

EVALUACIÓN Y POLÍTICAS EDUCATIVAS: ¿DOS CAMINOS PARALELOS?

El desarrollo e incremento de las actividades de evaluación de aprendizajes a gran escala han sido características de las últimas dos décadas en casi toda la región latinoamericana. Todos los países han desarrollado algún tipo de sistema de evaluación externa a gran escala.

Algunos lo han hecho en forma continua desde 1995 o antes. Ejemplos de ello son Brasil, Chile, Colombia, Cuba, El Salvador, México, Perú, República Dominicana y Uruguay. Otros han comenzado en forma más tardía o han tenido momentos de discontinuidad (entre ellos, Argentina, Ecuador, Guatemala y Panamá). Solamente Bolivia y Venezuela han detenido el funcionamiento de sus sistemas de evaluación de aprendizajes, bajo el argumento de que desarrollarían un enfoque diferente a los vigentes, enfocándose en la formación integral. Al cabo de casi una década, aún no lo han logrado. También ha sido notorio el incremento en la participación en evaluaciones internacionales como PISA, TIMSS, SERCE y el Estudio Internacional de Educación Cívica.

Simultáneamente, la última década del siglo XX y la primera del siglo XXI han sido prolíficas en materia de iniciativas y “políticas educativas”, definiendo a estas últimas en un sentido restringido: acciones, proyectos y programas implementados desde las instancias de gobierno educativo nacional o local, con el fin de mejorar la educación.

La variedad de ejemplos es enorme: programas de distribución de libros de texto o de bibliotecas de aula; creación de aulas de informática y, más recientemente, distribución de computadores personales para los estudiantes; programas de formación en servicio para docentes; implementación de sistemas de evaluación de desempeño docente, de premios a docentes destacados y de pruebas de conocimiento para el ingreso al ejercicio de la profesión; diversas modalidades de entrega de dinero a familias pobres como contrapartida por la asistencia de sus hijos a la enseñanza primaria o media; extensión de la jornada escolar; reformas de los currículos escolares y planes de estudio; variedad de iniciativas de descentralización de la gestión escolar y de promoción de la participación de las comunidades en las escuelas; reformas en las normas y procedimientos para la evaluación de los alumnos; introducción de nuevos enfoques y metodologías para enseñar disciplinas específicas; programas para evitar la deserción; cursos “acelerados” para estudiantes rezagados; entre otros, la lista podría continuar.

En este contexto, una de las debilidades principales del debate educativo durante este tiempo, ha sido la ausencia de evaluaciones sistemáticas de la mayor parte de estas iniciativas o “políticas educativas”. Muy pocos programas fueron acompañados desde el inicio con diseños apropiados de evaluación de sus impactos.

En algunos casos, se intentó suplir esta carencia recurriendo a los datos de las evaluaciones a gran escala mencionadas en el primer párrafo. Para un observador ajeno al campo de la evaluación (político, ciudadano, docente), debería ser posible identificar si los alumnos aprenden o no más y mejor, habiendo tantos resultados de evaluaciones. Sin embargo, esta asunción es errónea, por varios motivos, entre los cuales cabe destacar cuatro:

- A) La falta de continuidad en la actividad y en los enfoques de los sistemas de evaluación a nivel nacional (si los sistemas de evaluación se interrumpen o modifican su enfoque, no se contará con suficiente información).
- B) Dificultades de orden técnico para asegurar la comparabilidad de los resultados a lo largo del tiempo (técnicamente se denomina equiparación de las puntuaciones).
- C) El foco en la medición de logros al final de un determinado grado a nivel educativo (una verdadera medición de aprendizajes requiere contar con información para al menos dos momentos en el tiempo para los mismos alumnos, por ejemplo, al inicio y al final del año lectivo o de la participación en un determinado programa).
- D) Para realizar una asignación válida de causalidad a un programa (es decir, para poder afirmar que determinados avances o retrocesos en los aprendizajes son el efecto de un determinado programa) se requiere de un diseño metodológico especialmente concebido para ello, dado que los factores que pueden estar incidiendo en los resultados son múltiples.

En relación con los dos primeros problemas se han registrado avances significativos en varios países de la región, pero es poco lo que se ha avanzado en los dos últimos, que son los más importantes para valorar el impacto de programas y “políticas” específicos. Esto último requiere pensar el diseño de la evaluación junto con el propio programa, de modo que la misma se realice desde el comienzo de su implementación.

"El Programa de Promoción de la Reforma Educativa en América Latina y el Caribe - PREAL - es un proyecto conjunto del Diálogo Interamericano, con sede en Washington, y la Corporación de Investigaciones para el Desarrollo, con sede en Santiago de Chile. Su objetivo central es contribuir a mejorar la calidad y equidad de la educación en la región mediante la promoción de debates informados sobre temas de política educacional y reforma educativa, la identificación y difusión de buenas prácticas y la evaluación y monitoreo del progreso educativo.

Sus actividades se realizan a través de Centros Asociados de Investigación y Políticas Públicas en diversos países de la región y comprenden la realización de estudios, la organización de debates y la promoción de diálogos públicos sobre temas de política educacional y reforma educativa.

El Grupo de Trabajo sobre Estándares y Evaluación del PREAL fue creado con el propósito de informar, legitimar y utilizar espacios de la labor de medición de resultados del aprendizaje y promover un debate informado sobre el tema de la formulación de estándares. Se coordina desde el Grupo de Análisis para el Desarrollo -- GRADE -- de Lima, Perú, y cuenta con una amplia red de participantes con los que se vincula a través de un sitio virtual desde el cual también divulgan los resultados de estudios realizados en el marco de sus actividades y promueven foros y debates sobre su contenido. Ver: http://www.preal.org/Grupo2.asp?Id_Grupo=3

Las actividades del PREAL son posibles gracias al apoyo de la United States Agency for International Development (USAID), el Banco Interamericano de Desarrollo (BID), la International Association for the Evaluation of Educational Achievement (IEA), GE Foundation, The Tinker Foundation, entre otros."

Esta situación es la que pretende señalar el título de este boletín: tenemos gran cantidad de evaluaciones, por un lado, y gran cantidad de iniciativas, “políticas” e innovaciones, por otro lado. Pero no podemos cruzarlas de manera válida, con el fin de contar con información acerca de qué resultados están teniendo estas últimas. Son caminos paralelos, que no se encuentran (y, algunas veces, se intenta forzar dicho encuentro de un modo inapropiado).

En ausencia de evaluaciones específicamente diseñadas para responder la pregunta sobre en qué medida una determinada iniciativa o programa tiene un efecto positivo y verificable en los aprendizajes de los alumnos, los debates educativos se tornan puramente ideológicos, en el sentido de que no se sustentan en evidencia empírica sino en las posturas previas de los actores. Esto no es malo en sí mismo, pero una base de evidencia empírica sólida podría contribuir a que los debates fuesen más productivos. En ausencia de dicha evidencia, buena parte de la discusión sobre el éxito o fracaso de las mal denominadas “reformas” educativas suele tener poco sentido. Del mismo modo, es imposible discutir seriamente sobre la conveniencia de determinadas inversiones o sobre si el mismo dinero podría ser utilizado de manera más racional e inteligente.

Una nueva tendencia en el horizonte de la región: la evaluación específica de la implementación e impacto de los programas e inversiones en educación

En la región existen estudios y esfuerzos aislados de evaluación de impacto de programas educativos o, dicho de otra manera, de innovaciones educativas que cuentan desde el comienzo con un diseño de evaluación.

Recientemente, el Banco Interamericano de Desarrollo ha asumido esta línea de trabajo como prioritaria. En la sección de evaluación de la educación de su sitio WEB (<http://www.iadb.org/es/temas/educacion/evaluacion,6978.html>) se enfatiza que la labor del Banco está orientada a “*la evaluación de las políticas educativas, programas y reformas...*” con el objetivo de “*...utilizar la evidencia científica disponible para identificar programas eficaces de educación en América Latina y el Caribe... En la medida en que cada año se gastan miles de millones de dólares en programas de educación... se hace necesaria la continuidad de rigurosos programas de investigación y evaluación, con el propósito de comprender el impacto real de las diversas intervenciones... para garantizar efectos positivos en el aprendizaje de los estudiantes y asignar recursos de manera eficiente*”.

Enseña Perú: empezamos a registrar el cambio

El BID lanza un nuevo video sobre el programa "Enseña Perú", que recluta y capacita a graduados universitarios sobresalientes para que enseñen en escuelas de bajos recursos, por un período de dos años. Las primeras evaluaciones del programa realizadas por el BID, ya registran un impacto en el aprendizaje y las aspiraciones de los niños.



[Ver video](#)



APORTES
N° 1 - NOVIEMBRE 2009

Emma Näslund-Hadley
Pablo Ibarrarán
Marcelo Cabrol

● **MÁS ALLÁ DE LA TIZA Y LAS
● PALABRAS: EDUCACIÓN
● EXPERIMENTAL DE MATEMÁTICA
● Y CIENCIAS EN ARGENTINA**

En la sección de publicaciones de dicho sitio web se puede acceder a varios estudios de evaluación de impacto realizados recientemente, entre otros “Enseña Chile”, “Enseña Perú”¹, PROGRESA en México, “One laptop per child” en Perú y el “Programa de Mejora de la Educación Científica y Matemática” en Argentina.

En este último caso se trata de un estudio de investigación evaluativa que acompañó desde el comienzo la implementación del Programa².

Algunas de las características especialmente interesantes de este estudio, son las siguientes:

- I) Incluyó una medición inicial de conocimientos y habilidades al inicio del año lectivo, con anterioridad al inicio del Programa, y una segunda medición, con pruebas equivalentes, al final del año. Esto permitió constatar el progreso de los alumnos a lo largo del año.
- II) Se conformaron cuatro grupos de escuelas, pertenientes a poblaciones de alta vulnerabilidad social en dos provincias.

¹ En el caso de Perú lo que se ha evaluado es si la forma de selección efectivamente recluta a “los mejores” estudiantes (y concluye que sí) pero no se evalúa la efectividad de la enseñanza que pueden dar ellos en el aula.

² El estudio fue realizado en forma conjunta por la Oficina de Buenos Aires de IPE/UNESCO y el Instituto de Evaluación Educativa de la Universidad Católica del Uruguay. El Informe final pertenece al Ministerio de Educación de Argentina y al BID, y no ha sido publicado hasta el momento.



En tres de estos grupos se implementaron distintos programas de mejora de la enseñanza científica y matemática, que incluyeron materiales didácticos de apoyo, formación en servicio y acompañamiento didáctico a los maestros. En el cuarto grupo no se realizó intervención alguna. La asignación de las escuelas a cada grupo se realizó en forma aleatoria, lo que permitió asegurar la comparabilidad de las poblaciones. Simultáneamente, dado que la asignación de alumnos y maestros nunca puede ser aleatoria, se realizaron controles estadísticos adicionales para garantizar dicha comparabilidad.

III) El estudio no se limitó a indagar en los resultados, sino que incluyó un amplio operativo de seguimiento de la implementación en las escuelas de cada uno de los aspectos centrales de cada programa. Este aspecto es de fundamental importancia, por las razones que explicaremos enseguida.

Más información sobre las características y resultados de este Estudio, puede encontrarse en el Boletín APORTES, disponible en:

<http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=35435866>.

En la misma línea, el BID acaba de publicar el Informe PANORAMA DE LA EFECTIVIDAD EN EL DESARROLLO 2010, disponible en: <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=35804529>.



El documento está enfocado en la evaluación de la efectividad en distintas áreas, entre ellas la educación (páginas 75 a 83), en las que se incluye información sobre la evaluación mencionada en Argentina y la evaluación del programa “Enseña Chile”.

Enfatiza la política del Banco de exigir la inclusión de diseños de evaluación de impacto en los nuevos programas que ellos apoyan, y muestra cómo ha ido evolucionando el porcentaje de proyectos con diseños de evaluación en los últimos años: 8% en 2008, 11% en 2009, 27% en 2010.

Incluye, en su página 45, una definición de evaluación de impacto, que resumimos a continuación.

Recuadro 1 ¿Qué es una Evaluación de Impacto? *

Para conocer sobre el impacto en el desarrollo de nuestro trabajo, la pregunta metodológica clave a responder es: ¿qué hubiese ocurrido en ausencia del proyecto? A este escenario sin proyecto se le refiere como *counterfactual* (**este término suele ser traducido en la literatura técnica como “contrafáctico”**) y la tarea más desafiante en cualquier evaluación de impacto es encontrar uno que sea adecuado. La evaluación ideal compararía la situación de los beneficiarios de un proyecto con la de esos mismos beneficiarios sin la presencia del proyecto en un mismo punto del tiempo. No obstante, es imposible observar al mismo grupo al mismo tiempo, con y sin el programa. Debe encontrarse un sustituto razonable.

El “*counterfactual*” puede ser elaborado de diferentes maneras. Una alternativa es usar la situación antes del programa. Sin embargo, esta comparación antes-después no considera otros factores externos al programa que podrían haber cambiado en el tiempo y afectado el resultado que interesa. Otra alternativa es usar después del programa a no beneficiarios como un *counterfactual*. Esto es problemático, pues sabemos que si los no beneficiarios eran diferentes a los beneficiarios antes del programa, las diferencias entre aquellos y estos pueden deberse a diferencias inherentes a sus características más que al programa. Finalmente, el *counterfactual* puede ser construido asignando o no el proyecto aleatoriamente a un grupo de potenciales beneficiarios interesados y elegibles. Este es el *counterfactual* ideal puesto que garantiza que ambos grupos sean en promedio los mismos antes del programa, y que cualquier diferencia después de éste tendrá como causa el programa mismo. Pese a que no siempre es posible asignar aleatoriamente individuos a un programa, deben realizarse todos los esfuerzos para acercarse lo más posible a ello.

Una vez definido un “*counterfactual*”, el método más utilizado para registrar el impacto de un proyecto es la metodología de la diferencia-en-las-diferencias (DID). Este método compara la diferencia entre beneficiarios y no beneficiarios (primera diferencia), antes y después del programa (segunda diferencia); es decir, compara las tendencias en el resultado entre los beneficiarios y los no beneficiarios. Esta técnica puede identificar y estimar el impacto exacto del programa, en la medida en que, en ausencia del programa, beneficiarios y no beneficiarios hubiesen tenido la misma tendencia en las variables del resultado. En todo caso, suele ocurrir que las tendencias en las variables del resultado difieren significativamente entre beneficiarios y no beneficiarios. Para abordar este problema, se presta especial atención a identificar un subconjunto de no beneficiarios con características similares a los beneficiarios, mediante una técnica denominada “propensity-score-matching”.

.....

* Un Glosario de Términos de Evaluación de Impacto, basado en Gertler et al. (2010), se incluye en el DEO 2010, a partir de la página 261. Para más material sobre este tópico, ver también la sección “Publicaciones” en:

<http://www.iadb.org/en/topics/development-effectiveness/>.



La importancia de evaluar la implementación junto con el impacto

Más allá de la definición anterior, debe señalarse que la evaluación de impacto es insuficiente por sí misma. Una de las características particularmente relevantes en el estudio realizado en Argentina y mencionado más arriba es que incluyó la investigación en torno a la implementación de cada uno de los programas. Uno de los puntos débiles de la evaluación de impacto concebida exclusivamente como medición de los resultados es que presupone que los programas se implementan tal cual fueron concebidos. Pero, en la práctica, esto nunca ocurre así. Siempre y en forma inevitable, existe una distancia entre la concepción teórico-conceptual de un proyecto y su implementación en la realidad, dado que esta última depende del modo en que los distintos actores se apropiaron del mismo.

En el caso concreto del programa de educación científica y matemática, su implementación fue variable según el clima institucional en cada escuela, la formación previa de los maestros, la calidad del trabajo de los acompañantes, entre otros. El seguimiento de la implementación permitió dos cosas muy importantes. Por un lado, identificar las dificultades en la implementación y la diversidad de grados de implementación en las distintas escuelas. Esta información es clave a la hora de mejorar un programa o de avanzar en su implementación en una escala mayor.

Por otro lado, el estudio permitió mostrar que, a mayor grado de implementación, mayor impacto en los aprendizajes. Esta información también es clave. Primero, porque constituye evidencia adicional de que el impacto en los aprendizajes es resultado de la implementación del programa. Segundo, porque evita conclusiones equivocadas, basadas en resultados promedio. Muchas veces se subvalora el impacto de ciertos programas, porque se incluye en la medición escuelas en las que en realidad no hubo implementación. En otras palabras, el impacto siempre será reducido o nulo, si el programa no es adecuadamente implementado. Si no se evalúa la implementación, se corre el riesgo de considerar como “no-efectivo” a un programa que en realidad fue “no-implementado”.

Si un programa es implementado de acuerdo a lo previsto y no tiene impacto, no debería replicarse ni ampliarse. En cambio, si un programa tiene niveles de impacto diferentes según su grado de implementación, será clave analizar los problemas de implementación y las variables que inciden en la misma.

En resumen, el desarrollo de diseños específicos para la evaluación y seguimiento de la implementación y del impacto en los aprendizajes de las diversas iniciativas, innovaciones y “políticas” educativas, es un campo en estadio inicial de desarrollo en la región, pero clave para informar el debate público y la toma de decisiones en educación.



Fuentes adicionales de información sobre el tema

En los Estados Unidos hace ya más de una década que existe un debate importante sobre qué tipos de investigaciones son apropiadas y suficientes para formular juicios de valor válidos acerca de la efectividad de los diversos programas e intervenciones educativas.

Dos iniciativas interesantes en este sentido se presentan brevemente a continuación. En ambos casos se trata de materiales que solamente están disponibles en Internet en idioma inglés.

1. What Works Clearinghouse (<http://ies.ed.gov/ncee/wwc/>), que es algo así como un centro de información sobre “lo que funciona” en educación. Es parte del sitio web del Instituto de Ciencias de la Educación del Departamento de Educación de los Estados Unidos. La función del sitio es seleccionar estudios técnicamente apropiados de evaluación de impacto de distintas intervenciones y programas educativos, para darles difusión. El énfasis está puesto en la calidad de los diseños de investigación y evaluación que están en la base de cada estudio.

The screenshot shows the homepage of the What Works Clearinghouse. At the top, it features the 'ies WHAT WORKS CLEARINGHOUSE' logo and the affiliation 'U.S. Department of Education Institute of Education Sciences'. Navigation links include 'NewsFlash', 'Contact', 'Site Index', 'Help', and 'Tour the WWC'. A search bar with a 'GO' button is present. Below the navigation is a horizontal menu with categories: 'Topic Areas', 'Publications & Products', 'Reference Resources', 'WWC Help', 'What's New', and 'About Us'. The main content area includes a 'Welcome to WWC' message, a 'Find What Works' search section with a dropdown menu set to 'Academic Achievement' and a 'Search the Research' button, and a 'What's New' section. A large banner on the right side of the page says 'Find What Works' and encourages searching for programs or products that address school or district needs. Below this banner, there are links to 'WWC Releases New Study Design Standards', including 'Single Case Design Technical Documentation NEW (446 KB)' and 'Standards for Regression Discontinuity Designs NEW (174 KB)'.

2. The Campbell Collaboration (<http://www.campbellcollaboration.org/>) es una organización orientada a la recolección de estudios de tipo experimental o cuasi experimental (es decir, con asignación aleatoria de casos) en diversas áreas de actividad (entre ellas la educación). La sección Biblioteca de esta página incluye 81 estudios sobre educación.

The banner for The Campbell Collaboration features a globe icon with 'C2' on it. The text reads: 'THE CAMPBELL COLLABORATION' in large yellow letters, followed by the tagline 'What helps? What harms? Based on what evidence?' in white text on a blue background.