arrojan conclusiones válidas?**
  - b. ¿Qué tipo de concesiones técnicas son aceptables?**
  - c. ¿Cuáles son los requisitos metodológicos mínimos para que un estudio pueda considerarse una evaluación de impacto de buena calidad?**

En otras palabras, ¿qué alternativas tiene el inversionista social para evaluar el impacto de sus inversiones y cuáles son las limitaciones de cada una de ellas, cuando no es posible realizar una evaluación de impacto ideal? En el Módulo 7 se discuten algunos escenarios posibles y las aproximaciones metodológicas más apropiadas para adelantar evaluaciones de impacto rigurosas bajo distintos contextos de restricción. Un mensaje importante antes de cerrar este módulo, es el siguiente: en ocasiones, cuando no es posible asegurar una evaluación de impacto con una calidad mínima, es preferible optar por detenerse en una evaluación de resultados y aprender lo que es posible aprender de ella.



FIGURA 2. TIPOS DE EVALUACIÓN



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

¿Por qué evaluar  
**IMPACTO ES  
MÁS DIFÍCIL?**

61



# L

a evaluación de impacto es la forma de evaluación metodológica más sofisticada. Esó tiene que ver con el objetivo de establecer causalidad entre el programa de inversión y el resultado esperado. Es decir, se necesita una metodología más exigente—en términos de información, tiempo y recursos— para poder medir los efectos que son atribuibles exclusivamente al programa.

## **6.1 ¿CUÁL ES LA DIFERENCIA ENTRE CAUSALIDAD Y CORRELACIÓN?**

Dos variables están correlacionadas cuando se mueven de manera simultánea en la misma dirección o en direcciones opuestas. Por ejemplo, suela ocurrir que se observe simultáneamente un mayor ingreso per cápita y una mayor tasa de penetración del Internet. Son variables que tienen una correlación positiva; mientras más alta la una, más alta suele ser la otra. Esta correlación positiva no permite, sin embargo, establecer causalidad en ninguna dirección. Seña errado concluir a partir de la observación de la correlación entre las dos variables, que la mayor penetración del Internet contribuye a reducir la pobreza o

que la mayor riqueza aumenta la penetración del Internet; incluso si una de las dos afirmaciones resulta intuitivamente más atractiva que la otra.

Un ejemplo que hace aún más claro el error de asignar un efecto causal a una correlación: en la historia del mundo sucedió de manera simultánea que disminuyeron los ataques de piratas y aumentó la estatura promedio de la población. Estas son dos variables que tienen una alta correlación negativa: a menos ataques de piratas, mayor la estatura. Evidentemente, ni el control de la piratería determina el aumento en la estatura de las personas, ni la mayor estatura de las personas explica la disminución en los ataques de piratas.

Para establecer una relación de causa y efecto, es decir una relación de causalidad, entre las dos variables, es necesario realizar una evaluación de impacto. Una evaluación de resultados apenas permite establecer una correlación.

## **6.2 ¿DE QUÉ DEPENDE LA POSIBILIDAD DE ESTABLECER CAUSALIDAD?**

La primera que es necesaria para establecer una relación

causa-efecto es que existe ese grupo de control similar en todas sus características al grupo de población que será intervenido por el programa. El grupo de control es aquél cuyos resultados en las variables que se desea impactar con la inversión servirán como punto de referencia para comparar los resultados del grupo que en efecto se trata con la inversión. En la literatura se dice que el grupo de control servirá como "contra-factual". Para un ejemplo del modo en que la disponibilidad de un buen grupo de control permite establecer el impacto atribuible a un programa de inversión, VER RECUADRO 7.

### **EL GRUPO DE CONTROL**

Es un grupo idéntico al grupo de población beneficiado por el programa, en todas las variables relevantes, antes de la intervención de la empresa, pero que no participa del programa.

Identificar un grupo de control adecuado es más fácil cuando la evaluación de impacto se integra al programa de inversión social, desde la fase inicial de diseño. No obstante, existen métodos



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

## Recuadro 7

**"MÁS NUTRICIÓN, MÁS EDUCACIÓN"**

En los 10 años de implementación del programa, se ha observado que el indicador de desnutrición crónica que es la talla o estatura para la edad, ha tenido una mejora sustancial: mientras que en 2006 el 15% de los niños en las escuelas beneficiarias sufría desnutrición crónica, en 2016 esta proporción fue de apenas el 4%. ¿Es suficiente esta evidencia para concluir que el programa ha sido exitoso en términos de su impacto? Es decir, ¿es posible atribuir al programa este cambio en la proporción de niños con desnutrición?

Si absolutamente nada relacionado con la alimentación y el bienestar de los niños ha cambiado entre 2006 y 2016, aparte de la implementación del programa,

la respuesta es sí. Pero si se observan cambios en otras variables que puedan tener un impacto directo o indirecto sobre el estado nutricional de los niños en el mismo periodo, entonces la respuesta es incierta.

Este último escenario representa mejor el mundo real y el contexto en el que se deben evaluar los proyectos de inversión social. Si se sabe que la economía del departamento ha experimentado 10 años de prosperidad y que el ingreso promedio de los hogares se ha triplicado durante el periodo, es posible que apenas una pequeña parte de la diferencia en el indicador de nutrición sea exclusivamente resultado del programa de alimentación escolar.

Esta realidad hace que sea necesario un contra-factual que sirva como punto de referencia. En este caso particular, sería útil tomar como contra-factual a los estudiantes de otras 10 escuelas casi idénticas -en el mismo departamento y que atiendan a niños de hogares similares-, que no hayan sido beneficiarias del programa de alimentación. Estas últimas conformarían el grupo de comparación. De esta manera, se podría contrastar el cambio en el indicador de desnutrición en las 10 escuelas beneficiarias con el cambio en el indicador de desnutrición en las 10 escuelas del grupo de control que serviría como contra-factual.

Suponga ahora que se observan los siguientes datos:

POBLACIÓN	2006	2016	Diferencia
Grupo de beneficiarios	15%	4%	-11%
Grupo de control	15%	7%	-8%

¿Es suficiente esta evidencia para concluir que el programa ha sido exitoso en términos de su impacto? Si el grupo de control es realmente comparable, la respuesta es sí: 3 puntos porcentuales de la disminución en la nutrición son atribuibles al programa. Si sólo se mira el grupo beneficiario, se atribuiría al programa un impacto de 11 puntos porcentuales, mucho más alto que el que realmente tiene.



para "construir" grupos de control cuando esto no sucede. En la sección 7.2 de la guía, se discute en detalle cómo se asegura la disponibilidad de un grupo de control adecuado y los retos metodológicos que esto representa.

Uno de esos retos, que vale la pena mencionar desde ya, es la disponibilidad de una medición inicial que permita caracterizar a la población de potenciales participantes en el programa, antes de poner en marcha la inversión. De esa población saldrán posteriormente los grupos de beneficiarios y de control. Esta medición inicial es lo que en la literatura se llama una línea de base.

#### LA LÍNEA DE BASE

Medición inicial que permite caracterizar a la población de potenciales participantes en el programa, antes de poner en marcha la inversión.

Es esa medición del punto de partida o condición inicial de los indicadores que miden los impactos deseados, lo que luego permite esconder el cambio en ellos.

Recuadro 8

#### "MÁS NUTRICIÓN, MÁS EDUCACIÓN"

Antes de poner en marcha el programa de alimentación escolar "Más nutrición, más educación", se levantó una línea de base, que permitiera caracterizar con un buen nivel de profundidad al grupo de beneficiarios elegible para participar en el programa. Esto se hizo de la siguiente manera:

- Se identificaron los colegios potencialmente elegibles para participar en el programa en los tres municipios escogidos para la intervención. Las condiciones de elegibilidad fueron: **(1)** que tuvieran al menos 500 estudiantes cada uno; **(2)** que estuvieran ubicados en el casco urbano de los municipios o a una distancia máxima de 5 kilómetros del mismo; **(3)** que tuvieran un 80% de estudiantes provenientes de viviendas de estratos 1, y 2, y **(4)** que tuvieran un refrigerador con suficiente capacidad para almacenar los refrigerios de los niños durante al menos 2 días.
- Se identificaron las calificaciones trimestrales en matemáticas y comprensión de lectura del 100% de cada uno de los

ninos en estos colegios durante el último año escolar. También se obtuvieron mediciones de su talla y peso. El servicio de medición se contrató con un equipo externo a los colegios.

- Se recolectó y organizó la información de los niños y sus familias que tenían registrada los colegios: estrato de la vivienda, calificación SISBEN del hogar, nivel educativo máximo de madre y padre, número de personas que componen el hogar, ocupación de madre y padre, número de libros con que cuenta el hogar, disponibilidad de tablet o computador con conexión a Internet en el hogar.
- Esta información se organizó en una base de datos y se utilizó para seleccionar un grupo de 25 colegios estadísticamente comparables, en términos de la implementabilidad del programa, la relevancia del mismo y los valores de los indicadores de resultado que se querían afectar. De este grupo, se asignaron 10 colegios al grupo de beneficiarios y 15 colegios al grupo de control.

Otro reto es, por supuesto, la posibilidad de seguir en el tiempo a la población de ambos grupos, tanto de beneficiarios como de control, para tener una

segunda medición de las variables que se quiere impactar un tiempo después de la intervención. Esto se denomina un momento de "seguimiento".



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

¿Cómo se hace una  
**EVALUACIÓN  
DE IMPACTO?**

7



# E

Existe una amplia gama de libros, artículos académicos y manuales que, con diferente énfasis y nivel de profundidad, presentan las bases teóricas y prácticas para el diseño y aplicación de evaluaciones de impacto. En este módulo se presentan los cuatro pasos fundamentales que debe seguir toda evaluación de impacto de calidad: (1) establecer las bases para el diseño de la evaluación; (2) definir el método de evaluación; (3) recoger los datos para

la evaluación; y (4) analizar los datos para estimar el impacto.

El propósito de este módulo es explicar la importancia de estos pasos, y guiar al lector en su aplicación, entendiendo los diferentes contextos y restricciones de presupuesto, tiempo y datos. Se discuten también los retos, necesidades y limitaciones de los diferentes métodos de evaluación de impacto. Los contenidos se basan en la revisión sistemática de la literatura técnica más relevante sobre monitoreo y

evaluación de programas sociales y políticas públicas (VER REFERENCIAS BIBLIOGRÁFICAS PÁGINA 55).

## 7.1 BASES PARA LA EVALUACIÓN

Antes de comenzar a pensar en el diseño metodológico para la evaluación de impacto de un proyecto de inversión social, se deben resolver unas preguntas básicas. En esta sección se presentan las 4 actividades que son fundamentales para luego definir el diseño metodológico de una evaluación de impacto.

FIGURA 3:

### ESQUEMA BASE DE UNA CADENA DE RESULTADOS



#### a. DEFINIR LA PREGUNTA GUÍA

El eje principal de toda evaluación de impacto es la pregunta de causa y efecto que se quiere responder. En la etapa de diseño del programa de inversión, la empresa debe responder cuál es el resultado que espera tener al realizar esta inversión social y qué variables específicas quiere impactar.

Estas son las variables que guiarán el diseño de la evaluación de impacto.

#### b. DESCRIBIR LA TEORÍA DEL CAMBIO

El segundo paso en la preparación para una evaluación de impacto es entender cómo las actividades y acciones que comprende el programa pueden traducirse en los resultados deseados.

#### RECUADRO 9

### "MÁS NUTRICIÓN, MÁS EDUCACIÓN"

En el ejemplo del programa de alimentación escolar "Más nutrición, más educación", la pregunta que guía la evaluación de impacto es: ¿mejora el programa el desempeño académico de los niños participantes? Al definir esta pregunta guía, se definen tanto los impactos deseados del programa como sus beneficiarios.



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

Esto se conoce en la literatura sobre monitoreo y evaluación como teoría del cambio; la descripción de cómo se supone que un programa de inversión social puede llegar a los impactos deseados. La teoría del cambio describe la secuencia lógica de causas y efectos que conectan lo que se hace en el programa, con las transformaciones positivas que se espera obtener. Idealmente, este ejercicio debe apoyarse en evidencia científica publicada en revistas especializadas.<sup>3</sup> Cuando no hay evidencia disponible en la literatura, la experiencia previa del operador del programa puede servir para describir la teoría del cambio.

En el ejemplo del programa de alimentación escolar "Más nutrición, más educación", la pregunta sería: ¿por qué tendría el programa de alimentación escolar un impacto positivo sobre el desempeño escolar de los estudiantes? Es necesario entender los canales a través de los cuales la entrega de refrigerios puede conducir a que un niño tenga mejor desempeño escolar.

Por una parte, apoyándose en los estudios sobre el impacto de la escasez sobre el comportamiento humano del profesor Sendhil Mullainathan de la

Universidad de Harvard, se puede suponer que el hambre tiene un impacto importante sobre la capacidad de prestar atención y concentrarse.<sup>4</sup> Así, la teoría del cambio del programa "Más nutrición, más educación" puede seguir esta línea argumentativa: el hambre reduce los recursos mentales de los niños, de tal manera que un refrigerio –al eliminar el hambre– puede promover que los niños aprendan más.

Por otra parte, existe evidencia sobre el impacto positivo que tiene la entrega de desayunos, refrigerios o almuerzos sobre la asistencia escolar.<sup>5</sup> Dentro de la teoría del cambio, del mismo ejemplo se puede proponer también que la entrega de refrigerios es un incentivo para que las familias decidan enviar a los niños a la escuela ya que perciben un beneficio adicional, más allá de la educación del niño.

Es mucho más rentable hacer el ejercicio de definir la teoría del cambio durante el proceso de diseño del proyecto de inversión social, que una vez iniciada la implementación de este. Pensar en los mecanismos que conectan las actividades con los resultados deseados aporta información relevante para el diseño

y implementación del programa, y también para su evaluación. Por ejemplo, si la teoría del cambio propone que el desempeño escolar aumentará debido a la disminución del ausentismo producto de la entrega de refrigerios, será clave que el programa tenga un componente de comunicaciones diseñado para informar a los padres que a los niños se les dará ese refrigerio.

**C. DEFINIR LA CADENA DE RESULTADOS**

La cadena de resultados es un modelo esbozado para formalizar la teoría del cambio. Si bien en la literatura existen otros modelos, en esta guía se utiliza la cadena de resultados por ser un esquema claro y práctico para presentar la teoría de cambio de proyectos de inversión social.

Este tipo de esquema describe la secuencia de los insumos, actividades, productos y resultados (de corto, mediano y largo plazo), que están detrás de la lógica causal de un programa. Es decir, resume la cadena de factores del programa que, en teoría, causan los impactos observados.

**LA FIGURA 4 DEFINE CADA ESLABÓN DE ESTA CADENA.**



<sup>3</sup> Una revisada cautela: no todas las revistas especializadas son de la misma calidad. Las publicaciones científicas de calidad se distinguen porque los artículos que publican deben pasar el filtro de una revisión realizada por pares.

<sup>4</sup> Ver: Mullainathan, Shafir (2013).

<sup>5</sup> Ver: De Henriquez, Rieser (2004).

<sup>6</sup> Esquemas sencillos, modelizables y medianos de producto, socioeconómico.

FIGURA 4

## ESQUEMA BASE DE UNA CADENA DE RESULTADOS



Esta cadena tiene tres partes. La primera, son los **supuestos** y riesgos detrás de la teoría del cambio. Esta parte, contiene el cuerpo de evidencia ya existente que soporta la lógica de causalidad entre los insumos, actividades, productos y resultados. También incluye una revisión de evidencia sobre los riesgos que pueden comprometer el funcionamiento y efectividad del programa. La segunda parte está compuesta por los eslabones relacionados con la implementación del programa: **insumos, actividades y productos**. Estas son las áreas en las que el operador puede monitorear y medir directamente el desempeño del proyecto. Finalmente, la tercera parte de la cadena son los **resultados intermedios y finales**. A diferencia de los eslabones previos, los resultados no son controlados directamente por el operador. En efecto, los resultados de corto, mediano y largo plazo dependen del comportamiento de los beneficiarios y de la manera como éstos reaccionan al programa.<sup>7</sup>

<sup>7</sup> Para mayor detalle ver Bertar, Martínez, Pizmant, Ruvalcaba y Vennemann, (2011).



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

RECUADRO 10

**"MÁS NUTRICIÓN, MÁS EDUCACIÓN"****SUPUESTOS** - Endencia:

- Estudios clínicos sobre el impacto nutricional sobre el desarrollo cerebral y de habilidades cognitivas.
- Estudios sobre la efectividad de programas similares.

**INSUMOS**

Presupuesto para la compra de refrigerios, personal de apoyo logístico, equipo de monitores y evaluación, sistema de distribución.

**ACTIVIDADES**

Reuniones de planeación, reuniones de coordinación con rectores, talleres de manejo de alimentos, eventos con padres de familia, entrega semanal de refrigerios.

**PRODUCTOS**

Estudiantes atendidos y refrigerios entregados.

**RESULTADOS INTERMEDIOS**

Nutrición de los estudiantes.

**RESULTADOS FINALES**

Graduación de primaria y secundaria, vinculación a educación terciaria, resultados en las pruebas SABER de matemáticas y lenguaje del ICFES.

**d. SELECCIONAR LOS INDICADORES DE DESEMPEÑO**

Una vez se ha determinado la cadena de resultado, se debe seleccionar el conjunto de indicadores

para cuantificar los cambios (avances o retrocesos) en las variables de interés. La cadena de resultados sirve de guía para establecer la lista de variables de interés, y luego seleccionar

indicadores que midan el desempeño del programa en esas dimensiones. En la literatura se destacan 5 características fundamentales de un buen indicador de desempeño:



FIGURA 5:

## CARACTERÍSTICAS FUNDAMENTALES DE UN BUEN INDICADOR DE DESEMPEÑO



Es importante señalar que siempre es útil incluir indicadores correspondientes a todos los estabones de la cadena de resultados. De esta manera se logra tener información para que, además de medir el impacto, se pueda evaluar qué partes de la cadena explican el impacto. Por ejemplo, si en el caso del programa "Más nutrición, más educación" se encuentra que no hay un impacto sobre los indicadores de nutrición en la mitad de las escuelas beneficiarias, se debe revisar si hubo diferencias en los indicadores

correspondientes a los estabones relacionados con la implementación del programa. Con esta información se pueden responder preguntas como: ¿Llegaron los refrigerios a estos colegios dentro del cronograma planeado y con las condiciones de calidad requeridas? ¿Se entregaron en efecto los refrigerios a los niños? Si se encuentra que el programa no se ejecutó según el plan, no tiene mucha sentido evaluar su impacto. Se debería primero concentrar los esfuerzos en ajustar los procesos y la implementación.



## REFERENCIAS PARA LA CONSTRUCCIÓN DE INDICADORES

- Gertler, P., Martinez, S., Premand, P., Rawlings, B. y Vermeersch, C. (2011). *Impact evaluation in practice*. Washington, DC: World Bank.
- Roberts D. y Khaltri N. (2012). *Results Framework for Achieving Results: A How to Guide*. Independent Evaluation Group, Strategy, Learning, and Communication International Bank for Reconstruction and Development - World Bank.
- Yemini, Graham (2013). *Essential Guide to Leading Your Team: How to Set Goals, Measure Performance and Reward Talent*. Pearson Education, pp. 37-39. ISBN 0273772449. Retrieved 2013-07-05.
- Lawler, John; Bilson, Andy (2013). *Social Work Management and Leadership: Managing Complexity with Creativity*. Routledge, pp. 84-85. ISBN 1135247056. Retrieved 2013-07-14.



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

## 7.2 EL DISEÑO METODOLÓGICO DE LA EVALUACIÓN

Luego de presentar las cuatro actividades preparatorias, en esta sección se discuten los retos principales en el diseño metodológico de una evaluación de impacto. Sin duda, el reto principal es construir un grupo de control que sea válido. Tal como se ha explicado antes, para poder medir el impacto exclusivamente atribuible a un programa –y no a otros factores ajenos a la inversión social–, es necesario contar con un grupo de control que sirva de contra-factual. A continuación, se hace una explicación más profunda sobre la importancia del método a través del cual se define el grupo de control, para después pasar a describir 5 métodos de evaluación de impacto. Como se verá, estos métodos se diferencian precisamente por los criterios o reglas utilizadas para definir el grupo de control.

Para realizar una evaluación de impacto sería ideal poder observar: [1] el grupo de beneficiarios en presencia del programa y [2] el mismo grupo de beneficiarios en su ausencia. En esta situación ideal, se tendría un contra-factual perfecto: bastaría con comparar los resultados de los dos escenarios para medir el cambio atribuible al programa. Pero esta situación ideal no existe literalmente imposible, en

el mismo momento del tiempo, observar los dos escenarios.

Los diferentes métodos de evaluación de impacto proponen alternativas para encontrar un contra-factual válido. El criterio para establecer la validez de un contra-factual es la similitud entre el grupo de control y el grupo de beneficiarios, antes de que el programa se ponga en marcha. Lo que se busca es que, justo antes de que la población sea intervenida por el programa, los dos grupos sean estadísticamente idénticos. Cuando esta condición no se cumple, se dice que existe un sesgo de selección que hace imposible evaluar el impacto del programa.

El sesgo de selección no es un problema menor. Para poder concluir que las transformaciones positivas observadas son atribuibles al programa, debe asegurarse que los resultados no sean atribuibles a diferencias preexistentes entre los grupos de beneficiarios y de control. Si, a través del proceso de selección, se incluyen en el grupo de beneficiarios individuos con una mayor propensión a mejorar en las dimensiones relevantes a la evaluación, entonces el impacto será sobre-estimado. Por el contrario, si se incluyen en el grupo de beneficiarios individuos con una menor propensión a mejorar, el impacto

encontrado será sub-estimado (VER RECUADRO 11).

A continuación, se presentan brevemente 5 métodos de evaluación de impacto, comenzando por el que es considerado el "estándar de oro". El objetivo es presentar la lógica detrás cada método para construir el contra-factual, y explicar los contextos en los que se pueden usar, los supuestos necesarios y las limitaciones que tienen los resultados.

### a. EL "ESTÁNDAR DE ORO": ENSAYOS ALEATORIZADOS CONTROLADOS

El reto metodológico de cualquier evaluación de impacto es reducir al máximo el sesgo de selección. En un ensayo aleatorizado controlado (randomized controlled trial, o RCT por sus siglas en inglés) se logra este propósito al hacer la asignación de los grupos beneficiarios y de control usando un proceso probabilístico, como por ejemplo una lotería. Así, en un RCT se asegura que no haya diferencias sistemáticas entre los dos grupos antes de poner en marcha el programa: después de hacer una lotería lo único que explica la asignación al grupo de beneficiarios



y al de control es el azar. Bajo estas condiciones es extremadamente simple el cálculo del impacto:

basta con calcular la diferencia en los resultados entre los dos grupos. En este caso no hay duda

de que esta diferencia es atribuible al programa, y no a diferencias pre-existentes entre los dos grupos.<sup>4</sup>

#### RÉCUADRO 11

### "MÁS NUTRICIÓN, MÁS EDUCACIÓN"

Durante los primeros 2 años de funcionamiento del programa de alimentación escolar "Más nutrición, más educación" el proceso de selección de las escuelas beneficiarias siguió el siguiente protocolo:

- La gobernación comunica a todas las escuelas del departamento sobre la convocatoria del programa.
- Las primeras 40 escuelas que llenan los formatos en línea y entregan los documentos de soporte en la alcaldía, son pre-seleccionadas.
- A través de una visita a los escuelas pre-seleccionadas, se verifica que tengan la infraestructura necesaria para almacenar y refrigerar

los alimentos de por lo menos dos semanas.

En la primera evaluación de impacto del programa se decidió crear el grupo de control seleccionando el azar 40 escuelas que no fueron seleccionadas a través de este protocolo. En la siguiente tabla se encuentran algunas características básicas de los dos grupos de escuelas, las beneficiarias y las de control. Estos datos corresponden a lo observado antes de que el programa se pusiera en marcha en 2005.

¿Es el grupo de control realmente comparable con el grupo de escuelas beneficiarias? Es decir, ¿es un contra-factual

válido? Si se usa este contra-factual, ¿habría un sesgo de selección importante que lereste credibilidad a la evaluación de impacto?

La evidencia presentada es contundente: antes de comenzar el programa se observan diferencias importantes entre los dos grupos. Por ejemplo, el hecho de que una mayor proporción de estudiantes en el grupo de control sean socio-económicamente vulnerables puede llevar a que al usar este contra-factual se sobre-estime el impacto del programa: los estudiantes de las escuelas beneficiarias probablemente tienen hogares más educados donde se da un mayor apoyo al aprendizaje.

CARACTERÍSTICAS	BENEFICIARIOS	CONTROL	DIFFERENCIA
Número de estudiantes	800	400	-400
Estudiantes vulnerables	60%	80%	-20%

<sup>4</sup> Si bien PCT es el procedimiento metodológico apropiado para minimizar las diferencias sistemáticas entre el grupo de control y el de tratamiento, es posible que aún dentro de la misma, éstas se sigan presentando: sin embargo, esto algo que se puede "corregir" más adelante en el momento de la estimación del impacto del programa.



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

FIGURA 6



RECUADRO 12

**"MÁS NUTRICIÓN, MÁS EDUCACIÓN"**

En el tercer año de funcionamiento del programa de alimentación escolar "Más nutrición, más educación" se cambió el proceso de selección de las escuelas beneficiarias para hacerlo compatible con la implementación de un ensayo aleatorizado controlado para medir el impacto. El protocolo de selección quedó así:

- La gobernación comunica a todas las escuelas del departamento sobre la convocatoria del programa.
- Las primeras 20 escuelas que llenan los formularios en línea y entregan los documentos de

soporte en la alcaldía, son pre-seleccionadas.

- A través de una visita a las escuelas pre-seleccionadas, se verifica que tengan la infraestructura necesaria para almacenar y refrigerar los alimentos de por lo menos dos semanas.
- Se realiza una lotería para seleccionar el azar 10 escuelas beneficiarias a partir de las 20 pre-seleccionadas. Las 10 no seleccionados automáticamente se asignan al grupo de control.
- Usando la información recogida en la línea de base

se revisa que no existan diferencias significativas entre los dos grupos. La asignación por lotería se debe repetir si esta condición no se cumple.

En este caso hay certeza sobre la validez del contra-factual ya que se puede esperar que no exista sesgo de selección. Todas las diferencias observadas en los resultados académicos de los estudiantes en las escuelas beneficiarias, en comparación con las escuelas del grupo de control, serán completamente atribuibles al programa.



## RECUADRO 13

## RCT EN SÍNTESIS

Estas son las condiciones necesarias para poder realizar una evaluación de impacto usando un RCT:

- El operador –quien implementa el programa– debe tener completa autonomía para seleccionar a los beneficiarios. El núcleo de este método de evaluación es la asignación

aleatoria ex ante a los grupos beneficiario y de control.

- El universo de beneficiarios potenciales al programa debe ser por lo menos dos veces más grande que el número planeado de beneficiarios.
- Debe existir un estudio de línea de base, ojalá realizada antes

de asignar la población potencialmente beneficiaria a los grupos de tratamiento y control.

- Debe ser posible hacer por lo menos un seguimiento a los grupos de beneficiarios, una vez transcurrido el tiempo necesario para que el programa tenga impacto.

## RECUADRO 14

## PROS Y CONTRAS DEL RCT

- **PROS:** gracias a la asignación aleatoria, se puede esperar que no existan diferencias pre-existentes entre los grupos de tratamiento y control y, por ende, que el estimador del impacto no sufra de sesgos por selección. Lo más potente de este método es que: **[1]** no depende de supuestos difíciles de comprobar, **[2]** no se requieren una gran cantidad de datos [en comparación con los otros métodos] y **[3]** la estimación del impacto depende de procedimientos aritméticos básicos.

- **CONTRAS:** en la práctica el gran reto para implementar este método es lograr que la asignación al grupo de beneficiarios se haga a través de una lotería. Visualmente la razón detrás de esto es que el criterio de selección aleatoria va en contra de los intereses u objetivos de quien financia el proyecto de inversión social.

**b. EMPAREJAMIENTO POR:****PUNTAJE DE SIMILITUD:**

Si la decisión de realizar una evaluación de impacto se toma cuando el programa ya ha sido implementado, ya no es posible la construcción de un grupo de control a través de una lotería (como en el RCT). El emparejamiento por puntaje de similitud (Propensity Score Matching, PSM por sus siglas en inglés) permite construir un grupo de control apropiado para cada unidad o grupo

de tratamiento, mediante el emparejamiento de cada beneficiario con un "clon".

¿Y esto, cómo se hace? El criterio para realizar este emparejamiento es: la probabilidad de que un individuo, dadas sus características observables antes de ponerse en marcha el programa, fuese seleccionado como beneficiario. Esta probabilidad, que se conoce en la literatura como "puntaje de similitud" es una predicción,

y se obtiene usando métodos estadísticos avanzados.

Para identificar un grupo de control se requiere estimar ese puntaje de similitud para todos los beneficiarios y no beneficiarios con base en las características observadas en el periodo justo antes de que inicio el programa.

El PSM es válido sólo si se cumplen dos condiciones. La primera es que existe un soporte común entre los dos grupos. Esto quiere



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

dicir que en el universo de donde se prende elegir el contra-factual existan individuos que tengan una probabilidad de ser seleccionados al programa similar a la de los beneficiarios. La segunda condición tiene que ver con la existencia de

variables no observadas; esto es, características de la población que no están medidas y por ende no se encuentran en la base de datos disponible para la evaluación. En particular, la segunda condición exige que no existan variables no observadas fóbicamente

no incluidas en el proceso de emparejamiento, ya que no las medimos, que sean determinantes del proceso de selección al grupo de beneficiarios. El cumplimiento de estas dos condiciones se realiza utilizando pruebas estadísticas especiales.

FIGURA 7



RECUADRO 15

**"MÁS NUTRICIÓN, MÁS EDUCACIÓN"**

El proceso de selección de las escuelas beneficiarias del programa de alimentación escolar "Más nutrición, más educación" pasó en el cuarto año a manos de la Secretaría de Educación de la Gobernación de Cundinamarca. El equipo de monitoreo y evaluación, que entró a evaluar el programa un año después, no participó en ese proceso de selección. Sin embargo, existía una base de datos con información sobre características de las escuelas que le permitió usar el método de PSM para construir un contra-factual válido para hacer la evaluación de impacto del programa.

La información disponible de las escuelas del Departamento era la siguiente: (1) tamaño:

número de estudiantes y sedes; (2) infraestructura: número de baños, número de salones, disponibilidad de refrigerador; (3) puntajes de los estudiantes en la prueba estatal de matemáticas; (4) tasa de deserción; y (5) proporción de estudiantes de hogares vulnerables.

## PARA CONSTRUIR EL CONTRA-FACTUAL:

- Con base en la información disponible de beneficiarios y no beneficiarios, se estimó la probabilidad de que una escuela hubiera sido seleccionada para ser beneficiaria usando un modelo estadístico. Esta probabilidad es el "puntaje de similitud".
- Se comprobó que existían escuelas no beneficiarias que tenían una probabilidad de ser seleccionadas similar a la de las beneficiarias antes de la implementación del programa, y que no había variables no incluidas en el círculo del puntaje de similitud que hubieran determinado la selección al programa.
- Se asignó a cada una de las 10 escuelas beneficiarias una escuela no beneficiaria muy cercana en términos de su "puntaje de similitud". Ese grupo de escuelas no beneficiarias cercanas, es el que luego se utilizó en la estimación de impacto como contra-factual.



## RECUADRO 16

**PSM EN SÍNTESIS**

Estas son las condiciones necesarias para poder realizar una evaluación de impacto usando un PSM:

- Tener amplia información sobre la población beneficiaria y no beneficiaria en línea de base. Esta información es clave para construir el puntaje de similitud.

- Tener un universo suficientemente grande de potenciales beneficiarios del programa. ¿Qué tan grande es "suficientemente grande"? depende de qué tan similares son los beneficiarios y no beneficiarios en línea de base. En otras

- palabras, de qué tan grande es el soporte común.

- Que exista un soporte común.
- Que no haya variables no observadas, que hayan sido determinantes al momento de seleccionar al grupo de beneficiarios.

## RECUADRO 17

**PROS Y CONTRAS DEL PSM**

- **PROS:** mediante el empalmamiento por puntajes de similitud, es posible identificar el impacto del programa de reversión social sin necesidad de que el diseño de la evaluación sea experimental.

- **CONTRAS:** en ocasiones la población beneficiaria se elige de un modo tal que aquella que queda por fuera del programa, a partir de la cual se construiría el grupo de control, simplemente no es comparable. Es decir, hay casos en que no existe un soporte común. Cuando esto ocurre, la metodología es inaplicable. Una segunda desventaja es que su aplicación requiere información sustancial acerca de las características de la población beneficiaria y no beneficiaria.

**c. DIFERENCIAS EN DIFERENCIAS**

El método de diferencias en diferencias (differences-in-differences, o DD), tiene en cuenta las diferencias preexistentes entre los grupos de beneficiarios y de control para aislar los efectos atribuibles exclusivamente al programa. En contraste con el PSM, cuando se aplica una metodología de DD se asume que antes de la puesta en marcha del programa existían

diferencias importantes en características no observables (u observables) entre beneficiarios y no beneficiarios que no cambian en el tiempo. La metodología del DD permite controlar por estas diferencias directamente en la estimación, para asegurar que el impacto estimado del programa no sea "contaminado" por ellas. Por este se les conoce a veces como

variables contaminantes.

Para estimar el impacto con DD se debe poder observar a los mismos individuos en por lo menos dos momentos del tiempo: antes y después de la implementación del programa.

Una práctica común es la combinación de los dos métodos, el PSM y el DD, que da origen al método compuesto que se conoce en la literatura como



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

comparo/jarmento por puntaje de similitud con diferencias en diferencias (PSM-CD). Esta combinación permite aprovechar las fortalezas

de los dos métodos: por una parte, elimina las diferencias sistemáticas en las características no observables de los grupos

beneficiario y de control que no cambian en el tiempo y, por otra, utiliza el puntaje de similitud para construir un grupo de control comparable.

## RECUADRO 18

**"MÁS NUTRICIÓN, MÁS EDUCACIÓN"**

En el cuarto año del programa "Más nutrición, más educación" se decidió evaluar el impacto del programa sobre el peso y talla de los estudiantes que habían ingresado al programa en la primera oleada. Para esto se contrató una medición del peso y la talla de los estudiantes de los colegios que fueron elegibles en el primer año cuando el programa se puso en marcha. Esta medición se incluyó

tanto a los estudiantes pertenecientes a los colegios asignados al grupo de beneficiarios como a los del grupo de control.

El efecto del programa se estimó usando un modelo de DD. En esta estimación se atribuyó al programa el cambio en los indicadores de impacto entre el 1er y el 4to año que no estuvo explicado ni por efectos de tiempo (variables que cambiaron en el

tiempo de la misma manera para todos los estudiantes), ni por efectos fijos (características de cada estudiante que no cambiaron en el tiempo, como su sexo, el nivel máximo alcanzado de su madre y padre o el estrato socioeconómico de su vivienda). En algunas estimaciones se incluyó como control la edad de los estudiantes. Se encontró que esta variable no afectó el impacto del programa.

## RECUADRO 19

**DD EN SÍNTESIS**

Estas son las condiciones necesarias para poder realizar una evaluación de impacto usando un DD:

- Tener dos mediciones de los indicadores de impacto, una antes de

la puesta en marcha del programa y otra un período después, para la población elegible tanto beneficiaria como no beneficiaria del programa.

- Tener dos mediciones de las características de la población elegible que cambian en el tiempo, que pudieran incidir sobre el impacto del programa.



## RECUADRO 20

## PROS Y CONTRAS DEL DD

- **PROS:** mediante la estimación de un modelo de DD, es posible identificar el impacto del programa de inversión social sin necesidad de que el diseño de la evaluación sea experimental y sin que la población beneficiaria y de control sean idénticas en sus características observables antes de la puesta en marcha del programa.
- **CONTRAS:** en ocasiones la población beneficiaria y no-beneficiaria difieren en características que cambian en el tiempo y que tienen una relación con el indicador de impacto. Si estas características se omiten en la estimación, puede sobre-estimarse o subestimar el impacto del programa (dependiendo del signo de la relación que tenga la variable omitida con el indicador de impacto).

## d. VARIABLES INSTRUMENTALES

El método de variables instrumentales (Instrumental Variables o IV por su sigla en inglés) se basa en la identificación de una variable que esté altamente correlacionada con la probabilidad de ser seleccionado como beneficiario; y que no esté correlacionada con características no observables del individuo

que afecten directamente los impactos deseados. En otras palabras, la "variable instrumental" es una variable que no determina los resultados que se quieren afectar con el programa, pero sí la probabilidad de ser beneficiario. Esta variable se utiliza en la estimación de impacto como sustituto de la variable que indicaría participación en el programa y corrige la

estimación por posibles sesgos de selección.

El uso de IV es útil para asegurar que la estimación de impacto no se ve contaminada por este tipo de sesgos también cuando la asignación al programa es aleatoria, pero por algún motivo los grupos de tratamiento y control no resultan estadísticamente idénticos en el 100% de sus características.

## RECUADRO 21

## "MÁS NUTRICIÓN, MÁS EDUCACIÓN"

En el sexto año de funcionamiento del programa de alimentación escolar "Más nutrición, más educación", la Secretaría de Educación de la Gobernación de Cundinamarca continuó liderando el proceso de selección de las escuelas beneficiarias. A diferencia de los años anteriores, para este cohorte no se cuenta con un estudio de línea de base que describa las características de las escuelas

en el Departamento. Sin embargo, se sabe que para elegir los beneficiarios del programa se siguió el criterio de "mayor alcance". Así, la Secretaría decidió seleccionar a las escuelas más grandes, que además cumplieran con los requisitos previos de inscripción y trámite.

En este caso el equipo de monitoreo y evaluación del programa puede utilizar como variable instrumental, el número de

estudiantes de cada escuela. Por un lado, se sabe que esta variable está altamente correlacionada con la probabilidad de que una escuela haya sido seleccionada por la Secretaría. Por otro lado, se debe asumir que el número de estudiantes no es un determinante del desempeño escolar. Si la calidad de la educación es similar en colegios grandes y pequeños, este supuesto puede ser aceptable.



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

RECUADRO 22

## IV EN SÍNTESIS

Esta es la condición necesaria para poder realizar una evaluación de impacto usando IV:

- Identificar una variable que esté bien correlacionada con la probabilidad de ser beneficiario del programa, que no esté correlacionada con variables no observadas que puedan afectar los indicadores de impacto.

RECUADRO 23

## PROS Y CONTRAS DE LAS IV

- PROS:** cuando se dispone de una variable instrumental adecuada, es posible identificar el impacto del programa de inversión social sin necesidad de que el diseño de la evaluación sea experimental.

- CONTRAS:** la posibilidad de emplear una metodología de IV depende de que exista en efecto una variable bien correlacionada con la probabilidad de ser beneficiario del programa, que no esté correlacionada con variables no observadas que puedan afectar los indicadores de impacto. Este tipo de variables con frecuencia son difíciles de encontrar. Existen pruebas estadísticas que permiten identificarlas.

## c. REGRESIÓN DISCONTINUA

El método de regresión discontinua (Regression Discontinuity, o RD) es similar en espíritu al método de IV, para elegir el contra-factual se utilizan variables que están altamente correlacionadas con la participación en el programa, pero que no

explican las diferencias en los indicadores de impacto. Con frecuencia se utilizan reglas de elegibilidad que pueden ser usadas como instrumentos para predecir qué individuos son beneficiarios. En particular, en el método RD se utiliza como criterio de similitud o comparabilidad la cercanía

a un determinado umbral de elegibilidad. Así, este método busca comparar solamente a los individuos que están alrededor de un umbral que ha sido definido para determinar la localización del programa. El ejemplo práctico que se presenta en el RECUADRO 24 sirve para ilustrar esto.

FIGURA 8 REGRESIÓN DISCONTINUA



## RECUADRO 24

**"MÁS NUTRICIÓN, MÁS EDUCACIÓN"**

En el octavo año de funcionamiento de "Más nutrición, más educación" el equipo de monitoreo y evaluación del programa convenció a la Secretaría de

Educación de crear una regla objetiva para la selección de las escuelas beneficiarias. En particular, se decide que recibirán el programa de alimentación

escolar los colegios con más de 500 estudiantes. En este caso, el equipo de monitoreo y evaluación tiene un buen argumento para usar el método RD.

## RECUADRO 25

**RD EN SÍNTESIS**

Esta es la condición necesaria para poder realizar una evaluación de impacto usando RD:

- Que la elegibilidad para ser beneficiario del programa esté asociada con superar un umbral en un indicador observable, dado por una regla preestablecida

## RECUADRO 26

**PROS Y CONTRAS DE LA RD**

- **PROS:** cuando existe una regla de elegibilidad clara que permite identificar un umbral en algún indicador por encima del cual se participa en el programa y por debajo del cual no se participa, es posible identificar el impacto del programa sin necesidad de que el diseño de la evaluación sea experimental.

- **CONTRAS:** la limitación más importante de esta metodología es que el impacto que se identifica al usar RD corresponde únicamente a las diferencias entre los individuos que están alrededor de un umbral. Esto quiere decir que el método evalúa únicamente el impacto promedio para un grupo particular: quienes están justo por debajo y justo por encima del umbral. Cuando las reglas de elección no se cumplen sistemáticamente o varían en el tiempo, esta metodología es inaplicable.



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

### 7.3 CÓMO ESCOGER LA METODOLOGÍA DE EVALUACIÓN?

En esta sección se discuten los factores que se deben tener en cuenta para escoger, en la práctica, entre las diferentes alternativas metodológicas.

Usualmente, la regla de selección de los beneficiarios es el factor más relevante a la hora de escoger la metodología de evaluación de impacto. En particular, los pasos y criterios para decidir la población beneficiaria –es decir, la regla de focalización

del programa– limita las maneras en la que es posible construir el grupo de control. Por ejemplo, si el proyecto de inversión social está diseñado para beneficiar a todos los habitantes de un barrio en especial, es imposible hacer una asignación por lotería para definir un grupo de control a partir de hogares o personas del mismo barrio. En este caso la regla de focalización, además de definir el barrio, limita la capacidad de construir un contra-factual utilizando un criterio de aleatorización.

#### REGLA DE FOCALIZACIÓN

Pasos y criterios para definir la población beneficiaria

Casi siempre es posible encontrar una metodología para definir el grupo de control y evaluar impacto cuando la regla de focalización cumple con los siguientes tres criterios: equidad, transparencia y responsabilidad (VER FIGURA 9).<sup>2</sup>

FIGURA 9

#### CRITERIOS DESEABLES EN UNA REGLA DE FOCALIZACIÓN

##### TRANSPARENCIA

Una regla de focalización transparente permite que cualquier agente vedor pueda monitorear que ésta se esté cumpliendo. Para esto, la regla debe ser fácil de observar, medir y entender para quien lidera el proceso de monitoreo.

##### EQUIDAD

Una regla de focalización equitativa logra que todos los beneficiarios potenciales, que comparten las mismas características, tengan la misma oportunidad o probabilidad de ser seleccionados.



##### RESPONSABILIDAD

Una regla de focalización con responsabilidad está claramente asociada a los deberes e indicadores de desempeño del equipo, encargado de lograr que el proceso de selección sea exitoso.



<sup>2</sup>Véase Gerber et al. (2011).

Cuando ninguno de estos criterios se cumple, se pone en riesgo tanto la viabilidad del programa como su evaluación. Por el contrario, cuando por lo menos uno de éstos se cumple, generalmente es posible utilizar alguna de las metodologías de evaluación de impacto descritas en la sección 4.2.

Hay tres factores que usualmente determinan las reglas de focalización de un programa:

- **FACTOR 1: RECURSOS DISPONIBLES.** Es muy común que los recursos financieros no sean lo suficientemente abundantes como para lograr la cobertura universal de un programa. En otras palabras, es normal que el presupuesto disponible no sea suficiente para beneficiar a toda la población elegible. Cuando este factor está presente, es necesaria la focalización. Es decir, necesariamente se debe limitar la entrega del programa a un área geográfica particular (por ejemplo, un municipio, ciudad, barrio, colegio) o un grupo especial de la población (los hogares más vulnerables, los niños

en un grupo de edad particular o los trabajadores de una industria particular).

- **FACTOR 2: CRITERIOS TÉCNICOS DE ELEGIBILIDAD.** Si bien hay programas que buscan ser universales –por ejemplo, un programa de prevención de violencia intrafamiliar–, por su naturaleza y objetivos algunas intervenciones deben ser dirigidas únicamente a una población especial. De la misma manera que un programa de estimulación temprana es relevante solamente para niños menores de 5 años, un programa de salud mental es útil solamente para la población que muestra síntomas de enfermedad mental en regiones donde se tiene una alta probabilidad de sufrirlos (como los municipios más afectados por el conflicto armado). Con frecuencia, estos criterios se convierten en umbrales que marcan claramente quién es elegible para el programa. Muchas veces existe un índice o punto que ordena a la población según el

criterio de elegibilidad. En estos casos se determina un punto de corte a partir del cual se clasifican unos como elegibles, y el resto como no elegibles. Por ejemplo, la selección de beneficiarios del programa colombiano Familias en Acción depende del puntaje SISBEN, que es un indicador creado por el Estado utilizando información sobre los ingresos, activos y características demográficas del hogar para dar una medida objetiva de vulnerabilidad socio-económica. Este puntaje permite hacer un ordenamiento de los hogares elegibles según su necesidad de recibir el programa. Si, por ejemplo, se decide que el punto de corte son los 100 puntos, todos los hogares con un puntaje igual o superior a 101 serían excluidos del grupo de beneficiarios.

- **FACTOR 3: LOS TIEMPOS DEL PROGRAMA.** La linea de tiempo de entrega del programa es el tercer factor clave para la definición de una regla de focalización. Algunos proyectos de



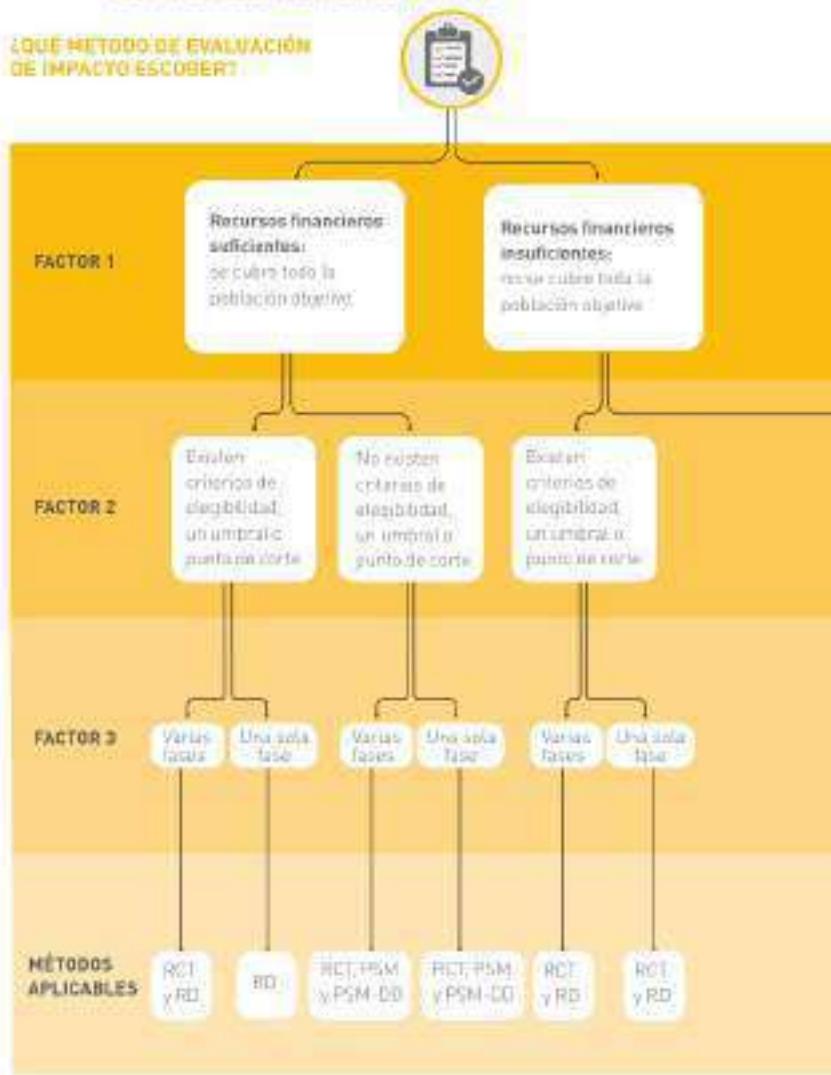
## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

inversión social se desarrollan secuencialmente por fases, por restricciones de capacidad administrativa u operativa. En estos casos, el diseño mismo del proyecto implica que existe un orden de entrega de los beneficios del programa, es decir, debe decidirse qué región o grupo de la población va primero, segundo, tercero, ... etcétera. En contraste, otros proyectos se desarrollan entregando los beneficios a la población seleccionada en un mismo momento, no por fases.

Suponiendo que los criterios deseables en una regla de focalización se cumplen (equidad, transparencia y responsabilidad), estos tres factores limitan el número y tipo de metodologías de evaluación de impacto que se pueden llevar a la práctica. Como herramienta para la elección del método de evaluación de impacto la FIGURA 10 presenta un árbol de decisiones que resume las posibles situaciones según la forma que tomen estos factores.

FIGURA 10 ÁRBOL DE DECISIONES PARA SELECCIONAR EL MÉTODO DE EVALUACIÓN DE IMPACTO

¿QUÉ MÉTODO DE EVALUACIÓN DE IMPACTO ESCOGER?



Según este árbol de decisiones, existen 8 situaciones posibles, en donde varía el conjunto de metodologías aplicables.

#### 7.4 RECOLECCIÓN DE DATOS

Una vez elegido el método -que a su vez define la manera en la que se construye el grupo de control- es el momento para determinar qué información se necesita para implementar la evaluación de impacto. En particular, es clave definir: **(1)** los indicadores necesarios para la evaluación y **(2)** la muestra para la cual se van a recoger los datos. En esta sección se discuten brevemente los criterios para decidir qué tipos de datos se necesitan y cómo se debe construir una muestra para dar validez a los resultados de la evaluación.

##### a. ¿QUÉ TIPO DE DATOS RECOGER?

Idealmente, en una evaluación de impacto se busca recoger datos que dan cuenta de todos los eslabones en la cadena de resultados (ver Sección 4.1). Sin embargo, existe un orden de prioridad para establecer cuáles son los datos más importantes para llevar a cabo este tipo de evaluación. Como se discute a continuación, algunos de



estos datos son esenciales:

##### • INDICADORES DE RESULTADO

Es imposible llevar a cabo una evaluación de impacto sin la información que permita medir las variables que se quieren impactar -los indicadores de resultado-. Recoger datos para medir indicadores de resultado en las etapas de seguimiento del programa es un mínimo necesario. No obstante, siempre es recomendable recoger esta información tanto en línea de base como en los seguimientos. La razón detrás de esta recomendación es que, solamente si se tiene una medición en línea de base (además del seguimiento), es posible aprovechar las ventajas del método de Diferencias en Diferencias (DD).

##### • INDICADORES SOBRE LA IMPLEMENTACIÓN

Es esencial contar con un mínimo de información sobre la implementación del programa. En particular, se deben tener datos que permitan identificar claramente los beneficiarios y las fechas de inicio y fin del programa. Además, es útil contar con indicadores de insumos y productos que permitan medir la "intensidad" de la



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

implementación. Es decir, la información recogida por el sistema de monitoreo permite identificar si existieron diferencias en las inversiones y beneficios correspondientes a cada beneficiario. Esta variación puede explicar en gran medida las variaciones observadas en los indicadores de resultados.

- INDICADORES DE FACTORES EXTERNOS QUE AFECTAN LOS RESULTADOS

Como se discute en la Sección 6.2, la existencia de variables

'contaminantes' puede dar lugar a imprecisiones en la estimación del impacto. Por este motivo, en la medida de lo posible, es deseable contar con datos que permitan construir indicadores de factores externos al programa que puedan afectar los resultados (características de los beneficiarios y sus hogares, del municipio o localidad en la que habitan, etcétera).

- INDICADORES PARA IDENTIFICAR IMPACTOS POR TIPO DE

BENEFICIARIO. Dependiendo del tipo de proyecto de inversión social, es posible que la rentabilidad o efectividad de la intervención sea mayor para un grupo particular de la población de beneficiarios. En estos casos es importante recoger datos que permitan identificar estos grupos, para luego poder estimar el impacto del programa para cada grupo por separado. Este ejercicio es muy útil en el proceso continuo de ajuste y mejoramiento de los programas.

## RECUADRO 27

**"MÁS NUTRICIÓN, MÁS EDUCACIÓN"**

- INDICADORES DE RESULTADO: medidas antropométricas (peso, talla, medidas de anemia, entre otros), resultados de las pruebas SABER 3, 5, 9 y 11, número de estudiantes que fueron regularmente a la escuela durante el año, número de estudiantes que regresaron en el siguiente año escolar; número de estudiantes que aprobaron el año.
- INDICADORES SOBRE LA IMPLEMENTACIÓN: número de refrigerios entregados a cada estudiante.
- INDICADORES DE FACTORES EXTERNOS QUE AFECTAN LOS RESULTADOS: nivel socioeconómico del hogar, choques al hogar (desempleo, catástrofes naturales, enfermedad, entre otros); migración, fecundidad y participación en otros programas de asistencia social.
- INDICADORES PARA IDENTIFICAR IMPACTOS POR TIPO DE BENEFICIARIO: género, edad, grupo étnico, orden de nacimiento del niño y educación de los padres.



La información que se emplea en la construcción de indicadores puede clasificarse según la fuente de la que se toma como información primaria o secundaria. La información primaria es la que se recoge especialmente para un estudio y se toma directamente de la fuente primaria de información –de ahí su nombre-. Por ejemplo, en una evaluación de un programa de prevención y promoción de salud se puede aplicar una encuesta de hogares para recoger datos primarios sobre el acceso y uso de programas de vacunación de menores. Esta información viene de la fuente primaria (los hogares) y se recoge precisamente para evaluar los resultados del programa.

La información secundaria, en contraste, es aquella que no se recoge especialmente para un estudio particular, pero que en todo caso es útil y relevante para ese

estudio. Una fuente común de información secundaria son las bases de datos administrativas, que recogen información útil para construir indicadores de insumos, productos y resultados. Por ejemplo, como parte de los requerimientos de las Secretarías de Educación, las escuelas solicitan a las familias pruebas que soporten la vacunación de los estudiantes. Esta información, que es recogida en un proceso totalmente ajeno al programa de prevención y promoción de salud, puede servir para construir indicadores de resultado. Otras fuentes de datos secundarios que son útiles son las bases de datos de organismos estatales, como el DANE, los Ministerios, las Secretarías y el ICFES.

Para optimizar el proceso de recolección, siempre vale la pena en primer lugar hacer un barrido de las bases de datos de fuentes secundarias que están públicamente disponibles:

Con respecto a la recolección directa de información primaria, a veces es necesario contratar su levantamiento con externos. Sin embargo, mucha de esta información puede obtenerse a partir de unos registros administrativos bien llevados del programa mismo o a través de mecanismos de recolección que puede implementar la empresa directamente. Lo importante es que la empresa tenga clara la importancia de contar con una información mínima del programa y de la población elegible. En la fase de diseño del programa –o, tardeamente, en el momento de intentar evaluar el impacto de un programa ya puesto en marcha– la empresa debe valorar las alternativas disponibles para la recolección de información y la construcción de indicadores y elegir la que resulte más cost-efectiva y más afin con los objetivos que persigue.



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

**b. ¿CÓMO SELECCIONAR LA MUESTRA?**

La analogía presentada en el próximo recuadro sirve para entender las grandes ventajas de utilizar una muestra -y no la población entera- al hacer una evaluación de impacto. El

objetivo de una muestra es poder representar a la población. Para que esto ocurra, se debe lograr que en la muestra estén proporcionalmente incluidas todos los tipos de individuos o agentes de la población. Es decir, que

todos los sub-grupos estén bien representados. Por ejemplo, si la población objetivo de un estudio son las empresas de Bogotá, y en ésta el 80% son microempresas, el 15% empresas pequeñas y medianas (PYMES) y 5%

## RECUADRO 28

**¿POR QUÉ TENEMOS QUE USAR UNA MUESTRA REPRESENTATIVA Y NO LA POBLACIÓN ENTERA?**

Una muestra representativa es un grupo particular de una población que ha sido seleccionado de tal manera que sus características son idénticas a la población que pretenden representar. En otras palabras, una muestra representativa tiene, en promedio, las mismas características que la población, aunque ésta compuesta por solo una

fracción de ella. En las ocasiones excepcionales cuando se recoge información de toda la población, se produce un censo poblacional.

Imaginemos que queremos evaluar el impacto de un medicamento para reducir el colesterol. Normalmente, para medir el nivel de colesterol en la sangre de una persona, se toma una

muestra de sangre. ¿Por qué no se hace la medición tomando toda la sangre del individuo? La respuesta es: ¡porque es imposible o demasiado costoso! Esta es la gran ventaja de utilizar una muestra representativa, y no un censo de toda la población: se reducen los costos de medición o, en otras palabras, se hace viable la medición.



grandes empresas, en una muestra representativa de las empresas de la ciudad se debería observar esta misma distribución por tamaño de empresas.

Si, por cuestiones del procedimiento de selección, en la muestra observamos solamente que el 20% son microempresas, podemos entonces concluir que la muestra no es representativa. En este caso es evidente que va a estar errada cualquier inferencia sobre la población que se haga a partir de esta muestra.

En una evaluación de impacto –al igual que en cualquier ejercicio de inferencia estadística<sup>11</sup>– interesa que la muestra cumpla dos principios. El primero, es que sea seleccionada a través de un método probabilístico (Ver Moore, McCabe y Craig, 2009). Es decir, que el subgrupo de la población que

se incluya en la muestra sea elegido siguiendo algún tipo de lotería. El gran poder de las loterías es que, si se hacen bien, garantizan que la muestra no tenga sesgo de selección.

El segundo principio es que la muestra permita tener suficiente poder estadístico. El poder estadístico es una medida que nos permite saber qué tan fuerte es la evidencia provista por la muestra, para sacar conclusiones sobre la población. En particular, en una evaluación de este tipo el objetivo es saber si el proyecto de inversión social ha tenido un impacto sobre la población beneficiaria. Así, al diseñar la muestra se debe garantizar que ésta tenga suficiente poder estadístico para concluir si existe o no un impacto en la población entera. Los factores que determinan el poder estadístico son:

- El tamaño de la muestra
- El tamaño del impacto esperado
- La diversidad de la población
- El nivel de error que se tolera en la inferencia estadística

Es importante subrayar que lo que se pretende es tener "suficiente" poder estadístico, no el mayor poder estadístico posible. El propósito es tener una muestra suficientemente grande para: **(i)** capturar suficientemente bien la diversidad de la población, **(ii)** reducir el error en la inferencia a un nivel deseado; e **(iii)** identificarse si existe o no un impacto de un tamaño que sea relevante. Por ejemplo, no vale la pena doblar el tamaño de muestra en la evaluación de "Más nutrición, más educación" para lograr identificar un aumento del 0.0001% en los indicadores de nutrición.

<sup>11</sup> La estadística es la ciencia de aprender de los datos. La inferencia estadística es la ciencia de aprender sobre una población, a partir de los datos obtenidos de una muestra de la población. Entonces para tratar de hacer una inferencia estadística es sacar una conclusión sobre la población, a partir de la información que tiene en su poder la muestra. Ver Moore, McCabe y Craig, 2009.



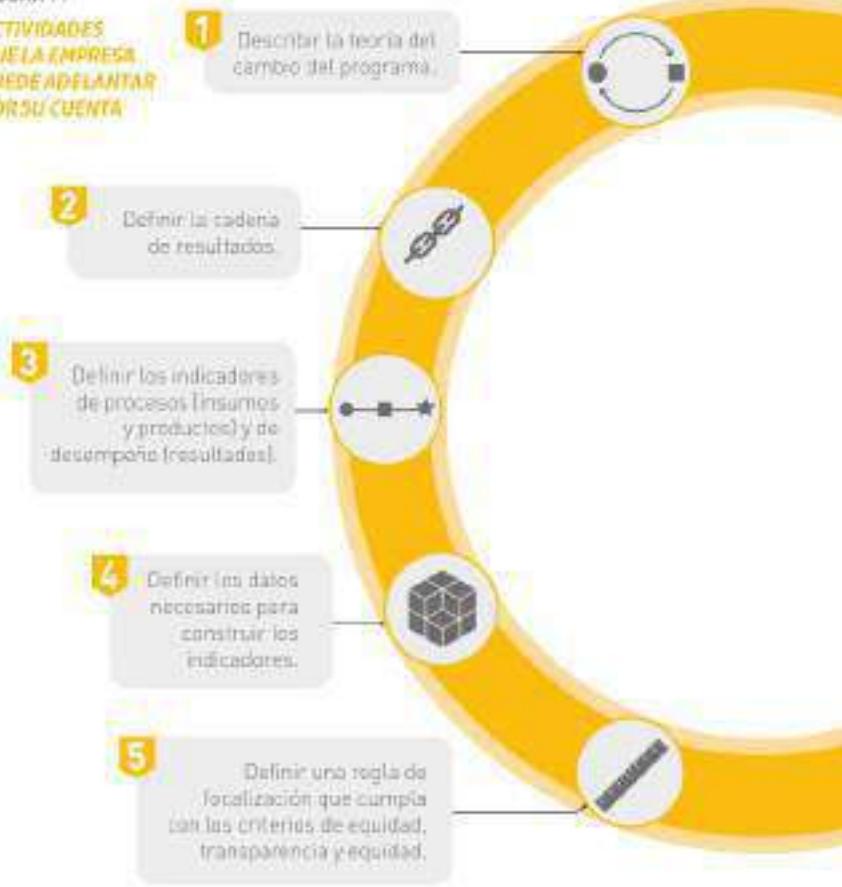
## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

**7.5 MANOS A LA OBRA**

A continuación, se presenta un recuento de las actividades y procesos que puede adelantar una empresa por sí sola para avanzar efectivamente hacia una evaluación de impacto. Más allá de si sea el momento adecuado para una evaluación de este tipo, estas actividades y procesos son en sí mismos útiles para reflexionar sobre el diseño e implementación del proyecto de inversión social. Adicionalmente, se destacan las tareas que, por su nivel de sofisticación, requieren el apoyo de consultores externos expertos en temas particulares y se hace una breve descripción del perfil de cada uno de estos consultores.

Por una parte, en la FIGURA 11 se muestran las actividades que cualquier empresa puede adelantar por su cuenta para avanzar hacia una evaluación de impacto:

FIGURA 11

**ACTIVIDADES QUE LA EMPRESA PUEDE ADELANTAR POR SU CUENTA**



Por otra parte, algunas actividades deberían ser acompañadas por un consultor especializado para asegurar la validez de los resultados de la evaluación de impacto. Esta recomendación se basa en su especificidad y complejidad. A continuación se presenta cada actividad y el perfil del consultor:

- a. DISEÑO DE INSTRUMENTOS PARA LEVANTAR INFORMACIÓN PRIMARIA:** para diseñar los instrumentos de recolección de información, tales como una encuesta; es útil contar con el acompañamiento de un experto con experiencia específica en este campo o con entrenamiento en psicometría.
- b. DISEÑO DE LA MUESTRA:** para determinar el tamaño y composición de la muestra es útil contar con

el acompañamiento de un experto con experiencia en diseño muestral que tenga un buen manejo de las herramientas de la estadística aplicada.

**c. ESTIMACIÓN DEL IMPACTO:** la estimación e interpretación de los resultados de una evaluación de impacto debe ser liderada por un economista experto en econometría aplicada. Si se trata de una evaluación de carácter experimental (un RCT), la tarea puede dejarse en manos de un economista junior que tenga un buen manejo de las técnicas econométricas para la evaluación de impacto. La aplicación de cualquiera de los otros cuatro métodos requiere un mayor nivel de sofisticación técnica.



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

*Negocios de alto  
impacto social:*  
**NUEVAS  
TENDENCIAS**

81



## L

a manera en la que se concibe el impacto social que generan las empresas viene transformándose. Las acciones aisladas a la operación del negocio, a través de las cuales las empresas contribuían a la transformación social de los territorios en los que operaban, que no representaban un retorno para la empresa al no estar alineadas con la estrategia y las dinámicas del negocio, y por tanto no garantizaban la sostenibilidad de la inversión social, vienen siendo sustituidas por nuevos modelos de negocio con impacto social. **VER RECUADRO 29**

Esto ha ocurrido porque el sector privado ha reconocido la interdependencia que existe en-

tra el impacto social y los resultados del negocio: los negocios se benefician de esquemas en los que se fortalece el desarrollo de la cadena de valor o el desarrollo de productos y servicios enfocados en resolver problemáticas sociales y al mismo tiempo, los mecanismos de mercado son una de las herramientas más poderosas para promover el desarrollo social. Adicionalmente, al dirigir la inversión social hacia temas en los cuales la empresa tiene mayor capacidad y conocimiento, se garantiza el aprovechamiento de la experiencia del negocio en pro del desarrollo social y se realizan inversiones pertinentes y más eficientes.



## ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

RECUADRO 29

## TEORÍAS SOBRE CÓMO GENERAR NEGOCIOS

**a. VALOR COMPARTIDO**

- Políticas y prácticas operacionales que mejoran la competitividad de una empresa a la vez que ayudan a mejorar las condiciones económicas y sociales en las comunidades donde opera. La creación de valor compartido se enfoca en identificar y expandir las conexiones entre los progresos económico y social a través de 3 vías:<sup>11</sup>
- Desarrollando nuevos productos o servicios que resuelven retos sociales.
- Logrando mayor eficiencia en la cadena de valor, al apoyar el fortalecimiento de proveedores o distribuidores locales.
- Participando en clusters o agrome-

raciones con empresas, instituciones, entidades públicas y gremios relacionados con un mismo sector para impulsar de forma colectiva la competitividad.

**EJEMPLO:** Una empresa de alimentos del sector lechero busca generar mayor eficiencia en su cadena de valor al fortalecer a las cooperativas lecheras de una zona desprimida del país, que son sus proveedores y así asegurar la calidad de la provisión. Además del fortalecimiento productivo, las cooperativas y sus familias reciben acompañamiento en términos técnicos, comerciales,

organizacionales y en mejoramiento de su calidad de vida.

**b. NEGOCIO INCLUSIVO**

Los Negocios Inclusivos son iniciativas empresariales económicamente rentables, ambiental y socialmente responsables, que en una lógica de mutuo beneficio incorporan en sus cadenas de valor a comunidades vulnerables y mejoran su calidad de vida.<sup>12</sup>

**EJEMPLO:** Una empresa de consumo masivo fortalece a las tiendas de barrio, que son en muchos casos su canal de venta final, para que mejoren en términos administrativos, comerciales y sociales. Gracias a este fortalecimiento las tiendas aumentan sus



Verde de Porter, M. y Werner, M. (2011). | Tomado de Decreto, Díal

## DE ALTO IMPACTO SOCIAL

ventas, y mejoran sus ingresos, al mismo tiempo que le compran más a la empresa de consumo masivo.

### c. NEGOCIO SOCIAL

Son modelos de negocio que buscan aliviar problemáticas sociales o ambientales a través de sus productos o servicios. Estos negocios además de contar con sostenibilidad financiera, maximizan el valor social al reinvertir las utilidades en el mismo negocio.<sup>11</sup>

**EJEMPLO:** Una empresa de alimentos crea un yogur nutricional hecho a la medida de las necesidades alimentarias de una población desnutrida, a un precio asequible. Adicional al beneficio en la salud, la empresa

instala la operación de la planta productora en esa misma población para generar empleo local, compra los insumos de pequeños productores locales y establece un sistema de micro distribución puerta a puerta que da empleo a mujeres de la misma comunidad. Todas las ganancias se reinvierten en el negocio, para ampliar el modelo de negocio. A través de este esquema se mejora la calidad de vida de la población a través de la generación de ingresos y la salud de la población, se incrementan los ingresos de la misma, y se crea un modelo de negocio rentable y sustentable en el tiempo.

En todos los modelos de inversión social –ya sean programas sociales o negocios de alto impacto social–, siempre es deseable la identificación de los resultados esperados, de la teoría del cambio y de la cadena de resultados. Asimismo, para medir si las inversiones avanzan de la dirección esperada y tienen la rentabilidad adecuada, es clave el diseño y construcción de indicadores que permitan, en su momento justo, el monitoreo y las evaluaciones de procesos, resultados e impacto.

Las formas de medición y evaluación que se han discutido en esta guía son relevantes e igualmente aplicables a los tres tipos de modelos de negocios de alto impacto social descritos en el Recuadro 29. La única diferencia en estos casos es que el conjunto de indicadores de resultado abarca dimensiones que van más allá de lo estrictamente social, ya que las transformaciones sociales deseadas están directamente asociadas con los resultados del negocio per se, tales como la productividad y el volumen de ventas (entre otros). Así, el mecanismo para lograr la transformación social es el resultado del negocio productivo y, en esta medida, la información para evaluar el impacto en esas dimensiones puede extraerse de los registros tradicionales de las empresas.

<sup>11</sup> Ver Yusac, M (2009).



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

# Referencias **BIBLIOGRÁFICAS**

9



## REFERENCIAS BIBLIOGRÁFICAS

- Bernal, R. y Peña, A. (2011). Guía práctica para la evaluación de impacto. Bogotá D.C.: Ediciones Uniandes.
- Cedoces. (2016, noviembre) ¿Qué es un negocio inclusivo? Recuperado de <http://negociosinclusivoscolombia.org/index.php/negocios-inclusivos/que-es-un>
- Duflo, E., Glennerster, R., y Kremer, M. (2008). Using Randomization in Development Economics Research: A Toolkit. NBER Technical Working Paper 0339.
- Bertler, P., Martínez, S., Premand, P., Rawlings, B., y Vermeersch, C. (2011). Impact evaluation in practice. Washington, DC: World Bank.
- Glennerster, R., y Takavarasha, K. (2013). Running Randomized Evaluations. Princeton University Press.
- Khandker, S., Koolwal, G., y Samad, H. (2010). Handbook of impact evaluation: Quantitative methods and practices. Washington, DC: World Bank.
- Kusek, J.Z. y Rist, R.C. (2004). Ten Steps to a Results-Based Monitoring and Evaluation System: A Handbook for Development Practitioners. World Bank Publications, Washington, USA.
- Mullainathan, S., & Shafir, E. (2013). Scarcity: Why having too little means so much. New York, NY: Times Books.
- Porter, M. y Kramer, M. (2011). La creación de valor compartido. Harvard Business School Publishing Corporation.
- Ravallion, M. (2008) Evaluation in the practice of development. Policy Research Working Paper 4547, The World Bank, Washington D.C.
- Vermeersch, Christel y Kremer, Michael. 2004. "School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation" Documento de trabajo del Banco Mundial sobre investigaciones relativas a políticas de desarrollo 3523. Banco Mundial, Washington, DC.
- White, H. (2006). DFID Economists' Manual: Impact Evaluation. Independent Evaluation Group, World Bank, Washington D.C.
- White, H. (2009). Theory-Based Impact Evaluation: Principles and Practice. Journal of Development Effectiveness, Vol. 1, No. 3, 271-284.
- World Bank (2006). Conducting quality impact evaluations under budget, time and data constraints. Independent Evaluation Group, The World Bank, Washington D.C.
- Yunus, M. (2007). Creating a world without poverty. Public Affairs, Nueva York.



ASOCIACIÓN NACIONAL DE EMPRESARIOS DE COLOMBIA • ANDI

# Glosario

101

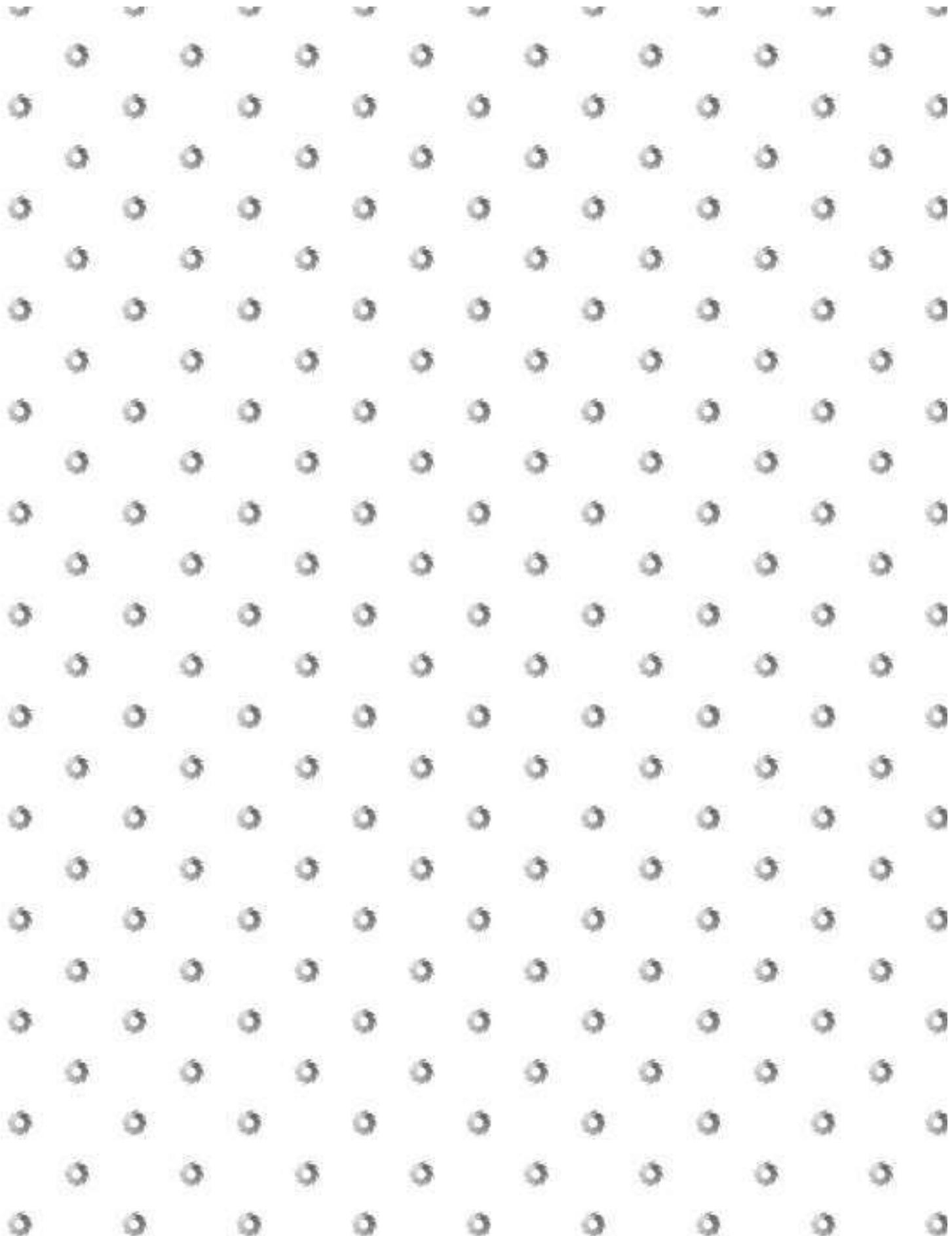


## EVALUACIÓN DE PROYECTOS SOCIALES

2013

TERMINO	DEFINICIÓN
POBLACIÓN BENEFICIARIA	Es la población que recibe los bienes o servicios que produce el proyecto de inversión social.
GRUPO DE TRATAMIENTO	Es el grupo de beneficiarios del programa, es decir, la población beneficiaria.
GRUPO DE CONTROL	Es la población que sirve de punto de referencia para comparar los resultados del grupo de tratamiento ya que es idéntico al grupo de tratamiento justo antes de que éste recibiera los beneficios del programa. En otras palabras, lo único que la diferencia del grupo de tratamiento es que sus miembros no han sido seleccionados para ser beneficiarios del proyecto de inversión social.
CONTRA-FACTUAL	Escenario hipotético en el cual se puede observar qué hubiera sucedido con la población beneficiaria, si no hubieran recibido los beneficios del programa.
OPERADOR	Organización, institución o entidad encargada de administrar e implementar el proyecto de inversión social.
LÍNEA DE BASE	Medición inicial que permite caracterizar a la población de potenciales participantes en el programa, antes de poner en marcha la inversión.
EVALUACIÓN DE PROCESOS	Evaluación en la que se miden y analizan las acciones adelantadas para establecer si se han alcanzado las metas de implementación y si las estrategias se han implementado de acuerdo con lo planeado.
EVALUACIÓN DE RESULTADOS	Evaluación en la que se miden y analizan los cambios en el grupo de tratamiento al comparar los indicadores de resultados antes y después de la aplicación del programa.
EVALUACIÓN DE IMPACTO	Evaluación en la que se mide y analiza la diferencia en el cambio en los indicadores de resultados entre el grupo de tratamiento y el grupo de control, para: (1) identificar la proporción de este cambio que es realmente atribuible al programa y (2) establecer causalidad entre el programa y el resultado observado.
CORRELACIÓN	Dos variables están correlacionadas cuando se mueven de manera simultánea en la misma dirección o en direcciones opuestas.
CAUSALIDAD	Conexión lógica entre una causa y un efecto. Esta conexión o relación puede tener un fundamento teórico o empírico. En el marco de una evaluación, usualmente se plantea la exposición a los beneficios del programa como la causa, y las transformaciones en los beneficiarios son el efecto.
VARIABLES NO OBSERVADAS	Características de la población que no están medidas y por donde no se encuentran en la base de datos disponible para la evaluación.
VARIABLES CONTAMINANTES	Diferencias en características que no cambian en el tiempo no observables u observables entre los beneficiarios y no beneficiarios.
SESCO POR VARIABLES NO OBSERVADAS	Error en la estimación del impacto resultante de la confusión generada por la existencia de variables no observadas que hacen que puedan explicar parte del impacto observado.
SESCO DE SELECCIÓN	Error en la estimación del impacto resultante de la confusión generada por la existencia de diferencias pre-existentes entre el grupo de beneficiarios y el grupo de control, que pueden explicar parte del impacto observado.
VARIABLE INSTRUMENTAL	Variable que no determina los resultados que se quieren afectar con el programa, pero si la probabilidad de ser beneficiario.
SOporte COMÚN	Cuando en el universo de donde se pretende escoger el contra-factual, existen individuos que tienen una "probabilidad de ser seleccionados al programa" similar a la de los beneficiarios.
UMbral DE ELEGIBILIDAD	Punto de corte que establece los individuos que, según un puntaje e indicador, son elegibles para participar en el programa.
REGLA DE FOCALIZACIÓN	Conjunto de pasos y criterios para decidir la población beneficiaria.





## ¿QUIÉNES HICIERON POSIBLE ESTA GUÍA?



La Fundación ANDI lidera acciones y proyectos encaminados a la construcción de una sociedad más equitativa. Por medio de herramientas teóricas y prácticas, acompaña a las empresas afiliadas a la ANDI a que realicen inversiones sociales cada vez más pertinentes, sostenibles y de impacto. Además, trabaja para promover empresas con propósito, en las que los temas sociales sean estratégicos para el negocio y una de sus premisas es ser un ejemplo mundial de un gremio comprometido con el desarrollo social.



CAF es un banco de desarrollo constituido en 1970 y conformado por 19 países, 17 de América Latina y el Caribe, España y Portugal, así como 13 bancos privados de la región. Promueve un modelo de desarrollo sostenible mediante operaciones de crédito, recursos no reembolsables y apoyo en la estructuración técnica y financiera de proyectos de los sectores público y privado de América Latina.



ECONESTUDIO ([www.econestudio.com](http://www.econestudio.com)) es un estudio pequeño pero sofisticado que se especializa en análisis económico y comunicación estratégica. Cuenta con un grupo de economistas muy calificados y profesionales expertos en temas de comunicación estratégica y sostenibilidad, asociados para ofrecer lo mejor de dos mundos: análisis económico de frontera con herramientas de comunicación que permitan su aplicación al logro de objetivos en contextos difíciles, y herramientas de diseño, monitoreo y evaluación de programas aplicadas a la optimización de la inversión social de las empresas. Activa desde 2009. Presta servicios en Colombia y en otros países.

---

SERVICIOS DE APÓYO DEL BANCO DE DESARROLLO PARA AMÉRICA  
LATINA, LA FUNDACIÓN ANDI Y ECONESTUDIO  
2017

---

GUÍA PARA LA  
EVALUACIÓN DE  
PROYECTOS SOCIALES

2017