



**REFLEXIONES
SOBRE EL SISTEMA
EDUCATIVO ESPAÑOL**

**EVALUAR PARA MEJORAR:
HACIA EL SEGUIMIENTO
Y LA EVALUACIÓN
SISTEMÁTICA DE LAS
POLÍTICAS EDUCATIVAS**

DANIEL SANTÍN
GABRIELA SICILIA
Universidad Complutense de Madrid

MONOGRAFÍAS SOBRE EDUCACIÓN

FUNDACIÓN
RAMÓN ARECES

Fundación Europea
Sociedad y Educación

REFLEXIONES

SOBRE EL SISTEMA EDUCATIVO ESPAÑOL

EVALUAR PARA MEJORAR: HACIA EL SEGUIMIENTO Y LA EVALUACIÓN SYSTEMÁTICA DE LAS POLÍTICAS EDUCATIVAS

DANIEL SANTÍN

GABRIELA SICILIA

Universidad Complutense de Madrid

FUNDACIÓN
RAMÓN ARECES

Fundación Europea
Sociedad y Educación

**EVALUAR PARA MEJORAR: HACIA EL SEGUIMIENTO
Y LA EVALUACIÓN SISTEMÁTICA DE LAS POLÍTICAS EDUCATIVAS**
REFLEXIONES SOBRE EL SISTEMA EDUCATIVO ESPAÑOL

Autores:
Daniel Santín
Gabriela Sicilia

DIRECCIÓN DEL PROYECTO
Mercedes de Esteban Villar
Fundación Europea Sociedad y Educación

El contenido y opiniones expuestos en este documento son responsabilidad exclusiva de sus autores.

Todos los derechos reservados

Este documento no podrá ser reproducido total o parcialmente en cualquier soporte impreso o digital sin la autorización de la Fundación Ramón Areces y la Fundación Europea Sociedad y Educación.

EDICIÓN 2014

© Fundación Ramón Areces
Vitruvio, 5 - 28006 Madrid
www.fundacionareces.es

© Fundación Europea Sociedad y Educación
José Abascal 57, 5º B
28003 Madrid
T 34 91 455 15 76
www.sociedadyeducion.org

© Autores

Diseño:
KEN / www.ken.es

ISBN 978-84-697-0518-6

Impreso en España

ÍNDICE

PRESENTACIÓN	7
RESUMEN EJECUTIVO	8
DOCUMENTO DE TRABAJO	13
SUMARIO	14
INTRODUCCIÓN	15
EL DISEÑO DE UNA EVALUACIÓN EDUCATIVA: CADENA DE RESULTADOS, INDICADORES Y LÍNEA DE BASE	19
Cadena de resultados	20
Indicadores	22
Línea base o estudio de diagnóstico	23
ÁREAS DE LA EVALUACIÓN EDUCATIVA: EFICACIA, EFICIENCIA E IMPACTO	23
Eficacia	23
Eficiencia	24
Evaluación de impacto	29
UN EJEMPLO: PLAN DE SEGUIMIENTO Y EVALUACIÓN (S&E) DEL PROGRAMA ESCUELA 2.0	34
Breve descripción del programa Escuela 2.0.	34
Elementos de un plan de S&E para el programa Escuela 2.0.	35
Definición de los indicadores de la cadena de resultados del programa	37
Evaluaciones del programa Escuela 2.0.	40
REFLEXIONES FINALES	42
REFERENCIAS	43

PRESENTACIÓN

Reflexiones sobre el sistema educativo español es el título de un proyecto impulsado por la Fundación Ramón Areces y la Fundación Europea Sociedad y Educación a lo largo de más de dos años. Su objetivo es profundizar en las relaciones entre rendimiento educativo y crecimiento económico, desde dos perspectivas fundamentales: por una parte, la que vincula el desarrollo y la competitividad de una sociedad con medidas que inciden directamente en la mejora y competencia del capital humano a través de la acción educativa; por otra, la que analiza el diseño de políticas educativas de éxito y sus efectos, a partir de evidencias procedentes de estudios comparados.

En su trabajo “Evaluar para mejorar: hacia el seguimiento y la evaluación sistemática de políticas educativas”, los profesores Daniel Santín y Gabriela Sicilia analizan la evaluación educativa y realizan una serie de propuestas para su articulación de manera sistemática e institucional. Este documento forma parte de una obra colectiva publicada en 2015 por las fundaciones Areces y Europea Sociedad y Educación, en la que un grupo de expertos, especializados en economía de la educación, aborda aspectos clave de las políticas educativas y sus efectos en el logro escolar.

Este proyecto, dirigido por la Fundación Europea Sociedad y Educación, y que ha contado con el apoyo de la Fundación Ramón Areces, pretende ofrecer algunas bases para una fundamentada reflexión sobre nuestro sistema educativo, apoyada en los resultados de la investigación y en la evidencia empírica, y dirigida tanto al lector no especialista como a los responsables de la toma de decisiones y a los profesionales de la educación.

Reconociendo y poniendo en valor los logros alcanzados, cada uno de los trabajos refleja un compromiso por la mejora de la educación en España y procura, desde su propia perspectiva de análisis, responder a una serie de cuestiones básicas que marquen el rumbo de reformas duraderas y estables en nuestro sistema educativo: ¿qué tipo de educación y qué tipo de competencias predicen un crecimiento sostenible y una más equitativa distribución de oportunidades para el capital humano de nuestro país?, ¿qué tipo de reformas educativas resultan exitosas y responden mejor a los desafíos que plantea hoy la educación?, ¿qué factores de contexto deben ser tenidos en cuenta para implantar políticas educativas que redunden en el desarrollo y competitividad de la sociedad española?

FUNDACIÓN RAMÓN ARECES

FUNDACIÓN EUROPEA SOCIEDAD Y EDUCACIÓN

RESUMEN EJECUTIVO

Este resumen destaca las ideas principales que desarrollan los economistas Daniel Santín y Gabriela Sicilia en el trabajo *Evaluar para mejorar: hacia el seguimiento y la evaluación sistemática de las políticas educativas*, que se integra en el proyecto *Reflexiones sobre el sistema educativo español*, promovido por la Fundación Europea Sociedad y Educación. Se ha publicado en 2015 como parte de una obra colectiva y cuenta con el patrocinio de la Fundación Ramón Areces.

EN CLAVE DE MEJORA

El futuro de cualquier sistema educativo debería estar guiado por una continua aspiración de mejorar sus resultados y una mayor transparencia, con el objetivo de ofrecer más información sobre las variables educativas relevantes para que los agentes implicados en las políticas educativas puedan tomar mejores decisiones. El seguimiento y la evaluación de las políticas educativas, realizados de forma sistemática en el tiempo, debe ser la herramienta fundamental que permita conseguir ambos objetivos. Los autores describen cuáles son las claves para poner en marcha un sistema integral de evaluación de la calidad, la *eficacia*, la *eficiencia* y la *equidad* del sistema educativo no universitario español.

MARCANDO UN RUMBO: TRES COORDENADAS

Tradicionalmente, la evaluación del sistema educativo ha sido parcial y ha estado centrada en el cálculo de los recursos invertidos y el número de alumnos beneficiarios. Si bien el control presupuestario resulta útil como mecanismo de fiscalización para garantizar que el presupuesto es ejecutado respetando la norma, resulta insuficiente como herramienta de evaluación. En particular, debe volver a subrayarse que de una correcta ejecución presupuestaria no se deriva un impacto en los objetivos a corto, medio y largo plazo perseguidos o que su aplicación haya sido eficiente. Por ello, junto al seguimiento de los costes del sistema educativo y del número de alumnos que pasan por las aulas existe, desde hace al menos un par de décadas, un creciente interés por medir los resultados educativos obtenidos.

El seguimiento y la evaluación de las políticas educativas debe ser el proceso generador de información relevante que permita relacionar los recursos, actividades y resultados a corto, medio y largo plazo de los centros y programas educativos de una administración. El seguimiento y la evaluación deben realizarse observando la evolución del sistema educativo en el tiempo para utilizar los resultados en la toma de decisiones racionales que permitan mejorar la eficacia, la eficiencia y la equidad del sistema educativo.

El sistema educativo español debe incorporar las principales técnicas de seguimiento y evaluación orientadas a resultados como pilar fundamental donde asentar el salto cualitativo que demanda la economía y la sociedad. Además, toda la información generada debe ser transparente para la sociedad civil y para los agentes que forman el sistema educativo de manera que estos puedan tomar las decisiones adecuadas.

QUÉ PROPONEMOS

La evaluación educativa debe estructurarse en una cadena de resultados para analizar tres grandes áreas: eficacia, eficiencia e impacto. La eficacia mide el grado de cumplimiento de todos los indicadores implicados en el proceso educativo pero, fundamentalmente, de los referidos a calidad, resultados, número de beneficiarios y equidad, definidos por la política educativa. La eficiencia permite saber si los recursos destinados para cumplir con dichos objetivos han sido, o no, los óptimos, y por qué y cómo evoluciona esta medida en el tiempo. Finalmente, la *evaluación de impacto* permite determinar si existe causalidad para concluir si determinadas políticas y programas educativos han permitido o no mejorar los indicadores de resultados educativos de la población destinataria. En definitiva, el seguimiento y la evaluación permiten informar a los gestores del sistema educativo si éste está funcionando adecuadamente o no. Con esta valiosa información los responsables técnicos y políticos podrán proponer cambios, plantear nuevas acciones, suprimir otras, introducir innovaciones y, en definitiva, contribuir a maximizar la eficacia, la eficiencia y el impacto de los recursos destinados a educación. Sin una evaluación propia no será posible tomar decisiones racionales basadas en evidencias; en el mejor de los casos, únicamente se aplicarán recetas que resultaron exitosas en otros países.

El primer paso que es necesario dar para poner en marcha una evaluación educativa es definir los indicadores que forman parte de la cadena de resultados que describan el marco lógico de cómo las políticas y los programas educativos van a conseguir alcanzar los objetivos planteados, que deben ser claros y cuantificables. La Figura 1 muestra de manera esquemática la articulación de esta cadena de resultados.

FIGURA 1. LA CADENA DE RESULTADOS EN UNA POLÍTICA O PROGRAMA EDUCATIVO



Fuente. Elaboración propia a partir de Gertler et al. (2011)

En la cadena de resultados educativa se pueden distinguir dos fases que se relacionan directamente con las tareas posteriores de seguimiento y evaluación. En primer lugar, los centros de educación no universitaria –colegios e institutos– llevan a cabo cada curso y en cada nivel educativo un proceso productivo en sus alumnos del tipo: *inputs* – actividades – *outputs*. A partir de los *inputs* con los que cuentan –recursos escolares, nivel socioeconómico y estructura familiar de los alumnos, conocimientos previos–, los colegios implementan una serie de actividades –metodología docente, implicación de los padres, tareas en el hogar, pruebas de evaluación, etc.– para transformar e incrementar los conocimientos de los alumnos a lo largo del curso. Debe tenerse en cuenta que el *output* final que se entrega a los alumnos es tan solo el resultado final del curso, el cual es controlable por la unidad productora, pero no el resultado final de la educación que se obtendrá una vez que cada cohorte finalice el proceso educativo en cada etapa. Cabe también destacar que esta tarea es llevada a cabo de forma repetida en el tiempo por parte de las unidades productivas; escuelas e incluso distintos profesores dentro de la misma escuela. Este proceso debe ser evaluado a través de técnicas de seguimiento y, en particular, a partir de la medición de la eficacia, la eficiencia y el cambio en productividad del proceso educativo.

En segundo lugar, una vez que los alumnos terminan el ciclo educativo se pueden empezar a medir los resultados a corto, medio y largo plazo de la educación en-

tregada por los centros. Estos resultados son los que produce el sistema educativo en forma de tasas de éxitos y rendimiento medio al final de cada etapa (corto y medio plazo) o acceso a la educación superior y al mercado de trabajo (largo plazo). Las técnicas de evaluación de impacto permiten analizar la relación causal entre los recursos del sistema educativo y los distintos programas educativos puestos en marcha por las autoridades educativas o por centros escolares innovadores y los resultados finales que se producen en los alumnos a corto, medio y largo plazo.

RECOMENDAMOS

España tiene un grave problema de abandono y fracaso escolar. Por ello, la evaluación continua es esencial para detectar a los alumnos con alto riesgo de fracaso escolar. Gracias a ella se podrá intervenir tempranamente desde la escuela para tratar de evitar, mediante la acción conjunta de familias y escuelas, que la contingencia efectivamente suceda, lo que fortalecerá a la postre la igualdad de oportunidades educativas.

La evaluación educativa debería ser realizada por una entidad externa lo más independiente posible con la colaboración de las instituciones evaluadas responsables del diseño y ejecución de la intervención educativa. En el supuesto de que la evaluación se lleve a cabo a nivel regional o local, una opción podría ser la creación de Agencias Autonómicas de Evaluación Educativa que dependan exclusivamente de la Presidencia de la Comunidad Autónoma y que por lo tanto tengan plena capacidad e independencia para llevar a cabo las evaluaciones. Otra opción sería la contratación de equipos de consultores externos independientes que realicen la evaluación en colaboración con las Consejerías de Educación.

DOCUMENTO DE TRABAJO

REFLEXIONES SOBRE EL SISTEMA EDUCATIVO ESPAÑOL

**EVALUAR PARA MEJORAR:
HACIA EL SEGUIMIENTO Y LA EVALUACIÓN
SISTEMÁTICA DE LAS POLÍTICAS EDUCATIVAS**

DANIEL SANTÍN
GABRIELA SICILIA

SUMARIO

El futuro de cualquier sistema educativo debería estar guiado tanto por una continua aspiración de mejorar sus resultados como por una mayor transparencia, donde se ofrezca más y mejor información sobre las variables educativas relevantes para que los agentes implicados en las políticas educativas puedan tomar mejores decisiones. El seguimiento y la evaluación de las políticas educativas realizados de forma sistemática debe ser la herramienta fundamental que permita conseguir ambos objetivos. Los autores describen cuáles son las claves para poner en marcha un sistema integral de evaluación sistemática de la calidad, la eficacia, la eficiencia y la equidad del sistema educativo no universitario español. Para ello se discute la información que sería necesario recoger y las principales técnicas de seguimiento y evaluación que deben acompañar el proceso educativo. Además de exponer la necesidad de

evaluar, los autores también dedican un espacio a presentar las ventajas de las dos metodologías que en los últimos años se han impuesto para realizar la evaluación y el seguimiento de los centros y los programas educativos: la evaluación de impacto y la medición de la eficiencia y la productividad total de los factores. Cualquier programa o intervención educativa que quiera ponerse en marcha y sea susceptible de ser evaluada debería ser diseñada de forma que su evaluación pueda ser implementada mediante estas técnicas. En este sentido, a modo de ejemplo se hace una reflexión sobre cómo debería haber sido evaluado el Programa Escuela 2.0. Finalmente, en el trabajo se explorará además cómo debería articularse institucionalmente la evaluación educativa dentro de un contexto donde las competencias educativas están descentralizadas.

INTRODUCCIÓN

Si acudimos al diccionario de la Real Academia Española evaluar es “estimar, apreciar, calcular el valor de algo”. Definido de esta manera, resulta del máximo interés conocer cómo debemos calcular el valor de las políticas públicas que se llevan a cabo en España. Por otro lado, la evaluación o el cálculo del valor de una política o programa público, debe realizarse teniendo en cuenta tanto los recursos invertidos como los resultados alcanzados. Por tanto, la evaluación de un programa educativo está íntimamente ligado con la medición de su eficiencia, que desde un punto de vista Paretiano implica la ausencia de desperdicio de los recursos aplicados para conseguir un objetivo determinado.

La importancia de la eficiencia en la articulación de las políticas públicas resulta un objetivo clave y más aún en el actual contexto de elevado déficit público y persecución de la estabilidad presupuestaria. En este escenario la sociedad simplemente no puede permitirse disponer de recursos ociosos que puedan aumentar su rendimiento en usos alternativos. Tal es la importancia de la eficiencia en la actuación pública que prácticamente todos los países recogen en sus Constituciones el mandato de ejecutar sus presupuestos de forma eficiente. En España, el artículo 31.2 de la Carta Magna afirma que: “El gasto público realizará una asignación equitativa de los recursos públicos y su programación y ejecución responderán a los criterios de eficiencia y economía”. Para cumplir con este mandato legal, la evaluación de las distintas políticas públicas es la herramienta de la que disponemos para medir el grado de cumplimiento de este objetivo de eficiencia dentro de un proceso que debería ser sistemático para racionalizar el gasto público y mostrar a los ciudadanos los resultados de las políticas aplicadas.

La evaluación de las políticas públicas en general, pero de la política educativa en particular, es un proceso no exento de problemas debido a distintos factores. En primer lugar, la educación formal reglada es un proceso productivo complejo que consiste en transmitir unos conocimientos en sentido amplio con el objetivo de que el alumno los aprenda y que los mismos le sean relevantes para adquirir nueva formación hasta completar su desarrollo personal y profesional. En segundo lugar, existen distintos agentes implicados en el proceso educativo como alumnos, padres, profesores, directores de centro o la administración pública, cada uno con sus propios objetivos, intereses e incentivos que a veces no están bien definidos y que no siempre coinciden. En tercer lugar, el producto de la educación carece de precios de mercado por lo que el valor de los resultados generados resultan difícilmente cuantificables. En cuarto lugar, los logros educativos generan valor de forma directa para el individuo que la obtiene en forma de capital humano (Becker, 2009) que se traduce en una mayor productividad que en media generará mejor salario y empleabilidad esperados a lo largo de su ciclo vital. Finalmente, la educación genera además distintas externalidades positivas tanto económicas como sociales. Por un lado, el sistema productivo resulta fortalecido en su conjunto a través de beneficios tales como mayor productividad, ocupación y crecimiento económico. De estos resultados se benefician no sólo los individuos sino también la sociedad, dado que los individuos con más ingresos serán también los futuros pagadores de más impuestos que contribuirán a desarrollar políticas de gasto más generosas que mejoren la cohesión social. Por otro lado, la educación favorece aspectos más sociales como son la participación y consolidación de los valores democráticos o la reducción de la marginalidad, la exclusión y la criminalidad (Lochner y Moretti, 2001) con los consiguientes ahorros en protección social.

Tradicionalmente la evaluación del sistema educativo ha sido parcial y ha estado centrada en garantizar el cumplimiento de los compromisos de gasto y el acceso al sistema. Si bien el control presupuestario sigue resultando útil como mecanismo de control y fiscalización para garantizar que el presupuesto es ejecutado respetando la norma, resulta insuficiente como herramienta de evaluación. En particular debe volver a ser subrayado que de una correcta ejecución presupuestaria no puede ser inmediato concluir que los recursos aplicados tengan algún impacto en los objetivos a corto, medio y largo plazo perseguidos, más allá del número de alumnos que han recibido el servicio, o que hayan sido aplicados de forma eficiente. Es por ello que junto al seguimiento de los costes del sistema educativo existe, desde hace un par de décadas, un creciente interés por medir los resultados educativos obtenidos en términos de las competencias que adquieren los alumnos. Para ello se han puesto en marcha proyectos de evaluación internacional del rendimiento a través

de pruebas de conocimiento a los que cada vez se adhieren más países. España se ha ido incorporando poco a poco a las evaluaciones de estos proyectos. En 1995 participó en la primera oleada del TIMSS (*Third International Mathematics and Science Study*) promovido por la IEA (*The International Association for the Evaluation of Educational Achievement*) y recientemente participó nuevamente en la oleada de TIMSS-PIRLS 2011 (*Progress in International Reading Literacy Study*). Además, España ha participado en las cinco oleadas que hasta la fecha han tenido lugar de la evaluación internacional que dirige la OCDE conocida como Proyecto PISA (*Programme for International Student Assessment*).

Este proceso de conversión de la evaluación orientada a los recursos en una mayor atención hacia el análisis de los resultados se puede observar en la evolución de la legislación educativa en España. Tanto la Ley General de la Educación (LGE) de 1970 como la Ley Orgánica de Ordenación General del Sistema Educativo (LOGSE) de 1990 eran leyes orientadas a garantizar la igualdad de los recursos educativos en todas las escuelas. Así, en aquellos años era más importante regular la escolarización obligatoria, el acceso a la función docente, la dotación mínima de recursos humanos y materiales o la mejora de la preparación del profesorado. En definitiva se trataba de que no existieran escuelas *de primera* y *de segunda* sino que todas ellas contaran con unos estándares de recursos adecuados. Posteriormente tanto la Ley Orgánica de Calidad de la Educación de 2002 como la Ley Orgánica Educativa de 2006 van más allá al proponer una evaluación orientada a los resultados (o evaluación de diagnóstico). La LOCE, que no llegó a desarrollarse, preveía incluso la realización de pruebas objetivas tanto en primaria como en la secundaria obligatoria y la LOE fijaba las edades de 10 y 14 años como los momentos en los que esta evaluación de diagnóstico era llevada a cabo. Recientemente, la Ley Orgánica para la Mejora de la Ley Educativa (LOMCE) de 2013 prevé cuatro evaluaciones individualizadas (no de diagnóstico) en tercero y sexto de primaria, en cuarto de la educación secundaria obligatoria y al finalizar el bachillerato.

Cabe destacar aquí que la realización de pruebas objetivas recoge solo una parte del verdadero producto de la educación recibida por un individuo que es muy difícil de medir debido a su intangibilidad inherente. La educación no se compone únicamente de la capacidad de repetir información y resolver problemas con más o menos aciertos sino que comprende además otras habilidades no cognitivas como la capacidad de razonar, los buenos modales en la sociedad, la adaptabilidad, el trabajo en equipo y la capacidad de juicio crítico, todos ellos de muy difícil cuantificación aunque sin duda relevantes (Levin, 2012). A pesar de esta naturaleza multi-producto de la educación, la mayoría de las evaluaciones en economía de la educación utilizan los resultados de rendimiento en pruebas objetivas acerca de los conocimientos

que el alumno debería tener en una determinada edad por distintas razones. Entre ellas cabe destacar que son difíciles de falsificar y responden a las actuales exigencias administrativas de calidad para pasar curso e incluso acceder a la universidad. Pero quizás, siguiendo a Hoxby (1999), la más importante sea que tanto políticos como padres de alumnos usan este criterio para evaluar el producto educativo y utilizan la información que se deriva para elegir escuela e incluso lugar de residencia. En este contexto de evaluación del producto educativo, existe un acuerdo prácticamente unánime en todos los países desarrollados en considerar que disciplinas como las matemáticas o la capacidad lectora, son una parte importante del currículo escolar y se consideran competencias vehiculares para la educación de los alumnos.

Teniendo en cuenta estas consideraciones previas podemos definir la evaluación educativa como el proceso generador de información relevante que permite relacionar los recursos, las actividades y los resultados a corto, medio y largo plazo de los centros y programas educativos de una administración observando su evolución en el tiempo para utilizar sus resultados para la toma de decisiones que permitan mejorar la eficacia, la eficiencia y la equidad del sistema educativo.

Las ventajas de la evaluación de cualquier política son numerosas. Por ejemplo la australiana *Steering Committee for the Review of Commonwealth/State Service Provision* (2013) señala las siguientes:

- Ayuda a definir y clarificar el cumplimiento de objetivos.
- Hace que el desempeño sea transparente y permite que tanto ciudadanos como políticos valoren si se cumplen los objetivos de las políticas y programas a lo largo del tiempo.
- Anima a la continua mejora en el desempeño.
- Anima a promover los valores de la eficiencia en el uso de los recursos.
- Promueve la coordinación de las políticas entre las distintas administraciones.
- Facilita la comparación de resultados (*yardstick competition*) entre unidades lo que incentiva la mejora del desempeño.
- Identifica a las mejores unidades para aprender de sus prácticas y tener así una referencia (*benchmarking*) de actuación.

El objetivo de este capítulo es introducir cuales son los aspectos clave y las principales herramientas disponibles de evaluación de la educación no universitaria. No se trata de desgranar el detalle de todas las metodologías de evaluación, pues para ello ya existen múltiples manuales y referencias en economía de la educación, sino de presentar el razonamiento lógico de cómo debería institucionalizarse la evaluación educativa y las técnicas más empleadas para llevar a cabo su correcta ejecución.

Para ello, el siguiente apartado está dedicado a presentar el diseño de una evaluación educativa, la cadena de resultados y los distintos indicadores implicados y necesarios para evaluar. A continuación, en el tercer apartado se detalla por un lado cómo realizar una evaluación de centros educativos utilizando para ello métodos de fronteras de producción y por otro lado los elementos esenciales de una evaluación de programas educativos mediante técnicas de evaluación de impacto. El cuarto apartado está destinado a presentar cómo se implementó el programa Escuela 2.0. y cómo se podría haber evaluado. Finalmente, el capítulo concluye con unas reflexiones finales exponiendo algunas recomendaciones para articular la evaluación educativa en España.

EL DISEÑO DE UNA EVALUACIÓN EDUCATIVA:

CADENA DE RESULTADOS, INDICADORES Y LÍNEA DE BASE

El sistema educativo español debe incorporar las principales técnicas de seguimiento y evaluación orientadas a resultados como pilar fundamental donde asentar el salto cualitativo que demanda la sociedad. Toda la información generada debe además ser transparente a la sociedad civil y a los agentes que forman el sistema educativo para que éstos puedan llevar a cabo una toma de decisiones adecuada.

La evaluación educativa debe estructurarse en una cadena de resultados para analizar tres grandes áreas: eficacia, eficiencia e impacto. La eficacia es una medida del grado de cumplimiento de los indicadores de calidad, resultados y equidad definidos por la política educativa. La eficiencia permite saber si los recursos destinados para cumplir con estos objetivos han sido, o no, los óptimos y por qué. Finalmente la evaluación de impacto permite determinar si existe causalidad para concluir si determinadas políticas y programas educativos han permitido o no mejorar los indicadores de resultados educativos de la población destinataria. En definitiva, el seguimiento y la evaluación permiten informar a los gestores del sistema educativo si éste está funcionando adecuadamente o no. Con esta valiosa información los responsables técnicos y políticos podrán proponer cambios, plantear nuevas acciones, suprimir otras y en definitiva contribuir a maximizar la eficacia, la eficiencia y el impacto de los recursos destinados a educación. Sin una evaluación propia no será posible tomar decisiones racionales basadas en evidencias y en el mejor de los casos se aplicarán recetas que resultaron exitosas en otros países. A continuación se desarrollan los principales pasos y herramientas que debemos utilizar para realizar una correcta evaluación de un sistema educativo¹.

1. Esta descripción no comprende todas las técnicas de evaluación disponibles. Es por ello que debe

CADENA DE RESULTADOS

El primer paso que es necesario dar para poner en marcha una evaluación educativa es una cadena de resultados que describa claramente el marco lógico de cómo las políticas y los programas educativos van a conseguir alcanzar los objetivos planteados que deben estar bien definidos y ser cuantificables. Una cadena de resultados está compuesta por los siguientes eslabones:

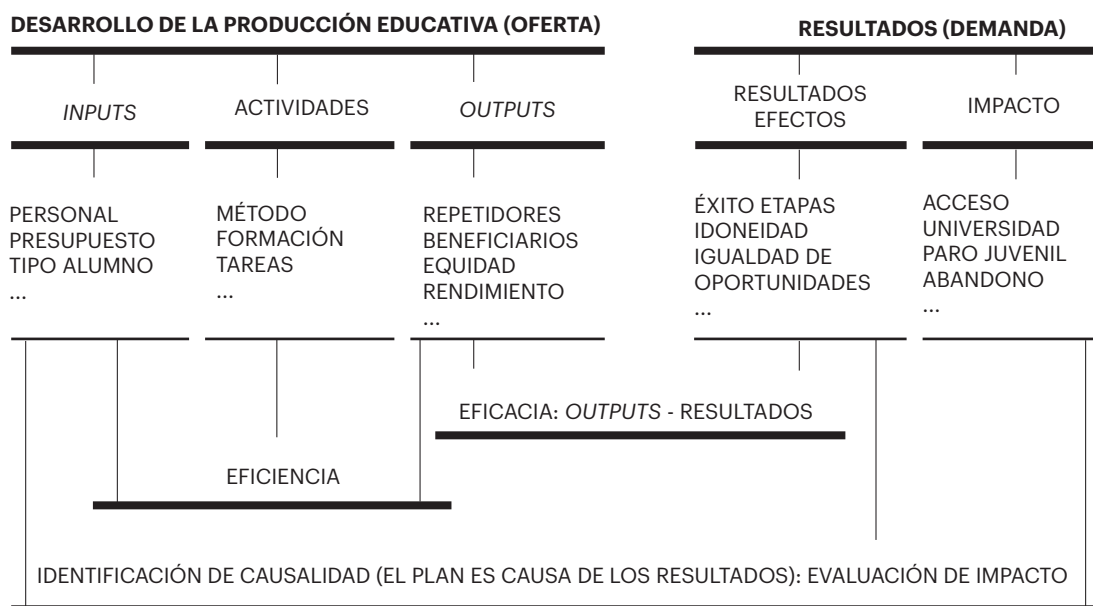
- *Inputs*: Son los recursos necesarios para obtener una cantidad de producción. Necesariamente deben incluir recursos humanos y de capital en sentido amplio.
- *Actividades*: Son aquellas tareas necesarias para transformar los recursos en resultados. No deben confundirse con los *outputs* ya que estas actividades no son un fin en sí mismas si no los medios para conseguir este fin.
- *Outputs*: Producción final perseguida por el programa o la unidad productiva. Se obtienen una vez que se han empleado los *inputs* y se han ejecutado todas las actividades.
- *Resultados*: Resultados a corto y medio plazo que alcanza la población beneficiaria una vez que hace uso de los *outputs* entregados.
- *Impacto*: Grandes y últimos objetivos del programa a los cuales contribuyen las políticas educativas.

La Figura 1 ilustra de forma esquemática cómo se organizaría la cadena de resultados para mostrar la lógica causal de la política educativa.

La cadena de resultados describe una causalidad del tipo “Si..., entonces...”. Por ejemplo, “Si en la escuela se refuerza el número de horas dedicadas a lectura en edades tempranas, entonces mejorará el rendimiento académico no solo en lectura sino también en otras disciplinas” (Machin y McNally, 2008). En este caso se necesitarían recursos para elaborar materiales específicos y una vez distribuido el material éste se entregaría a los alumnos realizando distintas actividades en clase. A partir de trabajar más horas de lectura en clase y en el hogar se esperaría que empiecen a mejorar los niveles de rendimiento académico tanto en cada curso como al final de cada etapa educativa. Cualquier política o programa de innovación educativa debería seguir una lógica similar a esta cadena dotada con sus indicadores correspondientes.

quedar claro que en este trabajo se exponen las técnicas cuantitativas que consideramos que más información aportarían a un proceso de evaluación. Sin embargo, otras técnicas de evaluación, fundamentalmente cualitativas, también pueden ser aplicadas para complementar los resultados del análisis cuantitativo.

FIGURA 1. LA CADENA DE RESULTADOS EN UNA POLÍTICA O PROGRAMA EDUCATIVO



Fuente. Elaboración propia a partir de Gertler et al. (2011)

Por otro lado, en la cadena de resultados educativa se pueden distinguir dos fases que se relacionan directamente con las tareas posteriores de seguimiento y evaluación. En primer lugar, los centros de educación no universitaria, colegios e institutos, llevan a cabo cada curso y en cada nivel educativo un proceso productivo en sus alumnos del tipo, *inputs* –*actividades*– *outputs*; de la siguiente manera. Los colegios a partir de los *inputs* con los que cuentan (recursos escolares, nivel socioeconómico y estructura familiar de los alumnos, conocimientos previos), implementan una serie de actividades (metodología docente, implicación de los padres, tareas en el hogar, pruebas de evaluación, etc.) para transformar e incrementar los conocimientos de los alumnos a lo largo del curso. Debe tenerse en cuenta que el *output* final que se entrega a los alumnos es tan solo el resultado final del curso, el cuál es controlable por la unidad productora, pero no el resultado final de la educación que se obtendrá una vez que cada cohorte finalice el proceso educativo en cada etapa. Cabe también destacar que esta tarea es llevada a cabo de forma repetida en el tiempo por parte de las unidades productivas; escuelas e incluso distintos profesores dentro de la misma escuela. Este proceso debe ser evaluado a través de técnicas de seguimiento y en particular de medición de la eficacia y la evolución de la productividad del proceso productivo.

En segundo lugar, una vez que los alumnos terminan el ciclo educativo podemos empezar a medir los resultados a corto, medio y largo plazo de la educación

entregada por los centros. Estos resultados son los que produce el sistema educativo en forma de tasas de éxitos y rendimiento medio al final de cada etapa (corto y medio plazo) o acceso a la educación superior y al mercado de trabajo (largo plazo). Además, la medición y evolución de los resultados a lo largo del tiempo nos sirve para evaluar la causalidad de los distintos programas educativos que hayan sido puestos en marcha por las autoridades educativas o por centros innovadores. Las técnicas de evaluación de impacto permiten analizar esta relación causal entre los recursos del sistema educativo invertidos por el programa y los resultados finales que se producen en los alumnos a corto, medio y largo plazo.

INDICADORES

Una vez definida la cadena de resultados en una evaluación orientada a resultados a continuación es necesario definir y clarificar los objetivos del sistema educativo, y crear indicadores que permitan su medición y evolución en el tiempo. A la hora de definir y construir indicadores podemos seguir el criterio de UNDP (2009, p. 58) que señalan que un buen indicador debería ser SMART (en inglés un acrónimo para *Specific, Measurable, Achievable, Relevant* y *Time-bound*), es decir debería cumplir las siguientes características:

- Específico: lo más cercano al concepto que se quiere medir.
- Medible: tanto si el indicador es cuantitativo o cualitativo, el indicador debe ser cuantificado de forma clara y no de forma ambigua en relación con lo que se mide.
- Alcanzable: que pueda ser alcanzable y atribuible al programa o política.
- Relevante: que el indicador aunque imperfecto realmente se aproxime a los objetivos perseguidos y sea relevante para la toma de decisiones.
- Con un plazo establecido: los indicadores deben poder obtenerse con la frecuencia deseada a un coste razonable.

El sistema de indicadores de resultados debería además, en la medida que sea posible, lograr el máximo consenso y aceptación por parte de todos los agentes implicados que participan en la evaluación. En particular, quien evalúa, quien financia la evaluación y quien es evaluado, deberían alcanzar un compromiso en el que quede claro que se evalúa para mejorar y no para fiscalizar. Por otro lado los indicadores deben ser multidimensionales, fáciles de entender, recoger y revisar y además recoger aspectos ligados tanto a la cantidad como a la calidad pero también a la “no-calidad”,

en forma de repetición de curso, quejas y reclamaciones por parte de los padres, etc. Con ello se trata de evitar que los gestores se centren únicamente en unos pocos indicadores que primen su desempeño y desprecien el resto (este efecto es lo que Smith (1995) denomina “visión de túnel”).

LÍNEA DE BASE O ESTUDIO DE DIAGNÓSTICO

Una vez definida la cadena de resultados y antes de empezar el proceso de evaluación resulta necesario establecer un diagnóstico o *línea de base* de los indicadores cuyos valores queremos modificar. Por tanto, la línea de base es la primera medición de todos los indicadores contemplados en el diagnóstico que permite establecer el ‘punto de partida’ y de comparación del proyecto o intervención. Suele tener un carácter cuantitativo y puede recurrir tanto a fuentes primarias, producidas *ad-hoc*, como a secundarias, generalmente utilizando otras encuestas o la información contenida en datos administrativos. Sin información de partida será difícil comprobar los progresos reales de las medidas implementadas. La línea de base permite, mediante los datos obtenidos, caracterizar de forma más precisa a la población objetivo del proyecto, y con ello incluso se podrían reformular los objetivos para ganar mayor pertinencia, eficacia, eficiencia e impacto.

ÁREAS DE LA EVALUACIÓN EDUCATIVA: EFICACIA, EFICIENCIA E IMPACTO

EFICACIA

El nivel más básico, aunque no por ello menos importante, en una evaluación educativa es la recogida de todos los indicadores definidos en la cadena de resultados pero especialmente de los referidos a los *outputs* y los resultados a corto plazo². La eficacia mide el grado en el cuál los *outputs* y los resultados deseados son alcanzados. Esta definición lleva implícita que cada indicador debe estar asociado a un objetivo esperado. En este sentido, en las últimas décadas la efectividad de un programa educativo se medía a través del cumplimiento presupuestario. Si bien es necesario seguir recogiendo información acerca de si los recursos asignados y las actividades

2. Debe advertirse que a veces resulta complicado delimitar perfectamente si un indicador es un *output* o un resultado. Es por ello que si bien la eficacia se corresponde con el cumplimiento de los *outputs*, en la Figura 1 y en esta discusión se extiende también el concepto a resultados a corto plazo.

educativas programadas se corresponden con las observadas en la ejecución, resulta aún más necesario poner la atención en los resultados obtenidos (cantidad, calidad, equidad entendida como igualdad de oportunidades educativas y posibles resultados no deseados) ya que la correcta ejecución del gasto es condición necesaria pero no suficiente para que el Plan tenga éxito.

El control de la eficacia debería realizarse mediante un seguimiento de los indicadores clave que permita comparar la realidad del sistema educativo con los objetivos establecidos por la política educativa en el marco de una Ley o programa. A modo de ejemplo la Figura 2 muestra un esquema para llevar el seguimiento de un conjunto de colegios de educación primaria.

FIGURA 2. SEGUIMIENTO DE LA EFICACIA DE COLEGIOS DE EDUCACIÓN PRIMARIA

INDICADORES (DEFINICIÓN)	LÍNEA DE BASE (LB) Y OBJETIVOS PRÓXIMOS DOS AÑOS			RECOGIDA DEL INDICADOR		
	LB	1	2	FRECUENCIA	INSTRUMENTO	RESPONSABLE
Presupuesto en personal (miles €)	49	50	52	Anual 31 Dic	Registros	Colegio
Presupuesto para gastos corrientes (€)	23	24	25	Anual 31 Dic	Registros	Colegio
Alumnos escolarizados (Nº)	280	285	295	Anual 30 Sep	Registros	Colegio
Alumnos por aula (Nº)	23	24	25	Anual 30 Sep	Registros	Colegio
Número de repetidores (%)	25	-	20	Bianual 31 Dic	Registros	Colegio
Rendimiento en matemáticas (prueba)	480	-	490	Bianual 1 Mar	Prueba Objetiva	Ministerio
Nivel estudios padres (años)	10	11	12	Anual 31 Dic	Encuesta	Consejería CC.AA.
...						

EFICIENCIA

La presencia de ineficiencias en la gestión educativa implica que sería posible una mejora paretiana de la misma ya sea mediante la reducción de los recursos presupuestarios sin disminuir los *outputs* y resultados, o alternativamente, incrementando los resultados con los recursos disponibles. Las fuentes de la ineficiencia son diver-

sas. En primer lugar, el marco legal puede condicionar los resultados si los objetivos no están claros y los incentivos del profesorado son limitados. En segundo lugar, los factores educativos pueden explicar parte de las diferencias si asumimos que, no todos los métodos pedagógicos, por ejemplo la organización de la clase, la cantidad de tareas para el hogar, el desdoble de grupos, el tratamiento de la diversidad o la política de decidir la repetición de curso, son igualmente válidos. En tercer lugar, puede existir una incorrecta organización y/o gestión de los centros en ausencia de un marco competitivo (Levin, 1974; Nechyba 2000; Woessman 2001). Finalmente, la calidad, experiencia y carisma del profesor a la hora de transmitir conocimientos también puede explicar estos resultados³.

Para combatir la ineficiencia es necesario cuantificarla. Así, la medición de la eficiencia permite clarificar y relacionar cuales son los recursos, actividades y *outputs* de los colegios e institutos. Las ventajas de la evaluación del desempeño o la eficiencia de los centros educativos a través de su seguimiento son numerosas pero en particular permiten:

- Fomentar y animar la continua mejora en el desempeño.
- Promover los valores de la eficiencia en el uso de los recursos.
- Facilitar la competencia por comparación de resultados (*yardstick competition*) entre colegios lo que incentiva la mejora del desempeño.
- Identificar a las mejores unidades (*benchmarking*) para aprender de sus prácticas y tener así una referencia de actuación.

El seguimiento de la eficiencia y la evolución en el tiempo de la productividad total de los factores para los centros educativos, si se realiza de forma sistemática, permitiría:

- Elaborar un documento de buenas prácticas de gestión que puedan ser trasladadas al resto de centros.
- Analizar de forma detallada las causas que afectan al mal desempeño de los centros y en su caso tomar las medidas para que éstas no se reproduzcan.
- Estudiar la evolución del desempeño en el tiempo de tal forma que el segui-

3. Las fuentes de la ineficiencia son de difícil concreción ya que finalmente estos factores son las razones desconocidas que explicarían por qué la producción es menor que la potencial dado un nivel de recursos. Leibenstein (1966) lo explicaba así: “For a variety of reasons people and organisations normally work neither as hard nor as effectively as they could. In situations where competitive pressure is light, many people will trade the disutility of greater effort, or search the utility of feeling less pressure and of better interpersonal relations”.

miento funcione como un sistema de diagnóstico y detección temprana de los problemas que puedan surgir en cada centro.

- Introducir innovaciones y mejoras en la gestión cuyos efectos puedan ser después cuantificados y en caso de éxito ser extendidos a otras unidades gestoras.

Para medir la eficiencia educativa es necesario estimar la función o frontera de producción a partir de los *inputs* o recursos escolares utilizados y las cantidades de *outputs* producidos⁴. Para ello, el enfoque teórico normalmente utilizado está basado en la función de producción educativa propuesta por Levin (1974), Hanushek (1979) y más recientemente por Hanushek et al. (2013) que se expresa en la Ecuación 1:

$$A_i = f(B_i, S_i) \quad (1)$$

donde el subíndice i se refiere al centro educativo. A_i representa el vector de *outputs* educativos, en general asociados al promedio de los resultados académicos obtenidos por los alumnos en pruebas objetivas. Por otra parte, los *inputs* educativos se dividen en B_i , que representa las características socioeconómicas medias de los hogares de los que provienen los alumnos; aproximadas normalmente por el nivel educativo y profesional de sus padres y el vector S_i que hace referencia a los recursos educativos del centro. Por las razones comentadas con anterioridad, la estimación de dicha función de producción educativa es llevada a cabo frecuentemente considerando la posible existencia de comportamientos ineficientes. En este caso, se estimaría una frontera productiva (Ecuación 2) basada en la observación de las mejores prácticas productivas, es decir, en aquellos centros educativos que maximizan los *outputs* de sus alumnos a partir de los *inputs* que utilizan. La medida de la ineficiencia sería por tanto, la distancia de cada centro a la frontera construida:

$$A_i = f(B_i, S_i) \times u_i \quad (2)$$

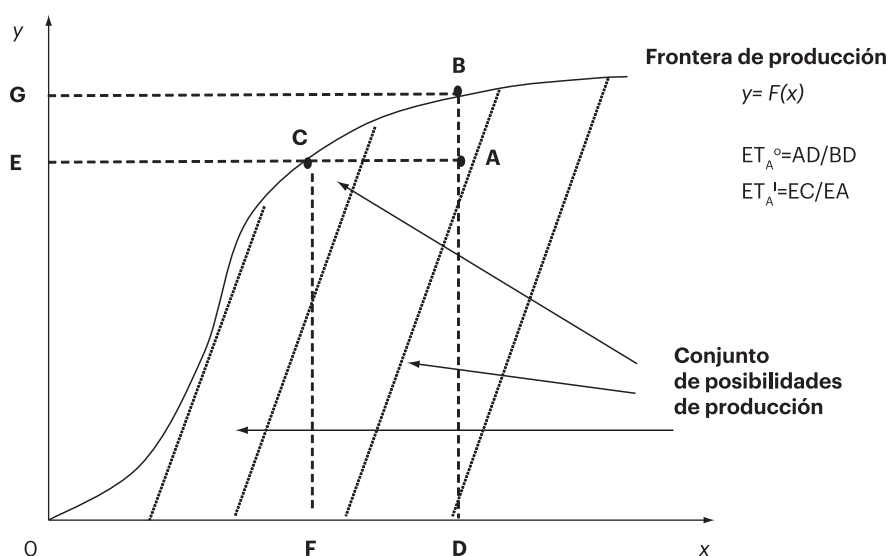
donde u_i representa el nivel de eficiencia de cada escuela. Si $u_i = 1$, ello supondría que el centro educativo i sería plenamente eficiente. Es decir, dado su nivel de *inputs* edu-

4. En ocasiones la estimación de la función de producción tiende a identificarse con una función de producción promedio que sería resultado de una aproximación econométrica sencilla mediante mínimos cuadrados ordinarios. Sin embargo, cualquier libro de microeconomía define una función de producción de la siguiente manera: "A single *output* technology is commonly described by means of a production function $f(z)$ that gives the maximum amount q of *output* that can be produced using *input* amounts $(z_1, \dots, z_{L-1}) > 0$ " (Mas-Colell et al., 1995, p.129).

cativos, el nivel socioeconómico de sus alumnos y los recursos escolares con los que cuenta, y dada la tecnología existente este centro estaría maximizando el nivel de *outputs* que obtiene y por tanto, llevando a cabo una buena gestión de los recursos con los que cuenta. Valores de $u_i < 1$, señalarían que el centro es ineficiente y por tanto la cuantía del índice de eficiencia indicaría la cantidad de *output* se incrementaría si se eliminara la ineficiencia⁵. Veamos un ejemplo donde se ilustra (Figura 3) una función de producción sencilla que únicamente utiliza un único *input* para producir un único *output*.

En la Figura 3 observamos una función de producción muy simple $y = F(x)$ que permite transformar el *input* x en el *output* y mediante una tecnología $F(x)$. Las unidades de producción B y C pertenecen a la frontera productiva esto es, dados los recursos que tienen disponibles no es factible que puedan producir más. Sin embargo, la unidad A pertenece al conjunto de posibilidades de producción. Es por tanto factible que existan unidades, como A, con un desempeño por debajo de la combinación de puntos definida por la frontera. El índice de eficiencia de la unidad A puede ser medido mediante una orientación *output* o mediante una orientación *input*. En el primer caso, $ET_A^O = AD/BD$ expresa la eficiencia en términos de *output* indicando en cuanto podría la unidad A aumentar su *output* dados los recursos de los que dispone.

FIGURA 3. CONCEPTO Y MEDICIÓN DE LA EFICIENCIA TÉCNICA EN TÉRMINOS DE INPUT Y DE OUTPUT



5. Una escuela con un índice de eficiencia; $u_i = 0,9$ indicaría que dados los *inputs* con los que cuenta la escuela y dado el desempeño que están obteniendo otros centros similares con niveles de recursos parecidos, la escuela evaluada podría aumentar en un 10% todos sus *outputs* mediante una mejor gestión.

En el segundo caso, $ET_A^1 = EC/EA$ indica la eficiencia en términos de *input*, es decir, dada la producción de E unidades en cuanto podría reducir sus recursos para llegar a la frontera de producción⁶.

La frontera de producción no es un concepto estático y puede cambiar a lo largo del tiempo. Así, el hecho de que la gestión de un servicio haya sido correcta durante un ejercicio no garantiza que lo será el siguiente. El sistema de seguimiento debe controlar el desempeño de las unidades de forma continuada en el tiempo para con ello poder anticiparse a los problemas que puedan surgir durante la ejecución del Plan. Para realizar este seguimiento continuado es necesario medir también la evolución de la productividad total de los factores, esto es, la evolución de la ratio entre la suma ponderada de *outputs* y de recursos utilizados por las unidades evaluadas.

Las técnicas analíticas empleadas para estimar la frontera de producción y medir la eficiencia y los cambios en la productividad total de los factores deben adaptarse a las peculiaridades de la producción educativa. Fundamentalmente esto pasa por reconocer la ausencia de precios para buena parte de la producción y el carácter multidimensional del *output*. Además, la tecnología de producción $F(\cdot)$ que permite entregar un servicio tan complejo es completamente desconocida lo que aconseja la utilización de aproximaciones flexibles en su formulación y que no impongan supuestos demasiado restrictivos sobre la frontera de producción.

Los principales métodos de medición de la eficiencia técnica pueden ser divididos en dos grandes grupos: los métodos no paramétricos, basados en modelos de optimización matemática y los métodos paramétricos o econométricos⁷. El principal método no paramétrico utilizado en evaluación de políticas públicas es el análisis envolvente de datos (DEA son las siglas en inglés correspondientes con *Data Envelopment Analysis*). Su principal ventaja es que al no asumir ninguna forma funcional acerca de la tecnología que relaciona factores productivos y resultados y con supuestos poco restrictivos, como convexidad y monotonocidad, es capaz de trazar la frontera productiva. Los métodos paramétricos son igualmente válidos pero necesitan hacer supuestos clave acerca de varios aspectos del análisis como la forma funcional de la tecnología, la función de distribución del error o la distribución de la ineficiencia. A cambio permiten calcular elasticidades para conocer la importancia de cada *input* (Crespo-Cebada et al, 2014).

6. Prácticamente la totalidad de los trabajos que estiman fronteras de producción educativa utilizan una orientación al *output*.

7. No es el objetivo de este capítulo realizar una descripción detallada de todas las metodologías disponibles. Para ello puede consultarse Coelli et al. (2005).

En cuanto a la medición de los cambios en la productividad total de los factores, la opción más empleada en el análisis de centros educativos es nuevamente la aproximación de fronteras de producción que tienen en cuenta explícitamente el posible comportamiento ineficiente de las unidades analizadas. El cálculo del índice de Malmquist es uno de los métodos más utilizados para analizar la evolución de la productividad total de los factores y sus componentes a lo largo del tiempo (Maragos y Despotis (2004), Brennan et al. (2014) y Essid et al. (2013)). El éxito de esta medida radica en que solamente necesita datos relativos a cantidades, no es necesario realizar supuestos sobre la forma funcional de la función de producción y permite descomponer la productividad total de los factores de los centros en cambios de productividad debidos a progreso tecnológico y a la mejora de la eficiencia (Coelli, et al., 2005).

Por las consideraciones anteriormente remarcadas relacionadas con la actividad del sector público, los métodos no paramétricos⁸ han sido mucho más utilizados para evaluar centros educativos (Worthington, 2001).

EVALUACIÓN DE IMPACTO

Las políticas educativas deberían estar basadas en evidencias científicas empíricas que muestren una relación causal entre la política o el programa que se desee aplicar y el efecto perseguido. Para cumplir con este objetivo las técnicas de inferencia causal permiten cuantificar el impacto que tiene una política pública sobre una población. Una evaluación de impacto es por tanto la medición del cambio en un indicador relevante (éxito escolar, rendimiento académico, etc.) para los beneficiarios que sea atribuible a un programa educativo. Una evaluación de impacto consiste en realizar un experimento o cuasi-experimento social y para ello requiere la definición de una situación contrafactual o también llamado grupo de control. Lo que se desea evaluar es el comportamiento del indicador de interés con y sin programa. Idealmente, requiere comparar la situación *ex-ante* (Línea de base), previa a la introducción de una política social concreta y *ex-post*, una vez finalizado el programa o en algún punto intermedio del mismo.

La gran ventaja de las técnicas de evaluación de impacto sobre otras técnicas de evaluación del tipo “antes–después” o “con–sin” (programa) es que permite distinguir correlación de causalidad. Obtener evidencia científica de causalidad del tipo $X \rightarrow Y$ es costoso (tiempo, recursos, análisis, etc.). De hecho, en muchas ocasiones

8. Para una descripción detallada sobre los pasos necesarios para aplicar en la práctica un análisis no paramétrico se recomienda la lectura de Emrouznejad y De Witte (2010).

sólo será posible hacer estudios de correlaciones basados en regresiones múltiples.

EL EXPERIMENTO SOCIAL

La idea básica del experimento social es la siguiente. Sea un grupo de individuos que por simplicidad sólo pueden recibir dos tratamientos (D):

- Si no recibe el tratamiento (grupo contrafactual) $D=0$
- Si recibe el tratamiento (grupo tratado) $D=1$

Los individuos son asignados a los dos grupos de forma aleatoria. Después del programa observamos:

- Resultado de los tratados Y_1 en el grupo tratado $D=1$ $E[Y_1|D=1]$
- Resultado de los no tratados Y_0 en el grupo no tratado $D=0$ $E[Y_0|D=0]$

Nuestro objetivo teórico es determinar el efecto que el tratamiento, o en nuestro caso el programa educativo, tuvo sobre el grupo tratado, es decir:

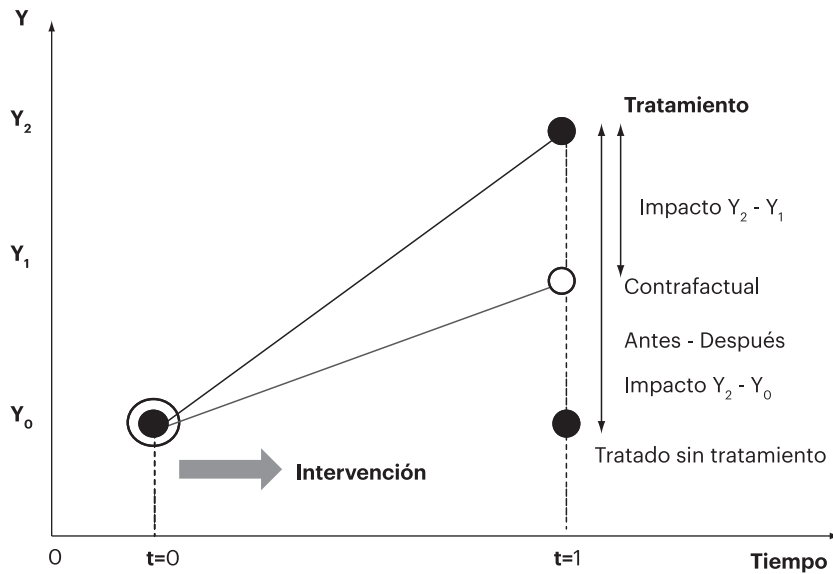
$$E[\Delta|D=1] = E[Y_1|D=1] - E[Y_0|D=1]$$

$E[Y_0|D=1]$ es inobservable, es decir, no es posible saber cuál hubiera sido el resultado del grupo tratado si no hubiera sido tratado. En la práctica la solución pasa por asignar de forma aleatoria a los individuos a los dos grupos mediante un experimento controlado con grupos formados por individuos idénticamente distribuidos en todas las variables. De esta forma se logra que Y (el resultado) y D (si el individuo recibió o no el tratamiento) sean independientes. Por tanto, mediante la aleatorización se consigue que:

$$E[Y_0|D=1] = E[Y_0|D=0]$$

La Figura 4 muestra este proceso. Inicialmente tenemos un grupo de escuelas o alumnos en $t=0$ y mediante un estudio de diagnóstico recogemos información de línea de base de la variable que queremos estudiar Y_0 . Al no haber ningún programa previo diferente los dos grupos presentarán el mismo valor medio. Posteriormente se produce una intervención educativa mediante la implantación de un determinado programa (reducción del tamaño de la clase, continuar con el mismo profesor va-

FIGURA 4. EVALUACIÓN DE IMPACTO DE UN PROGRAMA



rios cursos o cambiar cada curso, utilizar ordenador para enseñar matemáticas, etc.) en uno de los grupos. Tras realizarse esta intervención se vuelve a obtener el valor de la variable de interés que tiene el grupo de individuos tratado en el periodo $t=1$ que ahora presenta un valor Y_2 .

En un estudio de evaluación clásico “antes – después”, si sólo tenemos estos dos puntos (puntos negros de la figura) no sabemos si la variación en la variable de interés Y a lo largo de este tiempo $Y_2 - Y_0$ es debida a los efectos del programa o a otros motivos de la compleja realidad social (otras intervenciones públicas, factores exógenos, etc.).

Por esta razón en una evaluación de impacto se designa aleatoriamente a un grupo de la población como grupo de control (o contrafactual) que esencialmente tiene las mismas características que la población beneficiaria salvo que no recibe el programa para que nos indique que valor presentarían esos individuos en $t=1$ si el programa no se hubiera realizado. Esta situación viene dada por el punto Y_1 , por lo tanto el impacto que tiene este programa vendría dado por la diferencia de los valores entre Y_2 e Y_1 .

Determinar el grupo contrafactual que servirá para hacer la comparación no es tarea fácil y resulta por tanto el elemento clave en el diseño de la evaluación de impacto de un programa educativo. Es necesario recalcar que el grupo de control debe ser lo más parecido al grupo tratado salvo en que no resultó beneficiario del programa que se quiere estudiar.

EL SESGO DE SELECCIÓN

Uno de los problemas más relevantes que pueden afectar a la medición del impacto de un programa es el llamado sesgo de selección. Se produce sesgo de selección cuando los individuos no fueron asignados a los grupos de tratamiento y control de forma aleatorizada. El problema del sesgo de selección en una evaluación de impacto se produce cuando los individuos en los dos grupos (tratado y control) tienen determinadas características que no pueden ser observadas por el evaluador y afectan tanto a la decisión de participar en el programa como a sus resultados. Existen fundamentalmente dos tipos de sesgos:

1. Autoselección o heterogeneidad no observable: características individuales no observables entre los grupos como la motivación, habilidad, perspectivas, información, etc. que influyen tanto en pertenecer al grupo tratado como en los resultados del programa.
2. Heterogeneidad observable: por alguna razón hay diferencias significativas en características observables que afectan bien a recibir el tratamiento o bien a los resultados entre los grupos tratado y de control. Por ejemplo en edad, renta, nivel educativo, tamaño de población, etc. Los resultados del análisis se pueden controlar si se dispone de esta información.

En el caso de una evaluación tradicional, además del caso de la evaluación “antes – después” es habitual analizar individuos tratados y no tratados de forma *ex-post*; o evaluación “con – sin” el programa, sin atender a las causas que determinan el tratamiento. En este caso, se controla por características observables y mediante una regresión múltiple se introduce una variable *dummy* que indica si el individuo fue beneficiario o no del programa. Por ejemplo, consideremos un curso de formación para jóvenes desempleados. Una forma de evaluar este programa sería un año después calcular la tasa de empleo y el salario medio de los que siguieron el curso y de los que no siguieron ningún curso. Sin embargo, es posible que exista un sesgo de selección y que ambos grupos no sean comparables si los individuos que decidieron participar son aquellas personas que tienen más que ganar con el programa y/o están más motivadas con encontrar empleo. Existe por tanto una diferencia no observable entre los dos grupos que sesga los resultados ya que no sabremos si los mejores resultados del grupo tratado son consecuencia de su más alta motivación o del programa de formación seguido⁹. En la práctica, la determinación del grupo tra-

9. Además de la asignación aleatoria de la formación existen otras soluciones prácticas a este problema. Por ejemplo se puede establecer un programa con sobredemanda estableciendo un número de

tado y del grupo de control se puede realizar utilizando distintas metodologías que se dividen en dos grandes categorías: los diseños experimentales (aleatorios) y los diseños cuasi-experimentales (o diseños no aleatorizados)¹⁰.

Los diseños experimentales son los métodos más robustos de todas las metodologías de evaluación de impacto (Angrist, 2004). Partiendo de la población el estudio se realiza de forma bietápica en dos partes. En primer lugar se toma una muestra aleatoria representativa de la población (validez externa). A continuación, se realiza una asignación al azar de la pertenencia al grupo de beneficiarios y al grupo de control de los individuos objeto de estudio, generando dos grupos cuyas características socioeconómicas son estadísticamente equivalentes entre sí (validez interna). La principal ventaja de esta técnica es la simplicidad en la interpretación de resultados. El impacto del programa sobre los resultados se puede medir directamente como la diferencia entre las medias de las muestras del grupo de tratamiento y el grupo de control.

La aleatorización presenta algunos problemas en cuanto a su diseño. En primer lugar pueden existir problemas éticos o restricciones políticas si se considera que toda la población debe tener derecho al programa. En este caso las principales soluciones pasan por rotar los programas o por establecer programas piloto (*phase-in*) que resultan muy deseables ya que son baratos, permiten la innovación y la experimentación educativa y en caso de ser exitosos se pueden extender a las poblaciones que inicialmente sirvieron de grupo de control mitigando así los problemas éticos. En segundo lugar puede existir “contagio” si los grupos tratado y no tratado interactúan o si varios programas son aplicados simultáneamente en una población en cuyo caso será difícil dirimir posteriormente cuál fue el que causó los efectos medidos.

En el caso de que la aleatorización no sea posible se puede recurrir a las técnicas de cuasi-experimentación. Igualmente, si sabemos que los individuos no fueron asignados de forma aleatoria a los grupos tratado y control tenemos que acudir a estas técnicas basadas en la búsqueda de un contrafactual. En la práctica, se recurre a la cuasi-experimentación de forma *ex-post* justo cuando se quiere llevar a cabo la evaluación pero el programa no fue diseñado para ser evaluado.

Los métodos cuasi-experimentales tienen la ventaja de ser más rápidos y baratos ya que su implementación se basa en datos ya existentes donde se dispone

plazas limitadas o un punto de corte donde para el acceso al programa donde se pueda evaluar el programa en el entorno de este punto de corte.

10. Para una revisión detallada de las técnicas de evaluación de impacto tanto experimentales como cuasi-experimentales puede acudir a Gertler et al. (2011) y Khandker et al. (2010). Para el lector no familiarizado se recomienda la lectura de Schlotter et al. (2011) que realizan una excelente descripción de las principales técnicas de evaluación de impacto en referencia al ámbito educativo.

de información de beneficiarios y no beneficiarios. Sus desventajas son su menor fiabilidad respecto a los métodos de aleatorización si no se controla por todas las diferencias entre los dos grupos y requieren además métodos econométricos más complejos. Dependiendo del tipo de información disponible sobre los grupos tratado y contrafactual; “antes – después” de la intervención o sólo después, y de cómo se realizó la asignación de los individuos a los grupos se seleccionará la mejor técnica para cada caso concreto.

UN EJEMPLO: PLAN DE SEGUIMIENTO Y EVALUACIÓN (S&E) DEL PROGRAMA ESCUELA 2.0

En el presente apartado, utilizaremos un programa que fue emblemático en España, el programa Escuela 2.0, implementado entre 2009 y 2012 para ejemplificar, en términos generales, cómo debería diseñarse el proceso de S&E a la hora de implantar un programa educativo.

BREVE DESCRIPCIÓN DEL PROGRAMA ESCUELA 2.0

El Programa Escuela 2.0 tuvo por objetivo integrar las tecnologías de la información y la comunicación (TICs) en los centros educativos sostenidos con fondos públicos. Para ello, el programa consistía en dotar a los alumnos y profesores con un ordenador portátil, proveer la conectividad a internet y la interconectividad dentro del aula para todos los equipos, promover la formación del profesorado y generar y facilitar el acceso a materiales digitales educativos ajustados a los diseños curriculares tanto para profesores como para el alumnado y sus familias. A su vez, se buscaba implicar a los alumnos y a las familias en la adquisición, custodia y uso de estos recursos.

El programa comenzó en el curso 2009-10 con 5º de Primaria, y posteriormente, se extendería a 6º de Primaria y a los dos primeros cursos de la ESO. El presupuesto inicial del programa previsto era de 200 millones de euros por curso, cofinanciados al 50% entre el Gobierno y las Comunidades Autónomas (CC. AA.) que participaron. En 2012 el programa fue cerrado y sustituido por un nuevo programa de cooperación territorial de TICs. Por tanto, en los tres años que se implementó el programa se invirtieron alrededor de 600 millones de euros.

Existen numerosos estudios en la literatura de la Economía de la Educación que demuestran que más recursos tecnológicos (ordenadores, salas digitales, con-

xión a internet, etc.) no producen *per se* mejores resultados en el aprendizaje de los alumnos (Angrist and Lavy, 2002; Fuchs and Woessmann, 2004; Banerjee et al., 2007; Warschaver 2008). Una posible explicación radica en que introducir la tecnología para “informatizar” procesos y prácticas educativas existentes no implica una innovación real en el proceso de aprendizaje. Esto es, el éxito del programa Escuela 2.0 requiere además de la provisión de los recursos tecnológicos de acciones y actividades que impliquen un cambio en las prácticas educativas tradicionales que aseguren una adecuada y eficaz utilización de los recursos para lograr los resultados y que se definan. En este sentido, un diseño *ex-ante* de la evaluación, un oportuno plan de seguimiento durante la implementación del programa y una adecuada evaluación *ex-post* hubieran sido fundamentales para conocer si efectivamente el programa fue implementado eficientemente y ha logrado impactar positivamente en los resultados educativos de los estudiantes. A continuación, se presenta una propuesta de algunos elementos básicos que conforman un adecuado plan de S&E de una intervención educativa que hubieran podido aplicarse al programa Escuela 2.0.

ELEMENTOS DE UN PLAN DE S&E PARA EL PROGRAMA ESCUELA 2.0

IMPLEMENTACIÓN DE UN PLAN PILOTO

Los programas de alto contenido tecnológico como el Escuela 2.0 implican la inversión de una gran cantidad de recursos económicos. Hubiera sido por tanto adecuado instaurar el programa inicialmente solo en algunos centros seleccionados aleatoriamente y realizar un seguimiento y evaluación del plan piloto. Como hemos comentado, la idea de llevar a cabo una experiencia piloto es poder conocer efectivamente si el programa mejora el aprendizaje y las competencias de los estudiantes, y a su vez conocer cuáles son las formas de utilizar los recursos asignados que producen mejores resultados y cuáles no tienen efectos. Esta información permitiría en primer lugar, corroborar si el impacto es positivo y significativo, y en base a ello calcular cuáles serían los costes y beneficios de extender el programa a toda la población. Es decir, si tiene sentido o no aplicar el programa a nivel nacional o sobre qué colectivos tendría sentido aplicarlo. Por tanto, la evaluación piloto permite implementar el programa de forma eficiente.

Sin embargo, este plan piloto no se llevó a cabo con el programa Escuela 2.0. En ocasiones la idea de experiencia piloto se confunde con implementar un programa en etapas, y detectar y corregir sobre la marcha simplemente problemas de logística en la provisión de los recursos (ordenadores, instalación de redes y aulas digitales). Sin em-

bargo, el plan piloto requiere un verdadero S&E de la experiencia para obtener información de si efectivamente se alcanzan los resultados esperados con el programa.

DEFINICIÓN DE OBJETIVOS Y LÍNEA DE BASE DEL S&E, ¿QUÉ QUEREMOS EVALUAR?

El S&E de un programa o política debe tener objetivos claros, esto es, definir inequívocamente qué cuestiones se quieren analizar y dar respuestas. Así por ejemplo, un plan de seguimiento del programa Escuela 2.0 debería poder responder a las siguientes interrogantes: ¿cuántos ordenadores se han entregado efectivamente?, ¿cuántos se estropean y en cuánto tiempo?, ¿cuáles son los centros que mejor resultado en competencia digital obtienen dados los recursos asignados? (*Benchmarking*), ¿qué actividades llevan a cabo estos centros?, ¿qué acciones están implementando los centros menos eficientes?

En el caso de la evaluación de impacto, las preguntas debería enfocarse en los resultados de medio plazo, o en el impacto del programa: ¿qué impacto ha tenido Escuela 2.0 en el aprendizaje de competencia digital?, ¿qué impacto ha tenido Escuela 2.0 en el aprendizaje de otras disciplinas?, ¿ha tenido Escuela 2.0 algún impacto sobre la tasa de repetición o fracaso escolar?, ¿ha tenido Escuela 2.0 algún impacto sobre la tasa de abandono a temprana edad?, ¿cuál es la satisfacción de los padres y los profesores con el programa?

Junto a estas preguntas es necesario recabar información de la situación inicial antes de implementar el programa. Este hecho es crucial para poder llevar a cabo una evaluación de impacto del programa. En este caso, sería necesario disponer de información acerca del nivel de conocimientos en competencia digital (y otras competencias básicas) que los alumnos poseen antes de la intervención, así como de sus características socioeconómicas. A su vez, para poder llevar a cabo una comparación justa de los logros alcanzado por cada centro educativo, sería necesario controlar el impacto de la intervención por otros factores que también afectan al proceso de aprendizaje y por tanto al rendimiento de los alumnos. En este sentido, la línea de base debe proporcionar también información, por ejemplo, acerca del nivel de conocimiento y experiencia previo que tienen los profesores en materia de TICs así como de la disponibilidad inicial de TICs en los centros educativos.

DEFINICIÓN DE LOS INDICADORES DE LA CADENA DE RESULTADOS DEL PROGRAMA

INDICADORES DE RECURSOS (*INPUTS*)

Los *inputs* pueden definirse tanto en términos monetarios como en recursos físicos. En el caso del programa Escuela 2.0, no se asignó un presupuesto a cada unidad ejecutora (centro educativo), sino que se otorgaron los recursos físicos directamente y por tanto no se dispone de información presupuestal del programa a nivel de centro educativo. Los *inputs* deberían definirse entonces en términos físicos o no monetarios¹¹. Sin ánimo de exhaustividad, algunos ejemplos de indicadores de *inputs* para el programa Escuela 2.0 podrían ser:

- Número de ordenadores entregados a los alumnos.
- Número de alumnos potenciales receptores de ordenador en el centro.
- Número de ordenadores entregados a los profesores.
- Número de profesores potenciales receptores de ordenador en el centro.
- Número de aulas digitales instaladas.
- Número de horas de formación a profesores.
- Características del alumnado:
 - Índice socioeconómico medio de los alumnos del centro (aproximado por el nivel educativo y recursos económicos de los padres).
 - Otras características familiares.
 - Resultados en pruebas de competencia digital y competencias básicas antes de aplicar el programa.
- Características (previas al programa) del centro:
 - Número de ordenadores en el centro.
 - Si contaban o no con aula digital.
 - Número de horas de uso de los ordenadores por los alumnos.
- ...

INDICADORES DE ACTIVIDADES

Deberían reflejar las acciones emprendidas por cada centro con los recursos tecnológicos otorgados por el programa. Como hemos mencionado, los recursos en sí

11. Podría llevarse a cabo un ejercicio para recuperar esta cifra imputando el coste de los recursos otorgados a cada centro a partir del coste unitario de los ordenadores, aulas digitales, horas de formación, etc. que recibe cada centro educativo.

mismos no producen mejoras en el aprendizaje de los alumnos, sino a través de los métodos docentes que los emplean. Los indicadores de actividad son fundamentales sobre todo a la hora de llevar a cabo el seguimiento del programa, ya son estos indicadores permiten explicar o al menos disponer de hipótesis acerca de por qué se producen diferencias en los resultados del programa entre centros. Algunos ejemplos de actividades del programa Escuela 2.0 podrían ser:

- Número de horas de formación en cursos brindados a los docentes en el centro.
- Porcentaje de docentes que han asistido a los cursos de formación del centro.
- Resultados de los profesores tras una evaluación de sus competencias.
- El centro provee de contenidos educativos para el aprendizaje.
- El centro brinda apoyo institucional a proyectos educativos innovadores.
- El centro proporciona materiales digitales extras.
- Número de entrevistas, tutorías y contactos virtuales con las familias al mes.
- Número de horas de clase con portátil y/o en aulas digitales.
- El centro posee un coordinador TIC
- El centro posee un técnico informático o servicio técnico online.
- El centro ha tenido alguna experiencia previa en programas TICs
- Porcentaje de alumnos/docentes/familias que utilizan redes y portales.
- Uso de software libre / Uso de software convencional.
- ...

INDICADORES DE PRODUCTOS (*OUTPUTS*)

Estos indicadores deben reflejar y cuantificar la producción final obtenida a partir de los recursos tecnológicos proporcionados por el programa y de las acciones llevadas a cabo en las aulas y centros. Como ya se ha señalado muchas veces la distinción entre producto y resultado es difusa y, por tanto, su definición final dependerá de los objetivos establecidos en el programa y de qué se quiere evaluar. En este caso, si el objetivo del programa fuera mejorar las capacidades digitales de los estudiantes y que en el mediano plazo esto impacte positivamente en los resultados de otras competencias, los indicadores de productos deberían al menos cuantificar el aprendizaje en competencia digital:

- Resultado en pruebas de competencia digital al finalizar el curso.
- Porcentaje de alumnos por encima del umbral mínimo de competencia digital.
- Diferencia en competencia digital entre alumnos del decil más alto y bajo.

Un enfoque complementario sería fijar los objetivos en función del uso efectivo de los recursos tecnológicos y del grado de aceptación y satisfacción con los mismos de los principales agentes involucrados en el proceso educativo:

- Número de ordenadores entregados que se conservan en correcto funcionamiento.
- Grado de satisfacción de las familias / profesores.
- Grado de satisfacción de los alumnos con el uso de TICs.
- Grado de satisfacción del profesorado con los cursos recibidos.

INDICADORES DE RESULTADOS

Como hemos mencionado, el objetivo a medio plazo de un programa educativo debería ser mejorar los logros académicos de los estudiantes y por tanto la calidad del sistema educativo en su conjunto. En este sentido, posibles los indicadores de resultados del programa Escuela 2.0 serían:

- Resultado en pruebas de competencia digital al finalizar la etapa educativa.
- Resultados en pruebas objetivas de competencias básicas al finalizar la etapa educativa.
- Tasa de repetición en educación secundaria obligatoria.
- Diferencias en los resultados de las pruebas de competencias básicas y digitales entre distintos grupos de alumnos.

INDICADORES DE IMPACTO

Finalmente, deberían establecerse cuáles son los grandes y últimos objetivos del programa. Algunos ejemplos serían.

- Tasa de abandono educativo.
- Porcentaje de alumnos que acceden a carreras con base tecnológica.
- Aumento del empleo en sectores con base tecnológica

Una vez definidos los indicadores del plan de S&E es preciso determinar cómo se va a recabar la información necesaria antes, durante y después de la implementación del programa. Existen diversas fuentes y métodos de recopilación de la información y la selección final depende normalmente de los recursos disponibles. Para el caso del programa Escuela 2.0 la información podría obtenerse de las siguientes fuentes:

- Pruebas de conocimiento objetivas aplicadas a nivel nacional en competencia digital al inicio y al final del programa. Además de la prueba, sería adecuado aplicar una encuesta personal a los alumnos, directores y profesores para recabar las características iniciales comentadas anteriormente de centros y alumnos.
- Evaluaciones externas en competencias básicas, de la CC. AA o a nivel nacional (Evaluaciones Generales de Diagnóstico) que se apliquen antes y después del programa.
- Evaluaciones internacionales en competencias básicas (PISA, TIMSS, PIRLS) que se hayan aplicado antes y después del programa.
- Registros administrativos.
- Otros métodos: observación de aulas, grupos de discusión, etc.

EVALUACIONES DEL PROGRAMA ESCUELA 2.0.

Hasta el momento existen dos grandes evaluaciones del programa Escuela 2.0 llevadas a cabo por grupos externos de investigación financiadas por el entonces llamado Ministerio de Ciencia e Innovación.

PROYECTO TISCE 2.0

En el Proyecto *Las políticas de un “ordenador por niño” en España. Visiones y prácticas del profesorado ante el programa escuela 2.0. Un análisis comparado entre Comunidades Autónomas*¹², financiado por el Plan Nacional de investigación 2010 del entonces llamado Ministerio de Ciencia e Innovación, participan 13 universidades españolas y la U. de Coimbra de Portugal, y es impulsado por el grupo EDULLAB (U. de La Laguna). La evaluación está comprendida para ser ejecutada entre 2011 y 2014.

La evaluación tiene por objetivos en primer lugar conocer las opiniones, expectativas y valoraciones del profesorado. En segundo lugar, explorar el uso de prácticas de enseñanza o actividades didácticas con TIC en el aula, comparar entre CC.AA y finalmente crear un observatorio sobre políticas TIC (países iberoamericanos). Para ello el proyecto encuesta a profesores y alumnos y realiza estudios de casos de aulas en distintas CC. AA que aplican y no aplican el programa Escuela 2.0. con el fin de responder a las siguiente preguntas:

12. Para mayor detalle del proyecto ver <http://tecedu.webs.ull.es/ticse20/index.php?lang=es>.

- ¿Qué efectos e impacto pedagógico y organizativo tendrá esta abundancia de tecnologías en las aulas ordinarias?
- ¿En qué medida cambiará la metodología de enseñanza desarrollada en la clase y el aprendizaje del alumnado?
- ¿Qué impacto tendrá sobre los libros de texto y materiales didácticos tradicionales?
- ¿Qué visiones y opiniones tiene el profesorado sobre este programa?
- ¿Existen diferencias en estas visiones y prácticas docentes por Comunidades Autónomas?

PROYECTO GRUPO LOCALCOM

El proyecto *Las políticas públicas de impulso a la sociedad de la información en España: evaluación de las estrategias y actuaciones en nueve comunidades autónomas*¹³ forma también parte del Plan Nacional de investigación, financiado por el entonces llamado Ministerio de Ciencia e Innovación.

En este caso la metodología consiste en un modelo propio de análisis métrico de 24 indicadores (Índice Localcom, basado en tres categorías: oportunidades, infraestructuras y usos); y un modelo descriptivo para el estudio de las políticas autonómicas en TICs. Los indicadores se agrupan en siete dimensiones: pertinencia, efectividad, coherencia, permanencia, eficacia, impacto y eficiencia. Por ejemplo, en este caso los indicadores de eficacia, eficiencia e impacto son los siguientes:

Eficacia:

- ¿Se han obtenido distintos resultados según la tipología de los centros?
- ¿Cuál es la titularidad de los contenidos educativos digitales?
- ¿Se han implementado entornos de aprendizaje virtuales para el diseño, seguimiento y evaluación de actividades educativas?

Eficiencia

- ¿Se utilizaron herramientas o recursos existentes?
- ¿Se impulsó una entidad con participación de todos los agentes implicados?
- ¿El proyecto estuvo adscrito a una estructura institucional única?

Impacto

13. Para mayor detalle del proyecto ver <http://centresderecerca.uab.cat/oic/content/politicas-publicas-sociedad-informacion>.

- Número de ordenadores para estudiantes (acumulado en el tiempo).
- Número de aulas digitales (acumulado en el tiempo).
- Número de profesores formados (acumulado en el tiempo).

Como puede observarse, ambas evaluaciones son valiosas ya que pondrán a disposición nueva información relevante acerca de la implementación del programa Escuela 2.0 y de los resultados obtenidos. Sin embargo, en ninguna de ellas el objetivo final es llevar a cabo una evaluación de la eficiencia de los centros o una evaluación de impacto sobre las competencias desarrolladas por los alumnos. No obstante, de la información que se derive será posible construir algunos de los indicadores de *inputs*, actividades, productos, resultados e impacto comentados anteriormente, que den lugar en el futuro a evaluaciones cuasi experimentales sobre el impacto que realmente ha tenido el programa en los alumnos beneficiarios del mismo.

REFLEXIONES FINALES

La institucionalización de la evaluación de los centros y programas educativos debería ser el eje que vertebrase del sistema educativo español en los próximos años. Así, cualquier nuevo programa o intervención para la mejora de la educación debería ser acompañado desde su diseño de un documento que detalle cómo será evaluado.

El primer requisito para llevar a cabo una adecuada evaluación es disponer de información de calidad en diversos niveles del sistema educativo. En este sentido la reciente introducción de pruebas individualizadas debería contribuir a mejorar la transparencia del sistema educativo y proveer información cualificada para que los diversos agentes implicados en el proceso educativo tomen mejores decisiones basados en la evidencia empírica. Los padres deberían tener información de los conocimientos que sus hijos efectivamente adquieren en la escuela que eligieron teniendo además en cuenta el nivel medio que alcanzan otros niños en otros colegios del barrio, municipio o municipios cercanos. A su vez, las autoridades educativas deberían usar esta información para llevar a cabo el seguimiento y la evaluación de impacto de las medidas e iniciativas educativas que se vayan poniendo en marcha.

En segundo lugar, España posee un grave problema de abandono y fracaso escolar. Es por ello que la evaluación continua debe ser un mecanismo esencial a la hora de detectar a los alumnos con alto riesgo de fracaso escolar para intervenir tempranamente desde la escuela y evitar que la contingencia efectivamente suceda garantizando de este modo la igualdad de oportunidades educativas. La evaluación de impacto de los programas de intervención que se pongan en marcha en este sentido

deben ser la herramienta que garantice tanto la consecución de los objetivos como la eficiencia de los recursos invertidos.

Finalmente, el seguimiento y la evaluación educativa debería ser realizada por una entidad externa a la administración educativa lo más independiente posible pero contando con la colaboración de las instituciones evaluadas responsables del diseño y ejecución de la intervención educativa. En el supuesto de que la evaluación se lleve a cabo a nivel regional o local, una opción podría ser crear Agencias Autonómicas de Evaluación Educativa que dependan exclusivamente del Presidente de la Comunidad Autónoma y que por lo tanto tengan plena capacidad e independencia para llevar a cabo las evaluaciones. La segunda opción sería la contratación de equipos de consultores externos independientes que realicen la evaluación en estrecha colaboración con las Consejerías de Educación e incluso incorporando en el equipo evaluador algún miembro de la propia Consejería que conozca el detalle del programa evaluado.

En definitiva, las políticas educativas deberían estar basadas en evidencias científicas empíricas que permitan identificar cuáles centros son los que mejor desempeñan y gestionan sus recursos y por qué y a su vez, establecer la relación causal entre un programa y el efecto perseguido. La importancia de evaluar se puede resumir utilizando la frase del célebre físico y matemático William Thomson quien afirmó: “Lo que no se define no se puede medir. Lo que no se mide, no se puede mejorar. Lo que no se mejora, se degrada siempre”.

REFERENCIAS

- Angrist, J. y Lavy, V. (2002). “New Evidence on Classroom Computers and Pupil Learning*”. *The Economic Journal*, 112(482), 735-765.
- Angrist, J. D. (2004). “American education research changes tack”. *Oxford Review of Economic Policy*, 20(2), 198-212.
- Banerjee, A. V., Cole, S., Duflo, E. y Linden, L. (2007). “Remedying education: Evidence from two randomized experiments in India”. *The Quarterly Journal of Economics*, 122(3), 1235-1264.
- Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*: University of Chicago Press.
- Brennan, S., Haelermans, C. y Ruggiero, J. (2014). “Nonparametric estimation of education productivity incorporating nondiscretionary inputs with an application to Dutch schools”. *European Journal of Operational Research*, 234(3), 809-818.

- Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. y Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*: Springer.
- Crespo-Cebada, E., Pedraja-Chaparro, F. y Santín, D. (2014). "Does school ownership matter? An unbiased efficiency comparison for regions of Spain". *Journal of Productivity Analysis*, 41(1), 153-172.
- Emrouznejad, A. y De Witte, K. (2010). "COOPER-framework: A unified process for non-parametric projects". *European Journal of Operational Research*, 207 (3), 1573-1586.
- Essid, H., Ouellette, P. y Vigeant, S. (2014). "Productivity, efficiency, and technical change of Tunisian schools: a bootstrapped Malmquist approach with quasi-fixed inputs". *Omega*, 42(1), 88-97.
- Fuchs, T. y Woessmann, L. (2004). "What accounts for international differences in student performance? : CESifo working papers".
- Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B. y Vermeersch, C. M. (2011). *Impact Evaluation in Practice; La Evaluación de Impacto en la Práctica*: World Bank Publications.
- Hanushek, E. A. (1979). "Conceptual and empirical issues in the estimation of educational production functions". *Journal of human Resources*, 14(3).
- Hanushek, E. A., Link, S. y Woessmann, L. (2013). "Does school autonomy make sense everywhere? Panel estimates from PISA". *Journal of Development Economics*, 104, 212-232.
- Hoxby, C. M. (1999). "The productivity of schools and other local public goods producers". *Journal of Public Economics*, 74(1), 1-30.
- Khandker, S. R., Koolwal, G. B. y Samad, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices*: World Bank Publications.
- Leibenstein, H. (1966). Allocative efficiency vs. "X-efficiency". *The American Economic Review*, 392-415.
- Levin, H. M. (1974). "Measuring efficiency in educational production". *Public Finance Review*, 2(1), 3-24.
- Levin, H. M. (2012). "More than just test scores". *Prospects*, 42 (3), 269-284.
- Lochner, L. y Moretti, E. (2001). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports: National Bureau of Economic Research.
- Machin, S. y McNally, S. (2008). "The literacy hour". *Journal of Public Economics*, 92(5), 1441-1462.
- Maragos, E. K. y Despotis, D. K. (2004). "Evaluating School Performance over Time in the frame of Regional Socio-Economic Specificities". *WSEAS Transactions on Mathematics*, 3(3), 664-670.

- Mas-Colell, A., Whinston, M. D. y Green, J. R. (1995). *Microeconomic theory* (Vol. 1): Oxford university press New York.
- Nechyba, T. J. (2000). "Mobility, targeting, and private-school vouchers". *American Economic Review*, 130-146.
- Schlotter, M., Schwerdt, G. y Woessmann, L. (2011). "Econometric methods for causal evaluation of education policies and practices: a non-technical guide". *Education Economics*, 19(2), 109-137.
- Smith, P. (1995). "On the unintended consequences of publishing performance data in the public sector". *International Journal of Public Administration*, 18(2-3), 277-310.
- Steering Committee for the Review of Government Service Provision (2013). *Report on Government Services 2013*, Productivity Commission, Canberra.
- Taylor-Powell, E. y Henert, E. (2008). "Developing a logic model: Teaching and training guide". *Benefits*, 3, 22.
- UNDP (2009). *Handbook of Planning, Monitoring and Evaluating for Development Results*, New York.
- Warschauer, M. (2008). "Laptops and literacy: A multi-site case study". *Pedagogies: An International Journal*, 3(1), 52-67.
- Worthington, A. C. (2001). "An empirical survey of frontier efficiency measurement techniques in education". *Education Economics*, 9(3), 245-268.
- Wößmann, L. (2001). "Why Students in Some Countries Do Better: International Evidence on the Importance of Education Policy". *Education Matters*, 1(2), 67-74.



**FUNDACIÓN
RAMÓN ARECES**

Vitruvio, 5 – 28006 Madrid
www.fundacionareces.es
www.fundacionareces.tv

Fundación Europea Sociedad y Educación
European Foundation Society and Education

José Abascal, 57 – 28003 Madrid
www.sociedadeducacion.org