



Facultad
de Ciencias
Económicas y
Empresariales

Departamento
de Economía
Aplicada y
Estadística



**Revista de Evaluación de
Programas y Políticas Públicas**
JOURNAL OF PUBLIC PROGRAMS AND POLICY EVALUATION

Evaluaciones de valor añadido en el ámbito educativo

Núm. 2 (2014), pp. 55-77.

Jorge Calero¹

Rodrigo Travitzki ²

Recibido: Enero, **2014**

Aceptado: Febrero, **2014**

JEL Clasif: JEL-class: H.52, I.21

¹ Universidad de Barcelona e IEB. jorge.calero@ub.edu

² Departamento de Filosofía de la Educación y Ciencias de la Educación, Facultad de Educación, Universidad de Sao Paulo, email: r.travitzki@gmail.com

Abstract

This article aims at analyzing the potentialities of value-added models in the assessment of educational competences. First, a series of alternative functions which may be allocated to competence assessment are presented and discussed. These functions are defined as depending of, on the one hand, the agent which receives the information and, on the other, the intensity in the response of the agent (low-stakes vs. high-stakes assessment). The resulting ‘map’ of functions is useful to establish the field in which value added models work. These models allow, simultaneously, controlling for the effect of variables in the educational production function which are not under direct control of the school (e.g., the socio-cultural and economic background of students) and analyzing the long-term evolution of the student’s performance. The optimal characteristics of a value-added model are also discussed in the article. Taking into account the availability of results derived from value-added models, the article discusses, finally, which evaluations results should be published and in which format.

Key Words: value-added models, educational assessment, standardized evaluations, educational efficiency, competences.

Resumen

El artículo se plantea como objetivo analizar las potencialidades de la aproximación que proporcionan los modelos de valor añadido a la evaluación de competencias educativas. Se presentan y discuten, inicialmente, una diversidad de funciones asignables potencialmente a la evaluación de competencias, dependiendo del agente receptor de la información y de la intensidad de la respuesta del agente (low-stakes vs. high-stakes). El “mapa” de funciones descrito sirve para delimitar el campo en el que actúan los modelos de valor añadido. Éstos permiten, simultáneamente, controlar el efecto de variables de la función de producción educativa no controlables directamente por el centro educativo (como el origen sociocultural y económico de los alumnos, entre otros) y analizar la evolución a lo largo del tiempo del rendimiento del estudiante. En el artículo se plantean las condiciones deseables que debe reunir un modelo de valor añadido. Teniendo en cuenta la disponibilidad de resultados de modelos de valor añadido, el artículo discute, adicionalmente, qué tipo de resultados de la evaluación deben hacerse públicos y con qué formato.

Palabras clave: valores de valor añadido, evaluación educativa, evaluaciones estandarizadas, eficiencia educativa, competencias.

1. Introducción

La evaluación de competencias ha pasado a ocupar, en algo menos de dos décadas, una posición central en los sistemas educativos, hasta el punto que podemos hablar de una “colonización” de estos (a nivel global) por parte de la evaluación. El tipo de información que proporciona la evaluación de competencias es susceptible de cumplir funciones muy diversas, útiles para una diversidad de actores. Desde el punto de vista de la política educativa, por tanto, la implantación de evaluaciones de competencias puede estar guiada por intereses muy diferentes. Estos intereses pueden ser, por ejemplo, la disponibilidad de información en la que basar la rendición de cuentas de los centros educativos al sector público (en el extremo de esta visión, la responsabilidad del sector público se podría limitar a la evaluación, lo que nos lleva al concepto de “estado evaluador” - véase Elliot, 2002). Pero otro interés de la evaluación de competencias puede residir en la difusión de información a las familias para que estas puedan efectuar una elección de centro más eficaz.

Dentro de la tendencia expansiva de la evaluación de competencias hemos asistido, en los años más recientes, a la introducción paulatina de mejoras metodológicas en el tratamiento de los resultados, en un primer momento dirigidas a aislar el efecto-escuela del efecto de las variables de contexto. Posteriormente se han extendido las propuestas (con mayor o menor éxito cuando han sido llevadas a la práctica) de modelos de valor añadido, donde al control de variables contextuales se añade la visión longitudinal. Las exigencias de estas nuevas metodologías en términos de datos complejos y precisos son elevadas, por lo que por el momento asistimos, en muchos países, a procesos plagados de dificultades. No conviene, tampoco, olvidar los potenciales efectos negativos de la evaluación de competencias (véase Ravitch, 2010, Toch, 2006 y Calero y Choi, 2013), que pueden aparecer con independencia del nivel de sofisticación de la metodología con la que se analicen los resultados.

Teniendo en cuenta el marco descrito, en este artículo nos planteamos como objetivo analizar las potencialidades de la aproximación de los modelos de valor añadido a la evaluación de competencias educativas. Para proporcionar un marco adecuado a la metodología del valor añadido, en el apartado 2 analizamos las diversas funciones que puede cumplir la evaluación de competencias, dependiendo del agente destinatario de la información de la evaluación y del uso que de ella hace (low-stakes vs. high-stakes). El tercer apartado se dedica a la metodología que permite identificar

los efectos-escuela y, más adelante, a los modelos de valor añadido, describiéndose una serie de requisitos para tales modelos. A partir de tales requisitos, se propone un modelo de valor añadido simple que los satisface. El apartado 4 se centra en el debate acerca de la difusión de los resultados de la evaluación de competencias que, como veremos, viene en buena medida definido por el tipo de función a la que se orienta la evaluación. En el apartado 5 aportamos unas conclusiones globales.

2. Las funciones alternativas de la evaluación de competencias educativas

Proponemos aquí una clasificación de las funciones alternativas de la evaluación de competencias educativas establecida a partir del cruce de dos variables: por una parte, el agente destinatario de la información contenida en la evaluación y, por otra, la intensidad de la respuesta, por parte del agente, que se deriva de la evaluación. Los cinco agentes que consideramos son los siguientes: el sector público (como financiador), algunas instituciones internacionales, los centros educativos, las familias y, finalmente, la comunidad investigadora. Por lo que respecta a la intensidad de la respuesta, utilizamos la distinción anglosajona entre “low-stakes” y “high-stakes”. En el primer caso la intensidad de la respuesta es baja en el sentido de que ésta no se orienta a provocar un cambio inmediato en el comportamiento del resto de agentes (a través de la aplicación de incentivos o desincentivos, por ejemplo, en el caso de la actuación del sector público como financiador o de los centros educativos). En el segundo caso, “high-stakes”, la intensidad de la respuesta es alta.

El cuadro 1 sintetiza, a partir del cruce de las variables expuestas, un conjunto de funciones posibles para la evaluación de competencias; para cada una de las funciones se especifica, entre paréntesis, si el tipo de evaluación necesaria es de carácter censal (C) o puede también ser de carácter muestral (C/M). Adicionalmente, es preciso tener en cuenta que un único instrumento de evaluación puede tener diferentes funciones simultáneas, al ser utilizado por varios agentes con una diversidad de respuestas.

Cuando es el sector público el destinatario de los resultados de la evaluación de competencias y la intensidad de la respuesta que éste da es baja nos encontramos con las evaluaciones de sistema, que monitorean de forma general el sistema y su evolución.

Es posible, también en esta combinación de agente y respuesta, que el instrumento se utilice para evaluar el resultado esperado o real de intervenciones específicas, o que se emplee para identificar buenas prácticas. El sector público puede utilizar la información de una evaluación, en este caso siempre que no se implementen incentivos económicos, en el proceso de rendición de cuentas por parte de los centros. En todas las funciones, salvo en la última, en la que se requiere una evaluación censal, el tipo de evaluación requerida es de carácter censal o muestral.

El sector público puede dar una respuesta de alta intensidad ante los resultados de la evaluación. Puede, por ejemplo, utilizar la evaluación para “premiar” o “castigar” a centros educativos específicos mediante alteraciones en la asignación del presupuesto (véase Martínez Arias, 2009). Puede, también, “premiar” o “castigar” a profesores concretos, utilizando medidas de índole financiera y/o que afecten al destino, promoción, salario e incluso continuidad de las carreras laborales de los profesores. Esta función de la evaluación, tal y como se analiza en Baker et al. (2010), presenta graves problemas asociados a la falta de precisión e inestabilidad de los resultados obtenidos, incluso cuando se utilizan modelos de valor añadido.

Finalmente, el sector público puede utilizar la información individual sobre las competencias de los alumnos en procesos selectivos. Se trata de evaluaciones de tipo “reválida”, eliminatorias, como las que se establecen, al finalizar la ESO y al finalizar Bachillerato, en la LOMCE o como las que se realizan en Estados Unidos desde la implantación de la ley No Child Left Behind (2001). Conviene subrayar que en todas las evaluaciones de competencias cuyo destinatario es el sector público con respuesta de alta intensidad la información requerida es censal.

Un comentario aparte merece el papel de una serie de organismos internacionales, cuya actividad en la evaluación de competencias ha sido pionera. Nos referimos, por ejemplo, a la IEA (International Association for the Evaluation of Educational Achievement) que impulsa evaluaciones como TIMSS (Trends in International Mathematics and Science Study) y PIRLS (Progress in International Reading Literacy Study), a la OCDE (con PISA –Programme for International Student Assessment- y, más recientemente, la evaluación de adultos PIAAC, Programme for the International Assessment of Adult Competencies) o a la Comisión Europea (con su reciente evaluación de competencias de lengua extranjera ESLC (European Survey on Language Competences), de 2011. Estas instituciones utilizan la evaluación de competencias,

siempre en un contexto de “low-stakes” con la triple función de globalizar una “cultura” de evaluación, para monitorizar, en una perspectiva comparativa, la evolución de los sistemas educativos y para identificar y difundir buenas prácticas.

Un tercer agente, el centro educativo, puede ser también destinatario de los resultados de la evaluación de competencias. En este caso, lógicamente, el tipo de evaluación debe ser censal, con objeto de que la información pueda llegar a todos los centros. El centro educativo puede utilizar la información obtenida para reorientar recursos humanos o físicos y, también, para identificar buenas prácticas, todo ello en un contexto de “low-stakes”. Pero puede también basar en esa información una serie de incentivos económicos, destinados a la dirección del centro o al profesorado. En cualquiera de ambos casos, la llegada de información a las escuelas sobre los resultados, en términos de competencias, de los alumnos, supone un complemento muy valioso en los diseños de políticas educativas que dan énfasis a la autonomía de los centros.

Al destinar información a un cuarto agente, las familias, la evaluación de competencias puede ser simplemente informativa (“low-stakes”), proporcionando orientación a la familia sobre la situación y evolución del alumno. Sin embargo, cuando la información que recibe la familia se refiere a un centro en su totalidad, la evaluación se configura como un elemento esencial en el funcionamiento del sistema educativo como un mercado o un cuasimercado (véase Glennerster, 1991). Efectivamente, la información acerca del resultado de los centros es crucial para orientar la demanda de forma eficaz en un contexto de capacidad de elección. Otros elementos de política educativa, como la “desterritorialización” de los procesos de elección escolar (ampliación de los distritos hasta el distrito único, por ejemplo) sólo adquieren todo su sentido cuando están apoyados por la difusión de la información sobre los resultados de los centros.

Un último agente receptor de la información sobre las evaluaciones de competencias es la comunidad investigadora. El interés de ésta reside en la ampliación del conocimiento sobre el funcionamiento del sistema y de los elementos que lo constituyen y no tiene como función la alteración de los incentivos, por lo que muy frecuentemente resultan suficientes, en este caso, la aplicación de evaluaciones muestrales. La gran importancia de la disponibilidad de evaluaciones rigurosas, extensas e intensas, ha hecho que en los últimos años la comunidad investigadora en

este campo resulte muy dependiente de las iniciativas tomadas por otros agentes (muy especialmente, por las instituciones internacionales).

Cuadro 1. Posibles funciones de la evaluación educativa dependiendo del agente receptor de la información y de la intensidad de la respuesta del agente

Agentes	Low-stakes	High-stakes
Sector público (financiador)	Evaluaciones generales de sistema (C/M) Evaluaciones de programas, ex ante y ex post (C/M) Rendición de cuentas (sin incentivos económicos) (C) Identificación de buenas prácticas (C/M)	Asignación de presupuesto (C) Incentivos económicos a dirección, profesorado (C) Selección del alumnado (C)
Instituciones internacionales	Globalizar “cultura” de evaluación de competencias (C/M) Monitorizar la evolución de sistemas (C/M) Identificación de buenas prácticas (C/M)	
Centros educativos	Reorientación de recursos (C) Identificación de buenas prácticas (C)	Incentivos económicos a dirección, profesorado (C)
Familias	Información sobre evolución de las competencias (C)	Elección de centro (C)
Investigadores	Evaluación de sistema (C/M) Comparaciones internacionales e interregionales (C/M) Evaluaciones de programas, ex ante y ex post (C/M) Análisis de función de producción (C/M)	

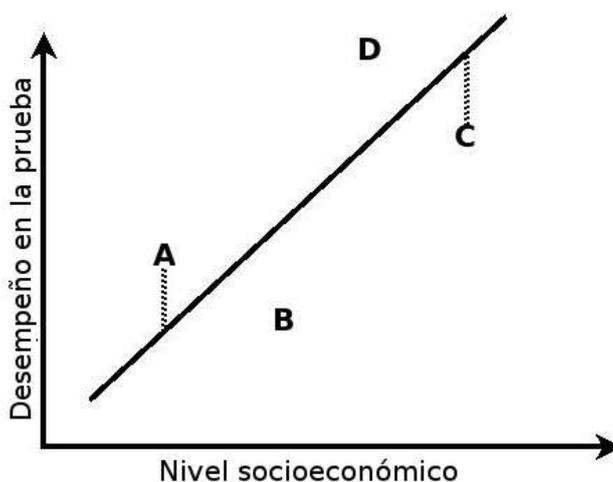
Nota. C: evaluación de carácter censal; M: evaluación de carácter muestral.

3. Efectividad de las escuelas: efecto-escuela y valor añadido

Las investigaciones sobre efectividad de las escuelas, a partir de la década de 1960 y 1970 en el Reino Unido y Estados Unidos se plantearon como objetivo permitir una comparación entre el desempeño de escuelas con características contextuales (usuarios, especialmente) muy diferentes. Los procedimientos estadísticos pretenden "eliminar" esas diferencias, intentando predecir, hasta cierto punto, lo que sería el rendimiento si todas las escuelas tuvieran las mismas condiciones de contexto.

El efecto de la escuela se estima por lo general a partir de una regresión multinivel. Técnicamente, es el residuo de cada escuela en un determinado modelo, es decir, la diferencia entre el rendimiento real y el rendimiento esperado de acuerdo con un modelo de regresión lineal (Thomas, 1998). Podemos identificar el efecto de la escuela como las líneas de puntos en la figura 1.

Figura 1. Efecto-escuela (línea de puntos) de cuatro escuelas hipotéticas (A-D), estimado a partir de un modelo con una variable explicativa (nivel socioeconómico).



La línea de regresión (diagonal) muestra el desempeño, en término de adquisición de competencias, que se espera de una escuela, de acuerdo con el nivel socioeconómico de sus alumnos, basado en el patrón identificado a partir de los datos de las cuatro escuelas. Podemos ver que la escuela C obtuvo un promedio superior a A. Sin embargo, teniendo en cuenta el contexto (en este caso, el nivel socioeconómico), la escuela A obtuvo mejores resultados de lo esperado, mientras que lo contrario ocurrió en C (líneas punteadas). Así, en términos de efecto de la escuela, dependiente en este caso exclusivamente del nivel socioeconómico, A sería “mejor” que C, en tanto que efectúa un mejor uso de sus recursos, conclusión opuesta a la obtenida cuando se comparan directamente los promedios. Esta metodología permite, por tanto, una comparación más justa de escuelas que operan en condiciones muy diferentes. Esto hace que sea posible identificar, por ejemplo, los centros que están haciendo un buen trabajo partiendo de condiciones desfavorables, como es el caso de la escuela A.

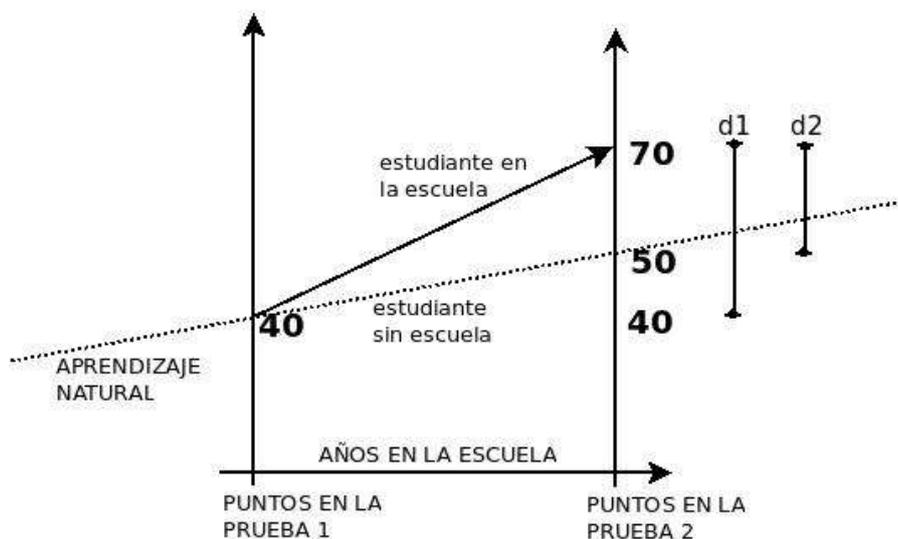
El efecto-escuela, como indicador de la eficacia de la escuela, presenta algunas propiedades estudiadas, como la estabilidad, perdurabilidad, consistencia, y la eficacia diferencial (Murillo Torrecilla, 2005). Inicialmente, los estudios en este campo se basaron únicamente en datos transversales, y produjeron resultados en términos de "rendimiento contextualizado".

Más recientemente, los estudios sobre la efectividad de las escuelas han puesto el acento en la evaluación longitudinal, a lo largo del tiempo, más que en la aproximación únicamente transversal (Martínez Arias et al., 2009). Este énfasis en los datos longitudinales ha dado lugar a una segunda etapa de los estudios de evaluación de la eficacia escolar, ahora agrupados bajo la etiqueta “de valor añadido”. La aplicación de este concepto, inicialmente económico, al contexto, pretende reformular la pregunta sobre lo que es una buena escuela. En este caso, una buena escuela sería la que añade valor a sus estudiantes, es decir, aquella en la que puede mejorar cualquier estudiantes con independencia de su nivel socioeconómico y cultural o de su nivel inicial de competencias.

El valor añadido pretende medir el efecto de las escuelas en el proceso de aprendizaje de los estudiantes, aislando factores que están más allá del control de los centros y que se encuentran distribuidos de forma no homogénea entre ellos. Una forma sencilla de calcular el valor que una escuela añade a los estudiantes sería, por ejemplo, realizar una sustracción entre los resultados de dos pruebas: una que el

estudiante realiza al ingresar en la escuela y otra que realiza cuando la abandona (d_1 en la Figura 2). Suponiendo que las pruebas son similares, este procedimiento parece tener sentido a primera vista; sin embargo, es importante tener en cuenta que la escuela no es el único factor que actúa sobre el estudiante durante el tiempo que transcurre entre las dos pruebas: no hay ninguna razón para suponer que el estudiante no aprendería nada si estuviera fuera de la escuela. En este sentido, el valor que la escuela ha añadido al estudiante, en relación a la ausencia completa de la escuela, podría ser representado por d_2 .

Figura 2. Versión simplificada del concepto de valor añadido aplicado a los efectos de la escuela sobre un estudiante.

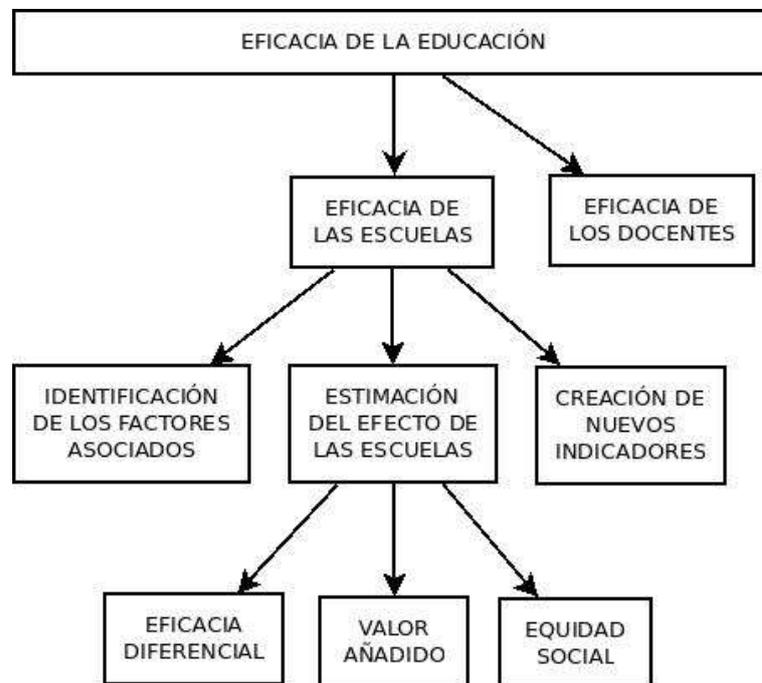


Los resultados de los modelos de valor añadido (VAM, en adelante) pueden describirse en función de algunas propiedades básicas tales como su mayor o menor nivel de estabilidad (a lo largo del tiempo) y su nivel de consistencia (entre alumnos). En cuanto a la primera, algunos estudios han encontrado una correlación cero entre los indicadores de valor añadido en cuatro años, lo que sugiere una alta inestabilidad (Linn y Haug, 2002). Por el contrario, en otras evaluaciones se ha identificado una estabilidad razonable y una gran coherencia entre las diferentes medidas de valor añadido (Ferrão y Couto, 2013), lo que sugiere la necesidad de realizar más estudios sobre este tema. También es interesante subrayar que, de acuerdo con Thomas et al. (1997), pocas escuelas tienen resultados estables y consistentes. Además, de acuerdo

con Luyten (1994), la estabilidad del efecto-escuela en el mismo estudiante es razonable, pero entre los estudiantes es muy pequeña.

Antes de enfocar nuestra atención en los VAM, aclararemos la relación entre el efecto-escuela y el valor añadido. Gran parte de la investigación sobre la efectividad de las escuelas, que en un principio trató del efecto-escuela, hoy se centran en el valor añadido. Esto no significa, sin embargo, que el efecto-escuela se haya dejado de estudiar y haya sido sustituido por los VAMs. Por el contrario, como describen Ferrão y Couto (2013), los estudios para estimar el efecto de la escuela pueden tener diferentes objetivos, tales como la eficacia diferencial (cuando diferentes tipos de alumnos responden de manera diferente a la acción de la escuela), la equidad social (cuando se trata de las diferencias socioeconómicas) y también el valor que la escuela ha añadido a cada estudiante. Por lo tanto, y como se observa en la figura 3, el concepto de valor añadido puede ser considerado una parte del concepto de efecto-escuela.

Figura 3. Estudios sobre eficacia escolar

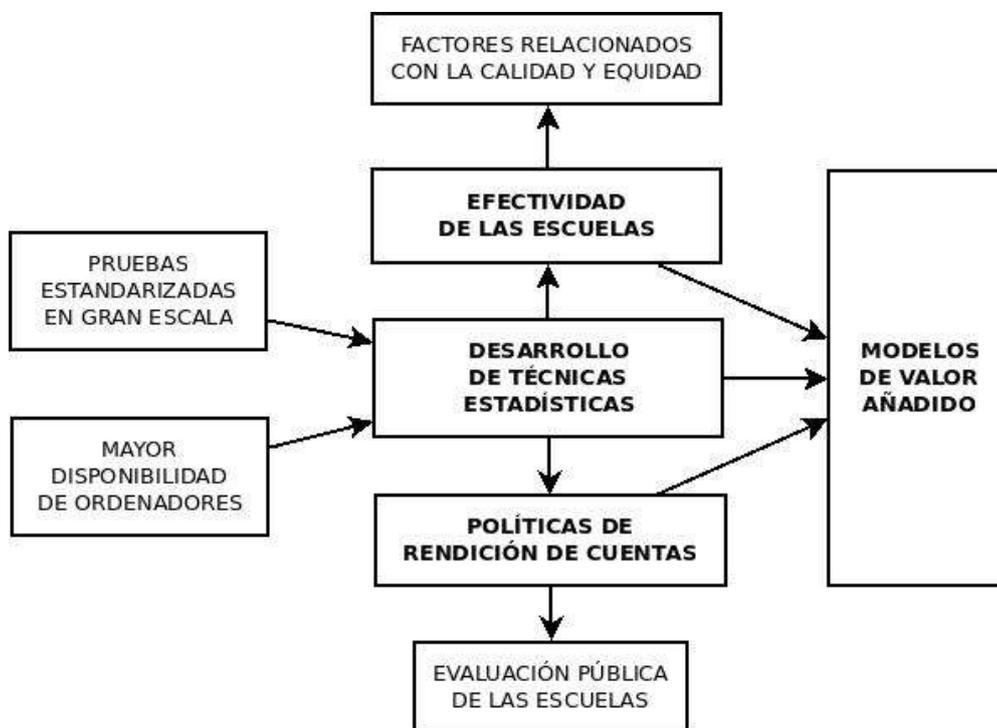


4. Los modelos de valor añadido

Aunque existen críticas a su uso en las políticas de rendición de cuentas y de establecimiento de incentivos, los modelos de valor añadido son ampliamente aceptados por los investigadores como un instrumento más apropiado que el simple resultado promedio de las evaluaciones estandarizadas de competencias (McCaffrey et al., 2004, Meyer, 1997; Tekwe et al., 2004; Thomas, 1998), en tanto que permiten suprimir el efecto de factores no controlados por la escuela, como los resultados previos o las características socioeconómicas y culturales de los usuarios

Los factores que han impulsado la proliferación de los VAMs son diversos. Entre ellos podríamos destacar la investigación sobre la efectividad de las escuelas, la implantación de políticas de rendición de cuentas, el desarrollo de técnicas estadísticas adecuadas y la disponibilidad de un número creciente de evaluaciones de competencias³. La figura 4 pretende sintetizar esta diversidad de factores.

Figura 4. Factores que impulsan la aplicación de los VAMs



³ Algunos de estos factores se discuten en Martínez Arias et al. (2009).

Observamos, pues, que los VAMs buscan solucionar problemas de gestión y economía - como la rendición de cuentas - y tratan, simultáneamente, de mejorar la calidad y la equidad de las escuelas. Se presentan como una de las mejores opciones disponibles para la construcción de indicadores de la efectividad escolar.

Es importante tener en cuenta, sin embargo, que la calidad de un VAM depende de que cumpla adecuadamente ciertos requisitos. De acuerdo con Ferrão y Couto (2013), los requisitos metodológicos mínimos para la construcción de un VAM correcto son los siguientes:

1. Que evalúe el nivel de competencias del estudiante al principio del proceso;
2. Que utilice datos longitudinales con al menos tres cohortes de estudiantes;
3. Que tenga en cuenta la estructura jerárquica de la población del estudio;
4. Que considere los factores contextuales de la escuela que inciden sobre los resultados de los estudiantes, tales como el nivel socioeconómico y cultural.

Se observa, por lo tanto, que las bases de datos utilizadas tienen que cumplir con varios requisitos previos para producir un buen modelo, lo que a menudo no se da en la práctica. Martínez Arias (2009) extiende los requisitos necesarios para un buen VAM a los siguientes; como puede observarse, los puntos 3, 4, 5 y 6 de las condiciones expuestas por Martínez Arias coinciden con las condiciones expuestas por Ferrão y Couto (2013).

1. Clara formulación de los objetivos;
2. Disponibilidad de muestras amplias y representativas de datos sobre los alumnos individuales, recogidos mediante tests estandarizados externos;
3. Recogida de datos de todos los niveles de rendimiento del alumnado;
4. Disponibilidad de al menos tres medidas del rendimiento (y un identificador único de cada estudiante que se mantiene);
5. Disponibilidad de información contextual de los alumnos y de la escuela;
6. Utilización de procedimientos de análisis estadístico multivariante de tipo multinivel;

7. Difusión de información adecuada sobre la interpretación de los datos a los equipos directivos de los centros, profesorado y otros agentes interesados;
8. Disponibilidad de un sistema de difusión de los resultados claro y transparente.

A continuación vamos a proponer un VAM simple que satisfaga los requisitos expuestos en los párrafos anteriores. La versión conceptual de este modelo simple consiste en que el resultado de un estudiante en un examen al final de la escuela (Y_2) se puede explicar en parte por el nivel de conocimientos previos, indicado por los resultados del examen inicial (Y_1), en parte por su situación socio-económica (SES), en parte por el trabajo de la escuela (ESC), en parte por su propio esfuerzo y el mérito (ESF).

$$Y_2 = Y_1 + SES + ESC + ESF \quad [1]$$

Para convertir esta idea en un modelo matemático de múltiples niveles, la forma más sencilla es la de transformar las dos variables de "mérito" en dos componentes de error, cada uno a su nivel. En un modelo multinivel, el esfuerzo del estudiante sería el componente de error en el nivel 1 (estudiante), mientras que el mérito de la escuela sería el error en el nivel 2 (escuela). Por otra parte, hay que añadir tres coeficientes (a_1 , a_2 , a_3) correspondientes a los efectos fijos. El modelo, por lo tanto, sería el siguiente:

$$Y_2 = a_1 + a_2*Y_1 + a_3*SES + ERR/escuela + ERR/alumno \quad [2]$$

Este sería un modelo sencillo que articula los datos longitudinales y contextuales jerárquicamente. La principal diferencia del modelo en relación con los primeros estudios sobre el efecto de la escuela es, precisamente, el elemento " a_2*Y_1 ", que se corresponde con el resultado previo del estudiante – por lo tanto, longitudinal - y entra como una variable de control adicional en el modelo.

La importancia del resultado previo como variable de control se ha confirmado sistemáticamente (Ferrão, 2009), lo que lleva a la búsqueda de la consolidación de bases de datos longitudinales sobre los estudiantes y las escuelas. Este tipo de datos presenta al menos dos ventajas en relación con el control realizado exclusivamente con las variables contextuales (por ejemplo, el nivel socio-económico): en primer lugar, permite superar el problema de la selectividad de las escuelas, es decir, en qué medida las escuelas seleccionan a los estudiantes a la entrada, y por lo tanto tienen menos responsabilidad sobre los buenos resultados al final del proceso; en segundo lugar

permite inferencias causales estadísticamente más consistentes, debido a que las medidas se establecen en diferentes momentos, y la temporalidad es uno de los supuestos básicos de las relaciones causales. Sin embargo, sólo la existencia de la relación temporal no es suficiente para que se pueda llevar a cabo una inferencia causal, como veremos a continuación.

Aunque existen diversas formas de definir las relaciones formales entre causa y consecuencia, un enfoque mínimamente seguro consiste en incluir el tiempo como una variable capaz de distinguir una de la otra, ya que la causa siempre debe producirse antes que el resultado. Por lo tanto, hay una cierta esperanza de que los VAM, utilizando datos longitudinales, tengan el potencial de detectar auténticos efectos causales.

Algunos autores consideran que la inferencia causal sólo puede alcanzarse experimentalmente. Pero existen razones para no llevar a cabo este tipo de procedimiento en el caso de los humanos, como el alto costo o las cuestiones éticas. Teniendo en cuenta estas limitaciones, junto con la gran cantidad de datos que se producen en las actividades humanas, como la educación, parece razonable utilizar las fuentes disponibles para el análisis. En estos casos, “parece más razonable estimar los efectos de los tratamientos a partir de estudios no aleatorizados que ignorar tales datos y soñar con el experimento ideal o tomar decisiones poco fundadas sin beneficiarse de los análisis de datos” (Rubin, 1974: 688). En otras palabras, el concepto que Rubin describe no viene a sustituir a la investigación experimental de las relaciones causales; por el contrario, se trata de generar información significativa en caso de que sea necesario, en los contextos en los que no es posible llevar a cabo experimentos. Así, Rubin describe su concepto de "causa" estadística del siguiente modo: “Intuitivamente, el efecto causal de un tratamiento E, sobre otro, C, para una determinada unidad y un intervalo de tiempo desde t_1 hasta t_2 es la diferencia entre lo que habría sucedido en el momento t_2 si la unidad hubiera sido expuesta a E desde el momento t_1 y lo que habría sucedido en el momento t_2 si la unidad hubiera sido expuesta a C desde el momento t_1 ” (Rubin, 1974: 689).

En esta versión expandida del método experimental, algunas variables siempre terminan siendo estimadas, en vez de medidas, ya que no se pueden obtener los resultados de la prueba de un estudiante de la escuela A y la prueba de este mismo estudiante hecha en la escuela B, o en ninguna escuela. Es decir, ciertos tipos de datos –

tales como los generados por las pruebas estandarizadas – no permiten una metodología estrictamente experimental: así, cuando falta cierta información, sólo podemos estimarla, de acuerdo con el patrón general que podemos identificar en el conjunto de datos.

Para que los VAM puedan ser utilizados para apoyar las políticas educativas (de rendición de cuentas o de otro tipo), es deseable que permitan inferencias causales. Con el aumento de la disponibilidad de los datos sobre la escuela y los estudiantes, se hace cada vez más posible realizar un buen análisis a partir de situaciones no experimentales. Con más datos sobre el contexto, es más fácil de aislar el trabajo realizado internamente en el centro de los determinantes relacionados con el contexto. Con más datos longitudinales, es más fácil separar las causas de los efectos. Así, los VAM pueden ser considerados como una técnica prometedora para identificar las causas de la efectividad (y falta de efectividad) de las escuelas.

Sin embargo, a pesar de que proporcionan una información más adecuada que el rendimiento promedio y que, a la hora de evaluar escuelas, las estimaciones realizadas por medio de VAMs presentan consistencia y una estabilidad razonable, estas presentan un punto débil y muy controvertido. De acuerdo con análisis recientes (véase Rothstein, 2007 y Baker et al., 2010) los VAM aún no están lo suficientemente maduros como para ser utilizados para proporcionar incentivos y desincentivos a los profesores, debido a la falta de estabilidad de sus resultados cuando se desciende al nivel del profesorado.

5. El debate sobre la difusión de los resultados de la evaluación de competencias

La diversidad de funciones que puede cumplir la evaluación de competencias educativas, que ha sido descrita en el apartado 2, conduce a que no pueda establecerse un criterio único a la hora de decidir en torno a la difusión de los resultados de la evaluación. Éste es un tema controvertido, en el que encontramos posiciones contrapuestas que oscilan entre, por una parte, la defensa de la publicidad completa de todos los resultados y, por otra, el apoyo a la prohibición de hacerlos públicos más allá

de los centros y los profesores implicados. Estas posiciones, como vamos a ver, se corresponden con la asignación a la evaluación de funciones alternativas.

La publicidad de los resultados (en el nivel del centro) de las evaluaciones censales está asociada con la función de proporcionar información a las familias en su proceso de elección de centro. La información que reciben las familias constituye un elemento clave en la configuración de los mercados y cuasimercados educativos y una política educativa que desee promoverlos deberá establecer las bases para que los procesos de elección estén basados en una buena información.⁴ Pero podemos preguntarnos qué constituye una “buena” información, una información adecuada para la elección; específicamente ¿necesitan las familias información sobre el nivel de competencias “en bruto” de cada centro o, por el contrario, medidas más sofisticadas, donde se identifique el efecto-escuela y se haya descontado, por tanto, el efecto del tipo de usuarios y los peer-effects que se establecen entre ellos? ¿Resultan de algún interés para las familias, en el proceso de elección de centro, los resultados de VAMs donde se hayan analizado longitudinalmente los efectos generados por el centro?

La información, en principio, más relevante para las familias es la que proporciona el nivel de competencias “en bruto”, ya que encapsula el efecto combinado del efecto-escuela, el contexto socioeconómico y cultural y los peer-effects, tres factores relevantes para identificar cómo puede afectar la escuela a la educación de los hijos e hijas. Sin embargo, no resulta descartable que haya familias que deseen incorporar en su proceso de decisión información más sofisticada, relativa al efecto-escuela y/o los resultados de VAMs. Este deseo puede estar relacionado con el nivel socioeconómico y cultural de la propia familia (pueden existir familias, por ejemplo, que confíen en el nivel sociocultural de la propia familia y consideren que puede compensar un nivel sociocultural bajo en la escuela siempre que ésta sea suficientemente eficaz). Por otra parte, la administración educativa puede decidir proporcionar información a las familias sobre el efecto-escuela o el resultado de VAM's en una práctica de transparencia en la prestación de los servicios públicos y no únicamente con intención de favorecer la elección de centro. En el Reino Unido, por ejemplo, la administración educativa proporciona a los ciudadanos valores que permiten comparar a las escuelas con respecto a las que tienen un tipo de usuario similar; se trata de una práctica en la

⁴ La capacidad efectiva de elección dependerá también, como es lógico, de la disponibilidad de oferta de centros educativos suficientemente diversos.

que confluye el doble papel del ciudadano como demandante de un “producto” y de contribuyente que tiene derecho a conocer el proceso de producción de un servicio financiado públicamente.

Otras funciones de la evaluación de competencias no implican en absoluto la difusión pública de información sobre los resultados de los centros. Si se ha establecido como receptor de la información al sector público o a los centros educativos, tanto en el caso de evaluaciones “low-stake” como “high-stake” (véase apartado 2), resulta coherente la imposición de una limitación o prohibición de la difusión pública de los resultados de los centros. Tal imposición se deriva del temor a que la difusión refuerce los procesos de segregación escolar. Una información inicialmente destinada, por ejemplo, a que los centros rindan cuentas ante la administración financiadora, podría ser utilizada por los centros con mejores puntuaciones para atraer demanda “selecta”, mientras que los centros con peores puntuaciones asistirían a la “huída” de aquellos pocos alumnos con mayores recursos socioeconómicos y culturales y, simultáneamente, con mayor capacidad e incentivos para la elección. Este argumento es el que guió la limitación de la difusión de los resultados de las evaluaciones establecidas en la Ley Orgánica de Educación de 2006⁵ y en la Ley de Educación de Cataluña de 2009⁶.

De este modo, el debate acerca de la difusión pública de los resultados de las evaluaciones de competencias es indisoluble de la función que pretenda, y en la que esté enmarcada, la evaluación. Los diseñadores de las evaluaciones deben ser conscientes de la falta de neutralidad de los procesos de difusión: una difusión pública de resultados “brutos”, por ejemplo, está vinculada única y exclusivamente a los procesos de elección de centro por parte de las familias y puede tener consecuencias inmediatas en términos de segregación y, por tanto, de equidad.

⁵ El redactado del artículo 140.2 de la LOU es el siguiente: “La finalidad establecida en el apartado anterior no podrá amparar que los resultados de las evaluaciones del sistema educativo, independientemente del ámbito territorial estatal o autonómico en el que se apliquen, puedan ser utilizados para valoraciones individuales de los alumnos o para establecer clasificaciones de los centros”.

⁶ El redactado del artículo 184, c) de la Ley de Educación de Cataluña establece que la evaluación educativa tiene que estar sometida al principio de “uso reservado de la información individualizada de los agentes y de los centros y servicios educativos respecto a la evaluación general del sistema”.

6. Conclusiones

La evaluación de competencias educativas puede estar orientada a cubrir las necesidades de información de una diversidad de agentes, que responden ante tal información con baja o alta intensidad (low-stakes vs. high-stakes). En el artículo hemos sistematizado tales funciones utilizando las necesidades de cinco tipos de agentes: administración pública, instituciones internacionales, centros educativos, familias y comunidad investigadora.

Posteriormente hemos concentrado la atención en los modelos de valor añadido; estos modelos suponen un avance indiscutible con respecto a la utilización simple de los resultados promedio de las evaluaciones. Esto es así pese a las dificultades técnicas y de disponibilidad de datos que entrañan, dificultades que impiden que puedan ser considerados como un instrumento adecuado a la hora de establecer incentivos y desincentivos en el nivel del profesorado. Después de discutir algunos de los requisitos que debe cumplir un modelo de valor añadido hemos propuesto un modelo sencillo que articula jerárquicamente datos longitudinales y datos contextuales (relativos al origen socioeconómico y cultural de los usuarios del centro).

En el artículo hemos abordado, también, la cuestión de la difusión de los resultados de las evaluaciones de competencias. Existe un debate en torno a esta cuestión, debate que depende, esencialmente, de la función que se desee asignar a la evaluación. Las posiciones más favorables a la difusión de los resultados asumen la función de facilitar la elección de centro en un contexto de mercado o cuasi-mercado. Los costes en términos de equidad de esta opción residen en un potencial incremento de la segregación escolar. Por contra, en el otro extremo, las posiciones que defienden una regulación que impida la difusión de los resultados más allá de los profesores y centros implicados están asumiendo una función para la evaluación centrada en la rendición de cuentas con respecto a la administración pública.

7. Referencias

Baker et al. (2010). "Problems with the use of student test scores to evaluate teachers". Economic Policy Institute Briefing Paper #278.

Calero, Jorge y Álvaro Choi (2012). "La evaluación como instrumento de política educativa" *Presupuesto y gasto público*, n. 67, pp. 29-42.

Elliot, John (2002). "The paradox of educational reform in the evaluatory state: implications for teacher education". *Prospects*, vol. XXXII, n. 3.

Ferrão, M. E. (2009). Sensibilidad de las especificaciones de los modelos de valor añadido: midiendo el estatus socioeconómico. *Revista de educación*, 348, pp. 137-152.

Ferrão, M. E. (2013). School effectiveness research findings in the Portuguese speaking countries: Brazil and Portugal. *Educational Research for Policy and Practice*. DOI:10.1007/s10671-013-9151-7

Ferrão, M. E. y Couto, A. (2013). Indicador de valor acrescentado, tópicos sobre consistência e estabilidade: uma aplicação ao Brasil. *Ensaio: Avaliação e Políticas Públicas em Educação*, pp.131-164.

Gairín, J. (2009). "Usos y abusos de la evaluación. La evaluación como autorregulación", en Gairín, J. (ed.) *Nuevas funciones de la evaluación*. Madrid, Ministerio de Educación.

Glennster, H. (1991). "Quasi-Markets for Education?" *The Economic Journal*, vol. 101, n. 408, pp. 1268-1276 .

Linn, R. L. y Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis*, 1522(310).

Luyten, H. (1994). Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. *International Journal of Educational Research*, 21(2), pp. 197-216.

Martínez Arias, R. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista de Educación*, 348, pp. 217-250.

Martínez Arias, R., Gaviria Soto, J. L. y Castro Morera, M. (2009). Concepto y evolución de los modelos de valor añadido en educación. *Revista de Educación*, 348, pp. 15–45.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A. y Hamilton, L. (2004). Models for Value-Added Modeling of Teacher Effects. *Journal of educational and behavioral statistics*, 29(1), pp. 67–101. doi:10.3102/10769986029001067

Meyer, R. H. (1997). Value-Added Indicators of School Performance: A Primer. *Education*, 16(3), pp. 283–301.

Murillo Torrecilla, F. J. (2005). La investigación sobre eficacia escolar. Eficacia escolar y factores asociados. Barcelona: Octaedro.

Ravitch, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Books.

Ray, Andrew; Helen Evans y Tanya McCormack (2009). “School Value Added Measures in England”. *Revista de Educación*, n. 348, pp. 47-66.

Rothstein, J. (2007). Do Value-added Models Add Value? Tracking, Fixed Effects, and Causal Inference. Princeton University, Department of Economics, Center for Economic Policy Studies. Working Paper No 1036.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5), 688–701.

Rubin, D. B., Stuart, E. A. y Zanutto, E. L. (2004). A Potential Outcomes View of Value-Added Assessment in Education. *Journal of educational and behavioral statistics*, 29(1), 103–116.

Schwartzman, S. (2005). As avaliações de nova geração nas sociedades contemporâneas. In A. de Melo e Souza (Ed.), SOUZA, Alberto de Mello e Souza (Org.). *Dimensões da avaliação educacional*. (pp. 15–34). Petrópolis: Vozes.

Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M. E., Roth, J., Resnick, M. B. (2004). An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance. *Journal of Educational and Behavioral Statistics*, 29(1), pp. 11–36. doi:10.3102/10769986029001011

Thomas, S. (1998). Value-added measures of school effectiveness in the United Kingdom. *Prospects*, XXVIII(1).

Thomas, S., Sammons, P., Mortimore, P. y Smees, R. (1997). Stability and Consistency in Secondary Schools' Effects on Students' GCSE Outcomes over Three Years. *School Effectiveness and School Improvement*, 8(2), pp. 169–197.

Toch, Thomas (2006). Margins of Error: The Education Testing Industry in the No Child Left Behind Era. *Education Sector Reports*.

Travitzki, R. (2013). ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar. Tese (doutorado). São Paulo: s.n.