

La evaluación educativa desde la perspectiva del valor añadido

La evaluación de los sistemas educativos constituye una herramienta fundamental para conocer el progreso en educación. La mayoría de los sistemas educativos informan, más o menos sistemáticamente, del estado actual del nivel de académico de los alumnos. Sin embargo, la comparación de este estado a través del tiempo, la valoración del incremento de los niveles de aprendizaje de nuestros alumnos suele ser un elemento ausente. En los últimos años aparecen propuestas legislativas como la inglesa y la norteamericana *No Child Left Behind* que proponen e impulsan sistemas de evaluación y rendición de cuentas de los sistemas educativos que se centran en el crecimiento de los logros de los aprendizajes. Este artículo describe los rasgos principales de los modelos de valor añadido en educación que permitan valorar simultáneamente el nivel actual de aprendizaje y la tasa de cambio o crecimiento que se ha producido por efecto de la acción de la escuela.

Palabras clave: modelos de valor añadido, modelos de crecimiento, evaluación de sistemas educativos, rendición de cuentas.

Educating People. Radical Educational Difficulties and a Proposal for Solving them

Evaluation of educational systems becomes a key element to know progress in education. Educational systems inform, more or less systematically, about current academic achievement. However, evaluating learning growth through time does not usually take place. Lately, some proposals have appeared such as the British and North American –No Child Left Behind– that help to develop evaluation and accountability systems which focus on learning progress and growth. This article describes the main features of value added models in education, which permit

Nb016

María Castro
Morera

Profesora Titular
Departamento de Métodos
de Investigación y Diagnós-
tico en Educación.
Universidad Complutense de
Madrid
maria.castro@edu.ucm.es

José Luis
Gaviria Soto

Catedrático
Departamento de Métodos
de Investigación y
Diagnóstico en Educación.
Universidad Complutense de
Madrid
josecho@ucm.edu.es

simultaneous evaluation of the current learning situation and the learning rate of growth due to school action.

Keywords: value added models, growth models, evaluation of educational systems, accountability.

1. INTRODUCCIÓN: LA EVALUACIÓN DE SISTEMAS BASADA EN EL VALOR AÑADIDO

Es muy difícil exagerar la importancia que la estrategia evaluativa basada en el valor añadido va a adquirir en los próximos años en el funcionamiento ordinario de la educación, aunque hoy todavía se trate fundamentalmente de un tema, importante desde luego, pero todavía de investigación. Y la razón para la progresiva relevancia de este enfoque reside en la importancia que la evaluación está ya adquiriendo como elemento que garantiza el dinamismo de los sistemas educativos.

La evaluación educativa comenzó a adquirir una gran importancia cuando las sociedades desarrolladas alcanzaron un cierto nivel de inversiones en el sistema educativo. Este crecimiento de los recursos dedicados a la educación respondía a un doble impulso: el producido por la democratización creciente de las sociedades y aquél cuyo origen estaba en la creencia en que el desarrollo económico está ligado de un modo u otro a la acumulación de lo que se dio en llamar como Capital Humano. Pero hay razones más radicales por las que la evaluación es importante. Y estas razones se relacionan con la naturaleza “entálpica” de la educación respecto de la sociedad y de la evaluación respecto de la educación.

En efecto, la educación es la institución social que permite la transmisión de conocimientos y valores de una generación a la siguiente, y por tanto, la que permite que todas las demás instituciones sociales funcionen. Sólo es posible la pervivencia de una sociedad abierta si una mayoría cualificada de sus ciudadanos se adhieren voluntariamente al sistema de reglas y convenciones que la hacen posible. Y aquí entra en juego la educación. Desde un punto de vista metaindividual, la pervivencia de la sociedad abierta sólo es posible si cada generación es capaz de transmitir a la siguiente la adhesión al conjunto de instituciones, convenciones y reglas abstractas que la definen. Y la educación es precisamente el mecanismo por el que esa adhesión transgeneracional ha de garantizarse. Por este motivo podemos decir que la educación es la institución que introduce orden, es decir, que reduce la entropía en el sistema social.

NOTASLA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

Por su parte la evaluación es a su vez el sistema de retroalimentación que permite que el sistema educativo evolucione en concordancia con las necesidades sociales cambiantes. En todos los casos la existencia de la evaluación como sistema de retroalimentación es la condición ineludible para la existencia de la conducta adaptativa en el propio sistema educativo. Por eso podemos decir que la educación introduce orden en la sociedad y la evaluación introduce orden en la educación.

Que la evaluación pueda cumplir ese papel depende, a su vez, de que la información que procesa sea la más relevante para la modificación del proceso educativo.

La evaluación basada en el valor añadido ha despertado enorme interés tanto entre las autoridades educativas responsables de la marcha de los sistemas en los distintos países como entre un amplio número de profesores sensibilizados con las consecuencias que la evaluación puede tener para su vida profesional. Para los primeros, la evaluación basada en el valor añadido trae la promesa de la posibilidad de tomar decisiones más acordes con la naturaleza del contexto en el que se produce la educación en una sociedad en la que, como a menudo se ha dicho, la única constante es el cambio. Para los segundos, la evaluación basada en el valor añadido permite pensar en una evaluación basada en resultados que tiene en cuenta las condiciones en las que estos se obtienen y, por tanto, en una evaluación más justa.

Pero lo que es realmente novedoso en este enfoque es el cambio de perspectiva que la propia metodología de la evaluación impone. En efecto, en las aproximaciones anteriores siempre primaba un cierto punto de vista sobre los demás. La evaluación podía entenderse como un proceso de control administrativo altamente tecnificado, donde el punto de vista predominante era el del administrador y los fines del sistema se planteaban en términos de variables de nivel macro. O podía entenderse como el proceso de mejora local, en el que lo importante era la perspectiva de satisfacción inmediata de los agentes locales implicados, perdiéndose la perspectiva de acción global del sistema.

En el enfoque del valor añadido la propia metodología obliga a tener en cuenta las dos perspectivas, el nivel macro y el nivel micro. Cualquiera de las aproximaciones al valor añadido se basa en las medidas de ganancia educativa, es decir, la diferencia entre los valores de entrada y los valores de salida. Pero para juzgar estos valores se utiliza la información que procede del conjunto de las

unidades evaluadas. Así el mérito individual de una unidad se juzga en comparación con otras unidades de características similares.

Una consecuencia de esta aproximación es que es posible detectar aquellas escuelas que, trabajando con *inputs* similares a otras, obtienen mejores resultados que las demás. Ya no se trata de analizar aquellas escuelas que obtienen en términos absolutos mejores resultados, sino de encontrar aquellas que con alumnos de ciertas características que les sitúa en una categoría de alumnos desfavorecidos, consiguen incrementos de aprendizaje comparables o superiores a otras con alumnos que presentan mejores condiciones de entrada.

Como veremos, esto es posible por la propia definición de valor añadido. Si como tal entendemos el residuo del incremento de aprendizaje una vez que los factores controlados han sido parcializados, es evidente que aquellas escuelas que presentan un valor positivo de ese residuo están rindiendo por encima de las expectativas, dadas sus características de entrada. En este nuevo paradigma lo importante ya no es sólo tener rendimientos altos en términos absolutos, sino presentar un residuo positivo.

Esta misma definición del valor añadido obliga y facilita tener además de la perspectiva macro, una perspectiva micro. En efecto, parcializados todos los factores conocidos, lo que produce un residuo positivo en una determinada escuela es algo idiosincrásico de la misma, que obliga a hacer una aproximación individual tratando de descubrir lo que son las “buenas prácticas” que permiten a esta escuela en particular destacar respecto de aquellas con las que puede compararse.

De este modo los resultados de la evaluación basada en el Valor Añadido se convierten en una importantísima fuente de información para diseñar y desarrollar programas de mejora en las escuelas en las que algunas de estas, que en principio parecerían tener resultados mediocres, son, por el contrario, modelos a seguir dada la ganancia en los aprendizajes que promueven en sus alumnos.

Todo esto es posible siempre que los factores conocidos se hayan parcializado. Es decir, hay que integrar necesariamente la perspectiva de la investigación. Se trata de identificar aquellos factores que afectan de modo general al rendimiento. Las escuelas con residuos positivos son precisamente aquellas que dados ciertos valores en estos factores producen rendimientos por encima de lo esperado.

Esta perspectiva de evaluación tiene, como vemos, una proximidad enorme con la investigación educativa. Podemos decir

NOTASLA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

que de algún modo la investigación se asume como punto de partida. Efectivamente, una definición de valor añadido que se basa en el cálculo de residuos asume que existen unos factores conocidos que explican parte de la varianza del rendimiento. La identificación de esos factores es precisamente el objetivo de la investigación educativa.

Pero es curioso cómo la evaluación basada en Valor Añadido interactúa con la investigación. Si una vez calculados los residuos de cada escuela identificamos aquellas con residuos positivos y tratamos de analizar cuáles son los elementos idiosincrásicos que actúan positivamente, una vez identificados, éstos se convierten a su vez en materia de hipótesis generales para contrastar.

En el límite, el objetivo es conseguir reducir el volumen de los residuos, es decir, tratar de trasladar las buenas prácticas desde las buenas escuelas a sus pares. Es evidente por tanto que este tipo de evaluación no sólo consigue establecer pautas de mejora para gran número de escuelas, sino que al hacer esto está también incrementando el conocimiento pedagógico. Hay pues una síntesis en este paradigma entre la investigación y la evaluación. Al mejorar los sistemas individuales, conseguimos aumentar nuestro conocimiento acerca del funcionamiento de las estructuras generales.

Este artículo constituye una primera aproximación a la evaluación educativa desde la perspectiva de los modelos de valor añadido y de crecimiento. Para ello, se describe en primer lugar el concepto de valor añadido y su operacionalización a través del modelo más básico, identificando los dos principales indicadores de progreso de un centro: el valor añadido y la tasa de crecimiento. A continuación se esbozan algunas sugerencias fundamentales para el diseño de evaluaciones basadas en el valor añadido. Y por último se realiza una breve síntesis de algunos de los sistemas de valor añadido implantados en la actualidad.

2. MODELO DE VALOR AÑADIDO Y LA FUNCIÓN DE CRECIMIENTO

El término “valor añadido” es llamativo y fuente de polémica dentro de la discusión sobre evaluación de sistemas educativos. Sus orígenes se enraízan con las nociones de bienes, riqueza y prosperidad en economía política (Saunders, 1999). Una referencia temprana del uso de este término en conexión con evaluaciones de intervenciones en un proceso de crecimiento educativo se puede encontrar en Bryk y Weisberg (1976). Su uso moderno transmite la idea de “progreso relativo”, tal y como señala Fitz-Gibbon (1997). Una interpretación más estadística reconoce que, independientemente de la forma que

adopte, los análisis de valor añadido se centran en modelos para “comparaciones ajustadas” (Goldstein y Spiegelhalter, 1996).

Conceptualmente, el valor añadido busca aislar el cambio en los niveles de aprendizaje del alumno, tanto de los factores relacionados con la historia económica y social del alumno, como de la composición de su escuela y su comunidad (Thum, 2003c).

Ahora bien, ¿qué se entiende operativamente por una medida de valor añadido?

De forma genérica, se puede decir que son expresiones empíricas de lo que se considera ganancia o crecimiento. Y son muchas las definiciones operativas a través de las cuales se puede definir una medida de ganancia (Thum, 2003b). Quizá la más sencilla sea el concepto de *ganancia bruta o absoluta*, que es una medida directa de la distancia entre el nivel actual de conocimientos y destrezas ($t=1$, Y_{i1}) y el nivel inicial ($t=0$, Y_{i0}), es decir ($Y_{i1} - Y_{i0}$).

Sin embargo, las medidas absolutas del crecimiento fallan en la más simple consideración de los distintos factores individuales y contextuales, haciendo de esta medida una apreciación “injusta”. Esta operacionalización se puede refinar a través del ajuste de covariables en el nivel actual o post-test, considerando que la principal covariable es el nivel inicial o rendimiento previo. La distancia queda entonces definida de esta forma: $Y_{i1} - E(Y_{i1}|Y_{i0})$, y se denomina *ganancia residual*. Esta medida de ganancia residual parece aportar una “medida relativa” de crecimiento.

Ahora bien, no puede considerarse como una medida adecuada de ganancia o crecimiento, ya que falla en informarnos cuánto ha cambiado un individuo en un atributo, además de violar el supuesto básico de la regresión de que las covariables están libres de errores, puesto que tanto el nivel inicial o pre-test como el post-test están medidos con error. Una alternativa más ajustada consiste en considerar tanto el nivel inicial como el actual como medidas de resultados, situándonos en el marco de las medidas repetidas. La *ganancia estimada* queda expresada de la siguiente manera: $E(Y_{i1}) - E(Y_{i0})$, que supone una mejora notable en la concepción del crecimiento, puesto que no se ven afectados artefactos del diseño y se considera la redundancia en las puntuaciones dentro del marco de las medidas repetidas. El valor añadido, entonces, hace referencia a las condiciones de equidad de este progreso relativo, al considerar que los determinantes que operan en el post-test también lo hacen en el pre-test, por tanto el efecto de las covariables es considerado en ambas

situaciones, quedando la medida definida de la siguiente manera: $E(Y_{i1}|X_{i1}) - E(Y_{i0}|X_{i0})$.

Los modelos jerárquicos lineales aportan el complemento metodológico para un tratamiento adecuado del crecimiento o del cambio en diseños de muestras estratificadas que son los más comunes en investigación educativa (medidas repetidas de estudiantes en el tiempo, estudiantes anidados en escuelas, distritos, etc.). Sólo los modelos lineales mixtos pueden responder simultáneamente a las cuestiones fundamentales en la evaluación del aprendizaje de los alumnos: cuál es su nivel actual de conocimientos, cuál es su tasa de crecimiento y cuál es la aportación específica de la escuela o valor añadido.

El desarrollo de los modelos jerárquicos lineales ha creado un potente conjunto de técnicas para la investigación del cambio individual dentro de organizaciones como las escolares (Raudenbush, 2001; Raudenbush y Bryk, 2002). Cuando se aplican junto a medidas válidas para diseños con múltiples puntos temporales, estos modelos ofrecen una herramienta adecuada para el estudio de la estructura y la predicción del cambio individual. Una ventaja de esta aportación es que es flexible con respecto a la estructura temporal de las observaciones. Igualmente importante es que los modelos multinivel facilitan el estudio de los efectos de las covariables en cada nivel de anidamiento (Bryk y Raudenbush, 1992; Gaviria y Castro, 2005; Goldstein, 1995; Laird y Ware, 1982; Snijders y Bosker, 1999).

2.1. Modelo básico de valor añadido

El modelo de valor añadido más sencillo refleja la existencia de una estructura jerárquica de, al menos, tres niveles: el tiempo (t), el alumno (i) y la escuela (j). Las puntuaciones recogidas en distintos momentos de medición se anidan en los alumnos que las producen (mientras aprenden y evolucionan), que a su vez están agrupados en centros escolares. Así, y_{tij} es el rendimiento en el momento t de un alumno i que asiste a la escuela j . El modelo nulo queda planteado como sigue (Ecuación 1),

$$y_{tij} = \beta_{0ij} + \beta_{1ij}(t - t_0) + \varepsilon_{tij} \quad (1)$$

donde $\varepsilon_{tij} \sim N(0, \sigma_{\varepsilon}^2)$

En este primer nivel, el término β_{0ij} representa el punto de corte en el momento inicial de la medición. Si, como es el caso, la variable

NOTAS

LA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

de respuesta es el rendimiento, β_{0ij} representa el rendimiento medio esperado del alumno i en la escuela j en el momento t_0 , es por tanto, el *nivel inicial de conocimientos* para ese alumno en esa escuela.

Dadas las condiciones de nuestro interés –el valor añadido– este modelo nulo precisa de la inclusión de un término adicional, β_{1ij} vinculado al predictor tiempo. Así, β_{1ij} representa la *tasa de crecimiento esperado en los aprendizajes* del alumno i en la escuela j debida al paso del tiempo y $(t-t_0)$ representa el paso del tiempo entre mediciones. El término ε_{ij} representa el error aleatorio en este primer nivel que muestra las variaciones aleatorias en las puntuaciones dentro de cada ocasión de medición, y muestra una distribución normal con media 0 y varianza constante como supuesto base.

Como se puede observar, este modelo nulo incluye un término referido al nivel de partida en los aprendizajes, β_{0ij} , y otro término referido a la tasa de crecimiento de los mismos, β_{1ij} . Según este modelo, los alumnos y las escuelas pueden variar en sus niveles iniciales así como en la distinta cantidad de aprendizajes logrados a lo largo del tiempo, reconociendo variabilidad en los niveles segundo y tercero. Si bien es cierto que se asume que la relación que hay entre el tiempo y el aprendizaje es lineal (y suponemos que positiva), esto supone afirmar que con el paso del tiempo los aprendizajes *siempre* irán creciendo.

El reconocimiento de la variabilidad entre alumnos y escuelas permite desarrollar los subsiguientes submodelos para cada uno de estos niveles. Así, tanto el nivel de alumnos como el nivel de escuelas tienen su propia representación para el nivel inicial de conocimientos y la tasa de crecimiento.

En el segundo nivel, la Ecuación (2) muestra la descomposición del estatus inicial del alumno, identificándose β_{0j} que representa la media en rendimiento de la escuela j en el momento de la primera medición, más un componente aleatorio μ_{1ij} que muestra el residuo específico de cada alumno con respecto a este valor central de su escuela. En cierto sentido, podríamos llegar a denominar a este componente el “valor añadido del alumno” pues representa la especificidad del nivel inicial de conocimientos de cada estudiante con respecto al conjunto de su escuela.

$$\beta_{0ij} = \beta_{0j} + \mu_{0ij} \quad (2)$$

$$\beta_{1ij} = \beta_{1j} + \mu_{1ij} \quad (3)$$

NOTAS

LA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

Del mismo modo, en la ecuación (3) incluye el término β_{1j} que muestra la tasa de crecimiento medio de la escuela j , más un término residual, μ_{1ij} , que es lo que se aleja la pendiente de crecimiento de un alumno específico con respecto a la tasa de crecimiento general de su escuela.

La parte aleatoria de este submodelo, compuesta por μ_{0ij} y μ_{1ij} , muestra una distribución normal bivariada, con media 0 y varianzas $\sigma_{\mu 0}^2$ y $\sigma_{\mu 1}^2$ respectivamente y covarianza $\sigma_{\mu 0\mu 1}$.

Y de forma análoga, en el tercer nivel (escuelas) podemos identificar en la ecuación (4) cómo el rendimiento inicial medio de una escuela tiene por ahora dos componentes, el nivel medio inicial del conjunto de la población de escuelas (β_{00}) y la diferencia específica de esa escuela en el momento inicial (μ_{0j}) que es el indicador del *Valor Añadido de la Escuela*. Igualmente, la ecuación (5) describe cómo la tasa de crecimiento de una escuela tiene un componente que muestra el nivel de crecimiento medio para el conjunto de escuelas (β_{10}) y la diferencia específica de esa escuela en la tasa de crecimiento (μ_{1j}) que es la *Tasa de Crecimiento* específica de ese centro educativo.

$$\beta_{0j} = \beta_{00} + \mu_{0j} \quad (4)$$

$$\beta_{1j} = \beta_{10} + \mu_{1j} \quad (5)$$

La presentación del modelo en tres niveles con submodelos específicos responde a criterios de claridad expositiva, pues el modelo es único, substituyendo en la Ecuación (1) las Ecuaciones (2), (3), (4) y (5), queda el modelo general como sigue:

$$y_{tij} = \beta_{00} + \beta_{10} (t - t_0) + \mu_{0ij} + \mu_{1ij} (t - t_0) + \mu_{1j} (t - t_0) + \varepsilon_{tij} \quad (6)$$

Los términos conceptualmente más relevantes del tercer nivel son los residuos del centro, vinculados al punto de corte y a la pendiente, pues cuantifican el valor añadido y la tasa de crecimiento de centro, objeto de interés de este artículo. Los estudios sobre crecimiento y valor añadido se suelen centrar en el ajuste a través de la inclusión de distintos predictores vinculados a las características del alumno, del centro y de la intervención escolar; pues el valor añadido y la tasa de crecimiento se definen como los residuos del modelo.

2.2. Forma de la función de crecimiento

La función de crecimiento más habitualmente utilizada en los modelos de valor añadido es la lineal. Sin embargo, es posible que con el paso del tiempo, una vez logrado un máximo de aprendizaje, se produzca un descenso en los niveles medios del mismo (las hipótesis para este comportamiento son variadas, bien por características propias del momento psico-biológico de los estudiantes, bien por las limitaciones de las pruebas de rendimiento que alcanzan a identificar un techo de rendimiento pero no tienen mayor sensibilidad). Y también es posible que, tras un tiempo que podríamos calificar “de olvido”, se vuelva a reiniciar un ascenso en nuestros niveles de conocimiento. Las hipótesis de tasas de crecimiento cuadráticas y cúbicas, respectivamente, son también planteables siempre que el número de ocasiones de medición lo permitan.

Si se asumiera una hipótesis cuadrática del crecimiento, que mostraría que el rendimiento de los centros y de los estudiantes alcanza un máximo para producirse un descenso posterior, el modelo nulo planteado en la ecuación (1) quedaría como sigue,

$$y_{ij} = \beta_{0ij} + \beta_{1ij} (t - t_0) + \beta_{2ij} (t - t_0)^2 + \varepsilon_{ij} \quad (7)$$

Si, por otro lado, se quisiera contrastar una hipótesis de crecimiento cúbica, que describiría un máximo, un mínimo y un nuevo ascenso del rendimiento, el modelo nulo a contrastar se especifica de la siguiente manera:

$$y_{ij} = \beta_{0ij} + \beta_{1ij} (t - t_0) + \beta_{2ij} (t - t_0)^2 + \beta_{3ij} (t - t_0)^3 + \varepsilon_{ij} \quad (8)$$

Como se ve en las Ecuaciones (7) y (8), la aditividad de los modelos de valor añadido permite incorporar y comprobar la significación de los términos cuadrático y cúbico independientemente.

3. CARACTERÍSTICAS DEL DISEÑO DE INVESTIGACIÓN

A la idea del valor añadido en educación le es consustancial una concepción dinámica de los logros de aprendizaje. El cambio y la evolución a lo largo del tiempo definen la esencia de los modelos de valor añadido que son sensibles a la verdadera noción de aprendizaje que implica crecimiento y cambio.

NOTAS

LA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

Desde este planteamiento, la inclusión del tiempo dentro de estos modelos es su marca de clase. Tanto es así que los modelos de valor añadido no se pueden entender fuera de un diseño longitudinal y sin un referente temporal dentro de la definición del modelo más básico (Ruiz de Miguel y Castro, 2006). Como señalan Zvoch y Stevens (2006) y antes Nesselrode (1991), el cambio precisa ser estudiado en la intersección de variaciones entre sujetos, variables y ocasiones de medida.

Los diseños longitudinales, al permitir evaluar la trayectoria del crecimiento de los alumnos durante un periodo de tiempo, proporcionan numerosas ventajas sobre otro tipo de diseños, hasta el punto que se han llegado a considerar los más adecuados para evaluar el progreso de los alumnos y la eficacia de las escuelas (Stevens, 2005; Thum, 2003c). Entre estas ventajas cabe destacar su adecuación a situaciones en las que queremos controlar la influencia de las características socio-familiares de los alumnos (al ser éste su propio control en sucesivas ocasiones) u otros factores incontrolables, así como aquellos estudios en los que el objetivo pase por medir un fenómeno cambiante a lo largo del tiempo como es el aprendizaje de los alumnos. Estas características adquieren especial relevancia en contextos como los escolares y en la interpretación de los índices de valor añadido, pues la distribución de alumnos en escuelas no responde a un patrón aleatorio (Raudenbush, 2004; Rubin, Stuart y Zanutto, 2004).

Aunque los diseños longitudinales sean el marco más apropiado para la evaluación del cambio y del crecimiento de los aprendizajes escolares, su aplicación no está exenta de problemas.

En este tipo de diseños la elección de la primera ocasión de medida de la serie temporal es un aspecto de vital importancia, especialmente cuando el número de ocasiones de medida es pequeño. La elección del punto de partida puede tener un gran impacto en la interpretación de los resultados (Stevens y Zvoch, 2006). Con el uso de sólo dos oleadas de medidas no se consigue extraer todo el potencial de los estudios longitudinales ni se permiten estimaciones precisas de la pendiente de la función de producción (Rogosa, 1979; Willet, Singer y Martin, 1998). Tal y como afirma Rogosa (1995), “dos oleadas de datos son mejor que una, pero puede que no mucho mejor” (p. 774). Con sólo dos oleadas de medida las puntuaciones en el pretest influyen en gran medida en la pendiente de la función de crecimiento, pero si se utiliza un conjunto más amplio de medidas, esta pendiente se estima

con mayor fiabilidad debido a las múltiples medias obtenidas (Rogosa, 1995; Willet et al., 1998).

En estos diseños en los que sólo se mide a los alumnos en dos momentos no es posible evaluar la forma de la función de crecimiento; ya que sólo se puede ajustar una función lineal. Con tres oleadas de datos ya es posible evaluar el ajuste de una función lineal y de una cuadrática. Y con cuatro puntos de medida, puede ajustarse una función lineal, cuadrática o curvilínea (Stevens y Zvoch, 2006). En concreto Gray, Jesson, Goldstein, Hedger y Rabash (1995) señalan que el número de periodos necesario para identificar el verdadero crecimiento de los alumnos y de la escuela está entre tres y cinco.

La utilización de ocasiones múltiples de medida permite además aumentar la fiabilidad de la estimación de la función de crecimiento e incrementar la validez interna de la valoración del rendimiento (Cook y Ware, 1983; Rogosa, 1979; Rogosa, Brandt y Zimowski, 1982; Singer y Willet, 1986; Willet *et al.*, 1998). La organización y gestión de los diseños longitudinales no deja de ser un asunto dificultoso, ya que requiere hacer un seguimiento individual de los estudiantes a lo largo de, al menos, dos años, teniendo importantes repercusiones las tasas de mortalidad experimental debidas, entre otras cosas, a la movilidad de los estudiantes. Sin embargo, los diseños que utilizan más de dos ocasiones de medida pueden proporcionar un control más exhaustivo sobre influencias confusas, cuantificar el artefacto de la regresión a la media, y proporcionar un mayor entendimiento de la forma de la función de crecimiento (Rogosa, 1995; Wainer, 2004; Willet *et al.*, 1998).

Los diseños de valor añadido ofrecen información valiosa sobre dos cuestiones fundamentales: los niveles de logro en los aprendizajes escolares y las tasas de crecimiento de estos aprendizajes por cada ocasión de medición. Parece claro en la literatura (Stevens y Zvoch, 2006; Zvoch y Stevens, 2003, 2006) que los factores relacionados con las características sociales individuales de los alumnos están más relacionados con los índices de nivel de conocimientos y menos con los índices de crecimiento en los mismos. Sin embargo, los niveles de conocimiento iniciales de los alumnos están estrechamente asociados con el ajuste de las tasas de crecimiento de los centros.

Es necesario señalar que mientras el interés por la medida del valor añadido es relativamente nuevo, los métodos y procedimientos para desarrollar estos análisis están relativamente bien establecidos

NOTASLA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

en la literatura. Algunas elecciones metodológicas son críticas en la construcción de un sistema para la medida del progreso académico. Referimos las pautas ofrecidas por Thum (2003a; 2003c; 2009) y Martínez Arias (2009) a modo de resumen para presurosos:

- a) Los objetivos de la implantación del modelo de Valor Agregado (VA) han de estar claramente formulados, pues diferentes objetivos pueden condicionar el modelo a desarrollar.
- b) La métrica adecuada para medir el cambio es la de intervalo, ya que es la única que permite comparaciones válidas para medir los cambios. No obstante también puede ser conveniente desde el punto de vista del informe la transformación a categorías ordinales.
- c) El uso de múltiples resultados para reflejar el nivel de aprendizajes y de cambio en los mismos está mejor representado por el conjunto de medidas que reflejan la actuación de los alumnos globalmente, encontrando un entorno natural en los modelos multivariados. El tratamiento simultáneo en los modelos multivariados aporta un conjunto de resultados más coherente que análisis separados para cada materia (Martínez Arias, Gaviria y Castro, 2009), ya que permiten analizar los cambios a lo largo del desarrollo y son más eficientes para el tratamiento de valores perdidos.
- d) La mayoría de los autores recomiendan la utilización de al menos tres ocasiones de medida. Como es obvio, es preciso disponer de un identificador único de cada estudiante que se mantiene durante su permanencia en el sistema y que permita vincularlo a la escuela (lo que supone un adecuado sistema de gestión de bases de datos y de control de calidad de las mismas).
- e) Las ganancias estimadas son las más adecuadas para la medida del cambio; puesto que como ya hemos señalado, es la única que ofrece una definición adecuada del cambio en los niveles de aprendizaje.
- f) El empleo de Modelos que reflejen distintas medidas de agregación permite que el cambio pueda ser seguido a través del tiempo, ya que define la unidad de medida individual. La información para las escuelas y las Comunidades Autónomas, por ejemplo, debe desarrollarse en un esquema de análisis que comience con el seguimiento de cada alumno individual para proseguir con un proceso inferencial apropiado para desagregar los resultados.

- g) Los modelos deben estar basados en la inferencia, informando de la precisión de las estimaciones. Cálculos, proyecciones, comparaciones, incluso rankings del nivel actual de nuestras escuelas y de sus tasas de crecimiento y valor añadido deben ir acompañadas de las fuentes de variabilidades muestrales y de medida. La precisión (error típico de medida) debe estar incluido, para reflejar la fiabilidad de nuestras estimaciones, siendo especialmente importante en el desarrollo de sistemas de evaluación de sistemas educativos.
- h) Los informes han de ser informativos socialmente, pero sin perder la rigurosidad que requieren. El modelo que subyace a las evaluaciones basadas en el valor añadido debe ser claro y discutible entre los profesionales, con el objeto de mejorar la metodología empleada. La consideración del informe a la sociedad o a sectores no técnicos no debe relajar los supuestos y demandas implicados en el desarrollo de estos sistemas, si bien será necesaria una adecuada traducción social de los mismos. Se recomiendan informes claros, que utilicen representaciones gráficas de los avances en los logros de aprendizaje, que señalen las diferencias estadísticamente significativas, para no atribuir significados inapropiados a diferentes ordenaciones de escuelas, por ejemplo. Suele ser conveniente reportar varios tipos de resultados: brutos o no ajustados, ajustados por características del alumnado, como el rendimiento anterior y resultados ajustados por factores contextuales del alumno y de la escuela. También es muy importante en los informes desagregar los resultados por grupos de alumnos, para analizar específicamente sus dificultades y sus ganancias de aprendizaje. Es imprescindible proporcionar una formación adecuada sobre la interpretación de los datos y sus posibles usos a los equipos directivos de los centros, profesorado y otras partes interesadas.

4. ALGUNOS MODELOS DE VALOR AÑADIDO IMPLANTADOS

Algunos modelos de VA ya han sido aplicados con éxito desde hace años para la monitorización de los rendimientos educativos y para la evaluación de escuelas y profesores. En general han sido valorados positivamente y han tenido una buena aceptación por las partes interesadas. En este apartado se describen brevemente algunos de los modelos con mayor tradición, fundamentalmente desarrollados en Estados Unidos e Inglaterra. Hay otros muchos

NOTAS

LA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

modelos de valor añadido que no se han implantado en procesos sistemáticos de evaluación y que se encuentran en fase de investigación (Goldstein y Behuniak, 2005; Martínez Arias, Gaviria y Castro, 2009), esto es debido a que el proceso de operacionalización y desarrollo de modelos de valor añadido está sometido a constante escrutinio y es una muestra del debate y evolución científica que se produce en este campo. Tres de las cuatro experiencias descritas son norteamericanas, y curiosamente, todas ellas han sido desarrolladas antes de la legislación *No Child Left Behind* (2002) que está siendo la principal impulsora del desarrollo de modelos de valor añadido para la evaluación de sistemas educativos.

Probablemente, el modelo de valor añadido más ampliamente citado, por ser el primero en aplicarse en amplia escala es *Tennessee Value Added Assessment System* (TVAAS, actualmente EVAAS). Fue desarrollado por William Sanders y sus colaboradores (Sanders y Horn, 1994; Sanders, Saxton y Horn, 1997). En la actualidad la aplicación del sistema se lleva a cabo desde la empresa SAS (que desarrolla el software estadístico), bajo la dirección del profesor Sanders y se denomina *Education Value-Added Assessment System* (EVAAS). EVAAS ha sido objeto de numerosos análisis (Amrein-Bearsley, 2008; Ballou, 2002; Bock, Wolfe y Fisher, 1996; Braun, 2005a, 2005b; Braun y Wainer, 2007; Kuppermintz, 2003; McCaffrey, Lockwood, Koretz y Hamilton, 2003). Permite obtener perfiles de las escuelas, estableciendo los resultados por quintiles de rendimiento. El método EVAAS proporciona además la denominada “metodología de proyección” que proporciona estimadores del nivel de rendimiento del estudiante individual en algún punto futuro, bajo el supuesto de que tendrá una experiencia de escolarización promedio (Wright, Sanders y Rivers, 2006). El modelo es muy eficiente en el sentido de que hace uso de toda la información para una cohorte de estudiantes en un período de cinco años. La utilización de tantas medidas y de varias materias minimiza los efectos del error de medida y de la fugacidad o temporalidad de los resultados. No utiliza ajustes de otras covariantes. Esta cuestión ha sido sometida a análisis empíricos (McCaffrey *et al.*, 2003). No obstante, Ballou, Sanders y Wright (2004) mostraron la posibilidad de inclusión de estos ajustes en el modelo.

El modelo *Dallas Value Added Assessment System* (Modelo DVAAS) comenzó a implantarse en 1984. A diferencia de EVAAS, DVAAS exige tener en cuenta para la evaluación las diversas características de los estudiantes y del contexto de la escuela. Desde sus orígenes

cuenta con la *Accountability Task Force*, un comité en el que está integrada toda la comunidad educativa. Una buena descripción del modelo de evaluación puede encontrarse en Webster y Mendro (1997) y Webster (2005). El modelo utiliza las puntuaciones longitudinales de los estudiantes de dos años consecutivos. Inicialmente se ajustan las medias utilizando procedimientos ordinarios de regresión. Posteriormente, se utilizan los residuos como base para el cálculo del valor añadido de profesores y escuelas, incluyendo en el esquema jerárquico lineal las variables de centro.

El modelo denominado de “*productividad*” de las escuelas públicas de Chicago evalúa los resultados y las ganancias en competencia lectora y matemáticas de los estudiantes del consorcio entre los cursos 2º y 8º. Considera el nivel inicial de los estudiantes y las tendencias de rendimiento, posibilitando la clasificación de escuelas por medio de un “perfil de productividad”. El modelo está descrito originalmente en Bryk, Thum, Easton y Luppescu (1998) y en la evolución en Ponisciak y Bryk (2005), donde se describen las mejoras introducidas en el modelo multinivel de tres niveles al añadir la clasificación cruzada, que permite tratar adecuadamente los datos perdidos debidos principalmente a la movilidad de los estudiantes.

En las escuelas británicas se han desarrollado un abanico de iniciativas para tratar de elevar los estándares de rendimiento. El *modelo de Valor Añadido Contextualizado en Inglaterra* es descrito por Ray, Evans y McCormack (2009). Una pieza fundamental ha sido el desarrollo de pruebas que pueden vincularse a los logros previos de los estudiantes y usarse para generar estimaciones de la contribución de la escuela al progreso de sus alumnos: indicadores de valor añadido. En Inglaterra, la disponibilidad de datos nacionales de las características de los estudiantes individuales desde 2002 ha permitido el desarrollo de puntuaciones “contextuales de valor añadido” para escuelas y grupos de alumnos dentro de cada escuela. La variable dependiente es un promedio de puntuaciones en una etapa educativa clave (distinta en primaria y en secundaria) y se incluyen en el modelo diez variables contextuales del nivel de alumno y dos del nivel de centro, además del rendimiento previo que es el predictor más importante. Más que usar la aproximación de Mínimos Cuadrados Ordinarios, los indicadores Contextuales de Valor Añadido se derivan de un modelo multinivel (Goldstein, 2003), obtenidos con MLwiN, un paquete estadístico que estima los efectos fijos –rendimiento previo y factores contextuales– y calcula los residuos de nivel 1 (variación de los resultados de los alumnos en

NOTASLA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

relación con sus escuelas) y 2 (variación de los resultados de las escuelas en relación con los resultados nacionales esperados, dados los factores medidos por los efectos fijos). Estos residuos del nivel 2 son las puntuaciones de valor añadido. Un rasgo propio del sistema inglés es que estas puntuaciones de valor añadido son utilizadas tanto por inspectores escolares y los gobiernos locales y nacional, como por los padres, profesores y directores escolares y para el público en general. Estas características pueden analizarse con un paquete estadístico llamado RAISEonline, desarrollado conjuntamente por el gobierno central y la agencia independiente de inspección de las escuelas.

Como se ha podido observar, el desarrollo de sistemas de evaluación basados en la idea del valor añadido constituye un reto social, educativo, intelectual y metodológico. Sin embargo, la información ofrecida por estos sistemas complejos permite responder a cuestiones de verdadera importancia: cuánto saben nuestros alumnos y cuánto evoluciona su aprendizaje debido a la acción de la escuela. Creemos que merece la pena dedicar nuestros esfuerzos a desarrollar un sistema válido y adaptable a nuestro contexto educativo y social.■

Fecha de recepción del original: 25 de octubre de 2008

Fecha de recepción de la versión definitiva: 14 de diciembre de 2008

REFERENCIAS

- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational researcher*, 37, 65-75.
- Ballou, D. (2002). Sizing up test scores. *Education Next*, 2, 10-15.
- Ballou, D., Sanders, W. y Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29, 37-65.
- Bock, R. D., Wolfe, R. y Fisher, T. H. (1996). *A review and analysis of the Tennessee value-added assessment system. Summary and recommendations*. Nashville, TN: Tennessee Office of Education Accountability.
- Braun, H. (2005a). Using student progress to evaluate teachers: *A primer on value-added models*. Princeton, NJ: Educational Testing Service.
- Braun, H. (2005b). *Value-added modelling: What does due diligence require?* En R. Lissitz (Ed.), *Value-added models in education: Theory and applications* (pp. 19-39). Mapple Grove, MN: JAM Press.
- Braun, H. y Wainer, H. (2007). Value-added modelling. En S. Sinharay y C. Rao (Eds.), *Handbook of statistics* (Vol. 26: Psychometrics, pp. 867-891). Amsterdam: North Holland.
- Bryk, A. S., Thum, M. Y., Easton, Q. J. y Luppescu, S. (1998). *Academic productivity of Chicago public elementary schools*. Chicago, IL: Consortium on Chicago School Research.
- Bryk, A. S. y Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Bryk, A. S. y Weisberg, H. I. (1976). Value-added analysis: A dynamic approach to the estimation of treatment affects. *Journal of Educational Statistics*, 1, 127-155.
- Cook, N. R. y Ware, J. H. (1983). Design and analysis methods for longitudinal research. *Annual Review of Public Health*, 4, 1-23.
- Fitz-Gibbon, C. T. (1997). *The Value-added National Project: Final report. Feasibility studies for a national system of value-added indicators*. London: SCAA.
- Gaviira, J. L. y Castro, M. (2005). *Modelos jerárquicos lineales*. Madrid: La Muralla.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Goldstein, H. (2003). *Multilevel models*. London: Arnold.
- Goldstein, H. y Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A*, 159, 384-443.
- Goldstein, J. y Behuniak, P. (2005). Growth models in action: Selected case studies. *Practical Assessment, Research & Evaluation*, 10(11). Extraído de <http://pareonline.net/pdf/v10n11.pdf>
- Gray, J., Jesson, D., Goldstein, H., Hedger, K. y Rabash, J. (1995). A multilevel analysis of school improvement: Changes in schools' performance over time. *School Effectiveness and School Improvement*, 6(2), 97-114.
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis*, 25(3), 287-298.
- Laird, N. M. y Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrika*, 65, 581-590.

NOTAS

LA EVALUACIÓN EDUCATIVA
DESDE LA PERSPECTIVA DEL
VALOR AÑADIDO

- Martínez Arias, R. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista de Educación*, 348, 217-250.
- Martínez Arias, R.; Gaviria, J. L. y Castro, M. (2009). Concepto y evolución de los modelos de valor añadido en educación. *Revista de Educación*, 348, 15-46.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M. y Hamilton, L. S. (2003). Evaluating value-added models for teacher accountability. Santa Mónica, CA: *RAND Corporation*.
- Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. En L. M. Collins y J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future decisions* (pp. 92-105). Washington, D.C.: American Psychological Association.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- Ponisciak, S. M. y Bryk, A. S. (2005). Value-added analysis of the Chicago public schools: An application of hierarchical models. En R. Lissitz (Ed.), *Value-added models in education: Theory and applications* (pp. 40-79). Maple Grove, MN: JAM Press.
- Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, 52, 501-525.
- Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practices? *Journal of Educational and Behavioral Statistics*, 29(1), 121-129.
- Raudenbush, S. W. y Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. London: SAGE.
- Ray, A., Evans, H. y McCormack, T. (2009). El uso de los modelos nacionales de valor añadido para la mejora de las escuelas británicas. *Revista de Educación*, 348, 47-66.
- Rogosa, D. R. (1979). Causal models in longitudinal research: Rationale, formulation and interpretation. En J. R. Nesselroade y P. B. Baltes (Eds.), *Longitudinal research in the study of behaviour and development* (pp. 263-302). New York: Academic Press.
- Rogosa, D. R. (1995). Myths about longitudinal research. En J. M. Gottman (Ed.). *The analysis of change* (pp. 4-66). Mahwah, N.J.: Lawrence Erlbaum.
- Rogosa, D. R., Brandt, D. y Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 90, 726-748.
- Rubin, D., Stuart, E. y Zanutto, E. (2004). A potential outcome view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, 29(1), 103-116.
- Ruiz de Miguel, C. y Castro, M. (2006). Estudio multinivel basado en PISA 2003: factores de eficacia escolar en el área de matemáticas. *Education Policy Analysis Archives*, 14(29). Extraído de <http://epaa.asu.edu/epaa/v14n29>
- Sanders, W. L. y Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation*, 9, 299-311.
- Sanders, W. L., Saxton, A. M. y Horn, S. P. (1997). The Tennessee Value-Added Assessment System: A quantitative, outcomes-based approach to educational assessment. En J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137-162). Thousand Oaks, CA: Corwin Press.
- Saunders, L. (1999). A brief history of educational 'Value Added': How did we get to where we are? *School Effectiveness and School Improvement*, 10, 233-256.

NOTAS

MARÍA CASTRO MORERA Y
JOSÉ LUIS GAVIRIA SOTO

- Singer, J. D. y Willet, J. B. (1986). Methodological issues in the design of longitudinal research: Principles and recommendations for a quantitative study of teachers' careers. *Educational Evaluation and Policy Analysis*, 18, 265-283.
- Snijders, T. y Bosker, R. (1999). *Multilevel analysis*. Thousand Oaks, CA: Sage.
- Stevens, J. J. (2005). The study of school effectiveness as a problem in research design. In R.W. Lissitz (Ed.), *Value added models in education: Theory and applications* (pp. 166-203). Maple Grove, MN: JAM Press.
- Stevens, J. J. y Zvoch, K. (2006). Issues in the implementation of longitudinal growth models for student achievement. En Lissitz, R. W. *Longitudinal and value added models of students' performance* (pp. 170-208). Maple Grove, MN: JAM Press.
- Thum, Y. M. (2003a). Making AYP-NCLB: A sample analysis. Ponencia presentada en el Annual Meetings of the American Educational Research Association, Chicago, IL.
- Thum, Y. M. (2003b). Measuring progress towards a goal: Estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods & Research*, 32(2), 153-207.
- Thum, Y. M. (2003c). *No Child Left Behind: Methodological challenges & recommendations for measuring adequate yearly progress* (CSE Technical Report 590). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, UCLA.
- Thum, Y. M. (2009). No Child Left Behind: retos metodológicos y recomendaciones para medir el progreso anual adecuado. *Revista de Educación*, 348, 67-90.
- Wainer, H. (2004). Introduction to the value-added assessment special issue. *Journal of Educational and Behavioral Statistics*, 29(1), 1-4.
- Webster, W. J. (2005). The Dallas school level accountability model: The marriage of status and value-added approaches. En R. Lissitz (Ed.), *Value added models in education: Theory and applications* (pp. 233-271). Maple Grove, MN: JAM Press.
- Webster, W. J. y Mendro, R. L. (1997). The Dallas Value-Added Accountability System. En J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 81-99). Thousand Oaks, CA: Corwin Press.
- Willet, J. B.; Singer, J. D. y Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Developmental and Psychopathology*, 10, 395-426.
- Wright, S. P., Sanders, W. L. y Rivers, J. C. (2006). Measurement of academic growth of individual students toward variable and meaningful academic standards. En R. Lissitz (Ed.), *Longitudinal and value added models of student performance* (pp. 385-406). Maple Grove, MN: JAM Press.
- Zvoch, K. y Stevens, J. (2003). A multilevel, longitudinal analysis of middle school math and language achievement. *Educational Policy Analysis Archives*, 11(20). Extraído de <http://epaa.asu.edu/epaa/v11n20/>
- Zvoch, K. y Stevens, J. (2006). Successive student cohorts and longitudinal growth models: An investigation of elementary school mathematics performance. *Educational Policy Analysis Archives*, 14(2). Extraído de <http://epaa.asu.edu/epaa/v14n2/>