

# ¿Qué información proporciona el Examen Nacional de Enseñanza Media (ENEM) a la sociedad brasileña?

*Rodrigo Travitzki, Jorge Calero y Carlota Boto*

## RESUMEN

Se evalúan las limitaciones y posibilidades del Examen Nacional de Enseñanza Media (ENEM) como indicador de la eficacia de los colegios en el Brasil, así como los efectos de la introducción de variables de contexto. Se realizó un análisis de regresión multinivel en tres niveles (individuo, colegio y estado) utilizando microdatos para 2009 y 2010 sobre 17.359 colegios. El contexto permitió explicar el 79% de la diferencia entre colegios. El promedio bruto y el valor añadido (efecto aleatorio a nivel del colegio) presentaron evaluaciones opuestas en el 34% de los casos. El promedio se mostró más estable ( $r=0,8$ ) que el valor añadido ( $r = 0,5$ ) en ambos años. Se encontraron diversas limitaciones del ENEM como indicador de eficacia escolar. Los resultados señalan que esta clasificación informa más sobre las condiciones socioeconómicas que sobre el mérito de los colegios, es decir, respecto del valor añadido que estos supuestamente proporcionan a los alumnos.

---

## PALABRAS CLAVE

Enseñanza secundaria, exámenes, establecimientos de enseñanza, clasificación, calidad de la educación, medición, Brasil

## CLASIFICACIÓN JEL

I24, I28, C18

## AUTORES

Rodrigo Travitzki es doctor en Educación de la Universidad de São Paulo, Brasil. [r.travitzki@gmail.com](mailto:r.travitzki@gmail.com)  
Jorge Calero es catedrático del Departamento de Economía Pública, Economía Política y Economía Española de la Universidad de Barcelona, España. [jorge.calero@ub.edu](mailto:jorge.calero@ub.edu)  
Carlota Boto es profesora de la Facultad de Educación de la Universidad de São Paulo, Brasil. [reisboto@usp.br](mailto:reisboto@usp.br)

# I

## Introducción

En los años noventa, en algunos países se publicaron clasificaciones escolares con el objetivo de mejorar la calidad de los colegios y su rendición de cuentas a la sociedad, proporcionando a la vez información que ayudara a los padres a escoger el colegio para sus hijos (Karsten, Visscher y De Jong, 2001). Normalmente, este tipo de estrategia forma parte de políticas moderadas de responsabilidad o rendición de cuentas con que se procura informar al gobierno y a las familias e identificar las buenas prácticas del sistema educativo, pero sin vincular los resultados a recompensas o penalizaciones, como se hace en las políticas de rendición de cuentas con recompensas y sanciones (Martínez Arias, 2009). Aunque la formulación general de esas políticas sea importante, sus efectos dependen principalmente de las medidas de calidad utilizadas (Ladd y Walsh, 2002). En consecuencia, se producen controversias sobre cuestiones técnicas y metodológicas relacionadas con la elaboración de los indicadores empleados en las clasificaciones, y con su capacidad de promover de hecho la calidad en los colegios. En la práctica, los países adoptan diferentes estrategias.

En Inglaterra, por ejemplo, se publican clasificaciones escolares desde 1992 y a partir de ellas se crearon sistemas de incentivos (West y Pennell, 2000). En el otro extremo, se pueden situar las políticas educativas de España, que prohíben la publicación de resultados por colegios (Gobierno de España, 2006, art. 140). El Brasil, por su parte, se encuentra actualmente en una situación intermedia, ya que la publicación de clasificaciones no está vinculada a incentivos, si bien en algunos estados existen políticas de rendición de cuentas con recompensas y sanciones. Algunos autores defienden el fortalecimiento de esas políticas en el ámbito nacional (Andrade, 2008), mientras que otros mantienen posturas más cautelosas y críticas, subrayando por ejemplo la tendencia a incrementar la desigualdad (Franco y otros, 2007).

El primer aspecto polémico consiste en el propio concepto de calidad, que es marcadamente polisémico cuando se aplica al campo educativo (Murillo Torrecilla,

2005). En la historia reciente del Brasil, la idea de calidad fue tomando distintas formas. Inicialmente, se vinculó a la universalización del acceso, después al flujo y la tasa de repetición y luego al desempeño de los alumnos en exámenes a gran escala (Oliveira y Araujo, 2005). La capacidad de lograr buenos resultados normalmente se denomina eficacia, pero conviene recordar que un colegio eficaz no es necesariamente un colegio de calidad, sino que la eficacia es más bien una condición necesaria pero no suficiente (Murillo Torrecilla, 2005, pág. 31). Es decir, la calidad es un concepto más amplio que la eficiencia y tiene diferentes significados.

Aunque la calidad escolar es un concepto controvertido, con frecuencia se ha vinculado al desempeño de los alumnos en exámenes estandarizados, no solo por el deseo de encontrar medidas objetivas, sino también por motivos prácticos, como el costo y la viabilidad. En este trabajo no se discutirá esa cuestión, que es fundamental, sino que su desarrollo se circunscribirá a un aspecto de la calidad escolar: la eficacia en preparar a los alumnos para realizar exámenes estandarizados. Se trata de un enfoque bastante limitado, pero necesario, teniendo en cuenta que en el Brasil, al igual que en el resto del mundo, este tipo de indicador está cada vez más presente en las políticas educativas y en el imaginario de las personas.

Otras controversias sobre las clasificaciones de los colegios se refieren a su valoración excesiva (Brandão, 2000), al aumento de la exclusión social (West y Pennell, 2000), a la reproducción del privilegio de clases (Apple, 2001), a la retroalimentación que beneficia a los mejores y perjudica a los peores (Ladd y Walsh, 2002) y a la falta de atención a los exámenes propiamente dichos (Reckase, 2004). También se critica el uso de indicadores individuales agregados (realizando el promedio, por ejemplo) como forma de evaluar a las instituciones (Meyer, 1997) y el tipo de información que se selecciona para su publicación (Van Petegem y otros, 2005). Finalmente, se observa que hay pocos estudios sobre cómo utilizar esa información para que los colegios puedan mejorar el proceso de aprendizaje de los alumnos (Heck, 2000).

En el presente trabajo se realiza un análisis crítico de este tipo de indicadores, utilizando los datos cuantitativos del Examen Nacional de Enseñanza Media (ENEM), que sirve como criterio de selección para las universidades brasileñas y que podría convertirse en el indicador oficial de calidad escolar en el ámbito nacional

---

□ Este trabajo contó con el apoyo del organismo brasileño de fomento de la investigación: Coordinación de Mejoramiento del Personal de Nivel Superior (CAPES por sus siglas en portugués).

(Passarinho, 2012)<sup>1</sup>. Algunas de las conclusiones de este estudio se refieren al propio examen, mientras que otras remiten a cuestiones metodológicas más generales sobre los modelos de valor añadido en la educación. En particular, se desea evaluar qué tipo de información proporciona la clasificación del ENEM a la sociedad y cómo la introducción de variables de contexto interfiere en los resultados por colegios, tanto transversal como longitudinalmente.

### La importancia del contexto

Como muchas otras, la clasificación del ENEM se basa en la publicación de los promedios por colegios. Sin embargo, dado que esas instituciones parten de condiciones muy heterogéneas, especialmente en los países en desarrollo, es necesario tener en cuenta los factores contextuales para evaluar con mayor precisión el efecto obtenido por cada colegio (Heck, 2000), es decir, su mérito. Según Meyer (1997, pág. 298), el promedio de puntuación en el examen, uno de los indicadores más usados en la educación estadounidense, resulta altamente cuestionable como indicador de desempeño escolar y constituye un instrumento muy débil, o incluso contraproducente, para la rendición de cuentas.

Muchos autores de la literatura especializada defienden esta postura, especialmente en lo relativo a la responsabilidad (Willms, 2006), pero también hay quien sostiene que tanto el promedio bruto como el “neto” (es decir, el obtenido después de controlar el efecto de las

variables de contexto) pueden dar lugar a resultados distorsionados (Tekwe y otros, 2004). Por otra parte, incluso los modelos multinivel de valor añadido, la “última generación” de indicadores de calidad escolar, pueden presentar una gran variedad de resultados, dependiendo de las variables de contexto integradas (Keeves, Hungi y Afrassa, 2005; Ladd y Walsh, 2002).

A pesar de estas limitaciones, si se desea realizar comparaciones equitativas entre los colegios con el objeto de asegurar su responsabilidad, es fundamental tener en cuenta el contexto de cada uno de ellos. Como señala Thomas (1998, pág. 92), la publicación de clasificaciones basadas en promedios brutos no ayuda ni a los colegios de buen desempeño ni a aquellos cuyo desempeño es peor. En el primer caso, no se percibe la necesidad de perfeccionamiento y, en el segundo, puede producirse una grave desmoralización de educadores y funcionarios que no son responsables de la situación.

Según Thomas (1998), en un estudio publicado en 1992 en el periódico *The Guardian*, se llegó a la conclusión de que el 23% de los colegios se evaluaron de forma diferente en la clasificación “bruta” y en la “neta”. ¿En qué medida estas conclusiones confirman los datos del ENEM?

Luego de la Introducción, el presente artículo se desarrolla de la siguiente manera: la sección II se refiere a los modelos de valor añadido según diferentes conceptos y autores; en la sección III se plantea por qué es necesario evaluar los colegios; en la cuarta sección se presenta el Examen Nacional de Enseñanza Media (ENEM) como indicador de calidad escolar; en la sección V se entregan los principales resultados y en la sexta sección, las conclusiones pertinentes.

## II

### Modelos de valor añadido

Existen diferentes conceptos de valor añadido, que incluso pueden ser contradictorios entre sí (Saunders, 1999). Aunque algunos autores consideran que estos deben basarse en datos longitudinales (Martínez Arias, Gaviña Soto y Castro Morera, 2009)<sup>2</sup>, en este estudio se parte de un concepto más amplio de valor añadido

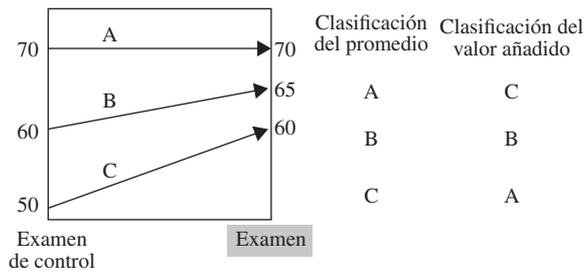
(Reckase, 2004), que incluye, por ejemplo, el efecto del colegio. La idea general es que se desea evaluar cuánto mejoran los alumnos gracias al trabajo del colegio y no debido a sus condiciones previas, intentando eliminar la influencia de factores más allá del control del colegio (McCaffrey y otros, 2004).

En el gráfico 1 se ilustra esa idea general, que puede aplicarse tanto a datos transversales como longitudinales. En el primer caso, que corresponde al efecto del colegio, solo se dispone de un examen

<sup>1</sup> El Índice de Desarrollo de la Educación Básica (IDEB).

<sup>2</sup> El argumento principal consiste en que en el rendimiento previo se condensan las variables referentes al nivel socioeconómico (Ferrão, 2009).

GRÁFICO 1

**Concepto general de valor añadido**

Fuente: elaboración propia.

realizado por los alumnos, ya que el “examen de control” se estima utilizando variables contextuales. Es decir, se procura determinar cuál habría sido la nota de los alumnos (o del colegio) si todos hubiesen partido de las mismas condiciones de contexto. En el segundo caso, el “examen de control” es real y, por ejemplo, fue aplicado a los alumnos antes de que entraran en el colegio. En este segundo caso, por lo tanto, se tienen al menos dos exámenes diferentes aplicados a los mismos alumnos. Cada una de las metodologías parte de sus propios presupuestos y presenta limitaciones, de manera que la opción por una u otra depende en gran medida de los datos disponibles<sup>3</sup>.

<sup>3</sup> El valor añadido con datos longitudinales presupone, por ejemplo, la similitud de los instrumentos de medida utilizados a lo largo de los años en lo referente a su finalidad, forma y grado de dificultad (en el caso de los exámenes). El valor añadido con datos transversales, por su parte, no solo presupone la existencia de datos de contexto, sino

Actualmente, se emplean modelos de valor añadido para garantizar la responsabilidad o rendición de cuentas en Tennessee, Dallas, Chicago (Estados Unidos de América) y en Inglaterra (Martínez Arias, 2009). Los modelos basados en datos longitudinales son recientes y su uso comenzó en Inglaterra en el cambio de milenio, mediante la creación de un número de identificación individual para los alumnos (Ray, Evans y McCormack, 2009). Los estudios sobre el efecto del colegio (valor añadido con datos transversales y contextuales) existen desde el comienzo de los años ochenta y en ellos se estima que entre el 5% y el 35% de la variación en las notas de los alumnos se puede explicar por el colegio en que estudian (Martínez Arias, Gaviria Soto y Castro Morera, 2009).

La estabilidad del efecto del colegio a lo largo de los años sigue siendo un tema polémico. Hay datos que muestran que pocos colegios tienen resultados consistentes (en los distintos alumnos) y estables (a lo largo del tiempo) (Thomas y otros, 1997). En diversos estudios se ha calculado el coeficiente de correlación del efecto del colegio en diferentes años. Según los casos, ese coeficiente era igual a cero (Linn y Haug, 2002), estaba en torno de 0,6 (Mandeville, 1988), o se encontraba entre 0,35 y 0,65 para los colegios de enseñanza primaria y entre 0,70 y 0,95 para los colegios de enseñanza secundaria (Luyten, 1994). De nuevo, cabe preguntar en qué medida los microdatos del ENEM confirman estos resultados.

también que esos datos representan adecuadamente las condiciones iniciales del alumno.

### III

## ¿Por qué evaluar a los colegios?

En lo atinente al uso de los datos producidos por este tipo de indicador (los basados en las pruebas estandarizadas), hay al menos dos cuestiones significativas, íntimamente relacionadas: ¿para qué sirven esos datos y a quién van dirigidos? Un enfoque mínimamente consistente sobre este tema requiere su propia investigación. Para los efectos de este estudio, basta con identificar dos tipos de uso: la rendición de cuentas de los colegios públicos y la elección de colegio por los padres. El primer aspecto se relaciona con los mecanismos estatales y el segundo con los de mercado.

¿Cómo se relaciona la finalidad con el instrumento de medida? Es posible considerar, por ejemplo, el tamaño de un objeto. Como el “tamaño” es un concepto objetivo y consensual, se puede utilizar la misma regla sea cual sea la finalidad de la medida. En el peor de los casos, se cambia el instrumento si el objeto es muy grande y el uso de una regla común resulta poco práctico y preciso. Sin embargo, dado que la calidad escolar no es un concepto consensual, parece razonable emplear distintos indicadores para finalidades diferentes. Existen investigaciones en esta línea con las que se procura

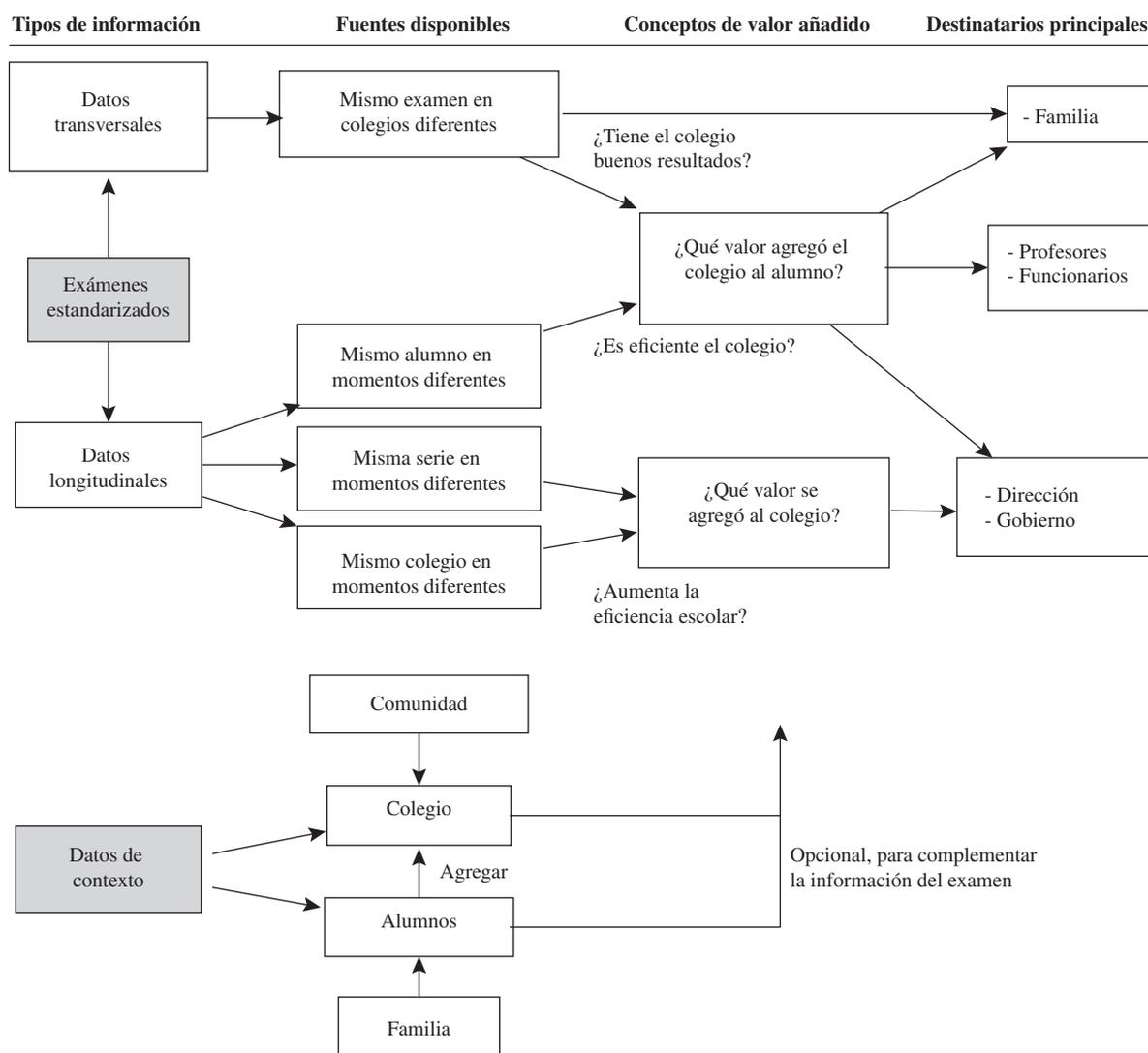
desarrollar formas diversas de estimar la calidad de los colegios, teniendo en cuenta su utilidad para la familia y para el estado, por ejemplo (Meyer, 1997). De manera general, se considera que los promedios brutos, si bien resultan informativos en cuanto al desempeño de los colegios, producen comparaciones injustas para los administradores, profesores y alumnos (Willms, 2006). En este sentido, se propusieron dos tipos de efecto del colegio, el primero (tipo A) más relacionado con el desempeño general y el segundo (tipo B), con el objetivo de aislar los factores sobre los que el colegio tiene algún control.

“El efecto de tipo B es el efecto que las autoridades escolares consideran al evaluar el desempeño de los que trabajan en las escuelas. Una escuela con un contexto desfavorable podría producir un gran efecto de tipo B mediante el esfuerzo y el talento de su personal docente. La escuela acertadamente ganaría el respeto de los evaluadores, no obstante, los padres al privilegiar un alto efecto tipo A podrían no elegir esa escuela para sus hijos” (Raudenbush y Willms, 1995, pág. 310).

Las principales ideas descritas se resumen en el gráfico 2, donde se introduce también otro concepto de valor añadido relativo a la calidad escolar. Hasta ahora,

GRÁFICO 2

**Visión general de los modelos de valor añadido**



Fuente: elaboración propia.

el valor que el colegio agrega al alumno se consideraba por medio de datos transversales o longitudinales. Sin embargo, también se puede estimar el valor añadido del colegio durante cierto tiempo, ya sea por el propio establecimiento, las políticas educativas, la comunidad o por los cambios económicos y sociales, entre otros. Este segundo concepto de valor añadido fue utilizado en el Brasil en la definición de las metas para cada colegio propuestas en el Plan Nacional de Educación<sup>4</sup>.

Resulta evidente, entre otras cosas, que la fuente principal de información para evaluar la calidad escolar la constituyen los exámenes individuales estandarizados (hecho que supone una clara limitación metodológica) y que para distintos sectores de la sociedad podrían utilizarse diferentes indicadores. Es necesario tener en

cuenta esta pluralidad, porque un énfasis excesivo en la elección por parte de los padres tiende a aumentar las desigualdades (Apple, 2001).

Según parece, existe una tensión constitutiva en el campo de la evaluación, que se aplica tanto a los colegios como a los alumnos de cada uno de ellos. Se trata del problema de la calidad frente a la equidad. Dubet describe esta tensión en las relaciones intraescolares, al reflexionar sobre lo que sería un sistema escolar justo, mostrando que no existe una solución perfecta, sino una combinación de opciones y respuestas necesariamente limitadas (Dubet, 2004, pág. 540).

Cabe preguntarse, por lo tanto, cómo sería un sistema justo de evaluación escolar que pudiese abarcar esa tensión y esa multiplicidad. Teniendo en cuenta que, en la actualidad, la evaluación de los colegios brasileños se realiza solo mediante exámenes individuales, sin tener en cuenta la situación de contexto, parece que aún hay mucho trabajo por hacer en ese sentido.

<sup>4</sup> En septiembre de 2012, el Plan todavía no había sido aprobado por el Congreso Nacional.

## IV

### El ENEM como indicador de calidad escolar

El Examen Nacional de Enseñanza Media (ENEM) fue creado en 1998 por el Instituto Nacional de Estudios e Investigaciones Educativas Anísio Teixeira (INEP por sus siglas en portugués), organismo vinculado al Ministerio de Educación del Brasil, con el objetivo explícito de proporcionar a los alumnos una autoevaluación al final de la educación básica (MEC, 2002), aunque es razonable suponer que también constituye una iniciativa para crear políticas de responsabilidad en el país. Esta herramienta era inicialmente poco ambiciosa, pero experimentó diversas modificaciones a lo largo de los años, tornándose cada vez más amplia y enraizándose en la cultura educativa. El ENEM se aplica a cerca de 5 millones de alumnos cada año y es el segundo mayor examen de enseñanza secundaria del mundo, después del *gaokao* chino, con sus 10 millones de alumnos (Zhang y Zha, 2010). Esto se debe a que, pese a ser voluntario desde su creación, el ENEM pasó a ser utilizado como examen de selección para las universidades públicas y para la obtención de becas en las universidades privadas, entre otras finalidades. En el cuadro 1 se observa el rápido crecimiento del examen y sus transformaciones a lo largo de los años.

La clasificación del ENEM abarca el 61% de los colegios brasileños de educación secundaria y el 7% del total, de manera que cada vez se utiliza más como indicador de calidad escolar<sup>5</sup>. Aunque el INEP no difunde una clasificación ordenada, la publicación de los promedios por colegios permite que la prensa divulgue una lista ordenada por promedios. La primera clasificación data de 2006 y tuvo al menos dos objetivos: movilizar a la sociedad para mejorar la enseñanza y ayudar a los educadores, directores y administradores a identificar las deficiencias y las buenas prácticas en el ámbito escolar (INEP/MEC, 2007). En 2011, junto con los promedios, empezó a publicarse la tasa de participación de cada colegio (en cuatro franjas) y se recomendó comparar únicamente a los colegios con tasas de participación similares (INEP/MEC, 2011a). El objetivo de esta iniciativa era minimizar el efecto de sesgo de muestreo provocado por el carácter voluntario del examen, pero en cualquier caso es necesario resaltar la falta de significatividad de la comparación de colegios con menos de un cuarto de alumnos participantes en este indicador.

<sup>5</sup> Basado en los datos de 2009 (véase el cuadro A.1 del anexo).

CUADRO 1

**Brasil: resumen histórico del Examen Nacional de Enseñanza Media (ENEM)**

Año	Participantes	Cambios
1998	157 221	Primer ENEM: voluntario, solo para autoevaluación.
2001	1 624 131	Introducción del número de identificación (ID) del colegio, igual al del censo pero diferente cada año.
2003	1 882 393	Grandes cambios en el equipo responsable debido a las elecciones presidenciales del año anterior.
2005	3 004 491	El ENEM empieza a utilizarse como forma de acceder a la enseñanza superior y como criterio para recibir becas del gobierno según el programa Universidad para Todos (ProUni) en las instituciones privadas.
2006	3 742 827	Publicación de la primera clasificación de colegios del ENEM.
2007	3 584 569	El número de identificación de los colegios (ID) pasa a ser el mismo a lo largo de los años, haciendo posible el análisis longitudinal.
2009	4 148 721	Cambios estructurales en el examen, incluida la matriz de competencias, el formato de la prueba y su duración.
2010	4 626 094	El ENEM empieza a utilizarse como certificado de finalización de la enseñanza media para cualquier ciudadano con más de 18 años. La mayoría de las universidades federales utiliza el ENEM como criterio de selección.

Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009.

## 1. Metodología

Debido a la estructura jerárquica de los datos disponibles, en este estudio se realizaron análisis de regresión multinivel, técnica que se considera más adecuada para estos casos (Raudenbush y Willms, 1995). En este análisis se incluyeron tres niveles (individuo, colegio y estado), mediante el uso de paquetes específicos de programas computacionales (*software*) de código abierto R (Bliese, 2012). Los microdatos del ENEM de 2009 y de 2010 y del Censo Escolar de 2009 se obtuvieron del portal del INEP.

Se ajustaron los modelos univariados y multivariados empleando diversas variables con respecto a los colegios y los alumnos (véase el cuadro A.2 del anexo), incluidos como efecto aleatorio no solo los interceptos en el ámbito de los estados y de los colegios, sino también el indicador del nivel socioeconómico de los individuos (componente NSE). En otras palabras, se desea controlar el efecto de los estados y aislar el efecto del colegio, permitiendo que cada colegio tenga una relación o pendiente diferente entre el nivel socioeconómico y la nota del alumno, es decir, su propia pendiente del gradiente socioeconómico (GSE), tal como se define en la Organización de Cooperación y Desarrollo Económicos (OCDE, 2010).

El valor añadido de los colegios (a partir de datos transversales) fue el efecto aleatorio de este nivel en los modelos completos, que incluyeron todas las variables significativas. El valor añadido longitudinal de los colegios se obtuvo mediante la simple resta de los resultados de 2010 y 2009. Para calcular la varianza explicada de los

modelos multinivel, se compararon sus residuos con los de modelos nulos (Raudenbush y Bryk, 2002).

Cabe resaltar una significativa limitación de la metodología utilizada. Debido al carácter voluntario del examen, es posible que exista un efecto de sesgo de muestreo, que justificaría un análisis de regresión en dos etapas (Heckman, 1976). Ese análisis podría realizarse en estudios futuros, ya que aún no existen versiones del programa R que realicen regresiones multinivel en dos etapas.

## 2. Cálculo de los promedios de los resultados por colegios

Según la documentación oficial (INEP/MEC, 2010), se elaboraron tres promedios para cada colegio: uno de la redacción, otro de la prueba objetiva y otro general, siendo el general el promedio ponderado de los dos primeros. Sin embargo, solo se consideraron en el cálculo de estos promedios los alumnos que:

- i) afirmaron que estaban en el último año de la enseñanza media;
- ii) cursaban enseñanza media regular o enseñanza media de jóvenes y adultos (consistente con los datos del censo), ya fuera en colegios tradicionales o por ciclos<sup>6</sup>, y
- iii) que estaban presentes en los días de la prueba.

<sup>6</sup> Los colegios por ciclos unen varias series en una sola clase, en general porque son más pequeños.

Dada esta muestra, continúa la documentación, los criterios para la divulgación de los promedios fueron:

- i) no publicar si la tasa de participación fue inferior al 2%<sup>7</sup>;
- ii) publicar el promedio de la prueba objetiva si al menos 10 alumnos hicieron las 4 pruebas;
- iii) publicar el promedio de la redacción si al menos 10 alumnos hicieron la redacción, y
- iv) publicar el promedio total si al menos 10 alumnos hicieron las 4 pruebas y la redacción.

En este estudio se utilizaron los mismos criterios en el proceso de limpieza de los datos y en el cálculo de los promedios, con el fin de reproducir los resultados divulgados en Internet<sup>8</sup>. Con respecto a ese proceso, fue necesario articular los microdatos del ENEM con los del censo escolar.

### 3. El componente NSE

El indicador de nivel socioeconómico (componente NSE) se creó empleando una metodología similar a la del Programa de Evaluación Internacional de Estudiantes (PISA por sus siglas en inglés) de 2009. Hubo que realizar algunas adaptaciones, sobre todo debido a las diferencias existentes en los datos disponibles en el cuestionario, no solo entre el ENEM y el programa PISA, sino también —inesperadamente— entre los exámenes brasileños de 2009 y 2010.

Después de analizar la matriz de correlación con diversas variables, se escogieron cinco variables y se agruparon en tríos para elaborar cuatro posibles candidatos a componente NSE, que solo se diferenciaban en las variables incluidas (véase el cuadro 2). Después, se calcularon las correlaciones entre los cuatro candidatos y los promedios del colegio, además de otras métricas, con el fin de comparar la calidad de los candidatos en lo referente a su poder explicativo de la nota en el ENEM.

Entre los cuatro candidatos, el componente NSE 0 corresponde al trío de variables más semejante al incluido en el índice socioeconómico del programa PISA de 2009: suma de las posesiones del hogar, máximo nivel educativo de los padres y máximo nivel profesional de los padres<sup>9</sup>. Sin embargo, el componente que mejor se ajustó según todos los criterios fue el componente

NSE 2. Como el objetivo era controlar al máximo las variables contextuales, se escogió este componente como variable socioeconómica del estudio. Curiosamente, está compuesto por dos variables económicas, a diferencia del indicador del programa PISA. El significado de este hecho resulta controvertido. Podría ser una cuestión de opción, una variación natural de los índices, o que el índice del nivel profesional —elaborado utilizando los resultados de países más desarrollados— no sea adecuado para el contexto de la sociedad brasileña, o simplemente que no resulte apropiado por algún otro motivo. También es posible que los antecedentes económicos de la familia interfieran más en las notas del alumno que la profesión de los padres. En este estudio no se han investigado esas hipótesis, sino que solo se buscó el modelo que mejor se ajustase a las notas de los alumnos del ENEM de 2009, después de filtrar los datos según los criterios del INEP.

En resumen, el componente NSE en el nivel individual es el primero de un análisis de componentes principales con tres variables:

- i) la suma de posesiones del hogar;
- ii) el promedio del nivel educativo de los padres, y
- iii) los ingresos familiares.

Una vez calculado en el nivel individual, el índice se agregó formando promedios en el ámbito de los colegios y de los estados (si bien este último nivel no fue significativo y, por lo tanto, no forma parte de los análisis).

Para la comparación entre 2009 y 2010 en el análisis longitudinal, se calculó el componente NSE con un trío diferente de variables, a saber, el nivel educativo de la madre, el nivel educativo del padre y los ingresos familiares. Esto fue necesario debido a la reducción del cuestionario de un año al siguiente, con la eliminación de la información sobre el nivel profesional de los padres y de las posesiones del hogar en 2010. Para comprobar la estabilidad de los datos longitudinales se realizaron correlaciones (correlación de Pearson,  $p < 0,01$ ) entre los dos años, utilizando la misma metodología (en términos del cálculo del componente NSE y de las variables incluidas en el modelo completo). También se calcularon las correlaciones empleando metodologías diferentes (el mejor modelo de 2009 y el modelo común a 2009 y a 2010), para comprobar la influencia de la elección de variables en la estabilidad de los datos longitudinales.

### 4. Los datos disponibles

Los microdatos del ENEM son bastante heterogéneos. Después del proceso básico de limpieza, quedaron cerca de un cuarto de los datos de los dos años (véase

<sup>7</sup> La tasa de participación de los colegios organizados por ciclos se dividió por tres.

<sup>8</sup> Debido a que el INEP no proporciona microdatos de los colegios en el ENEM, sino solo de los alumnos.

<sup>9</sup> Respectivamente, los índices HOMEPOS, PARED e HISEI, descritos en OCDE (2010).

el cuadro 3), en buena medida debido a la ausencia del número de identificación (ID) del colegio en los registros individuales. Por otra parte, incluso entre los colegios identificados, quedó poco más de la mitad de los datos después de la aplicación de los criterios de validez del INEP.

Este hecho no es problemático en sí mismo, pero puede provocar algunas limitaciones del ENEM en lo referente a la comparación de colegios, que depende de algunos criterios parcialmente arbitrarios, como un número mínimo de 10 alumnos o una tasa mínima de un 2% de participación.

CUADRO 2

**Brasil: diferentes formas de calcular el NSE. Según la metodología del programa PISA 2009**

Posibles componentes NSE	Variables incluidas	Coefficiente de determinación ajustado <sup>a</sup> (en porcentajes)	Datos que faltan <sup>b</sup> (en porcentajes)	Proporción de varianza en el primer componente (en porcentajes)	Verosimilitud (criterio de información de Akaike) <sup>c</sup>
NSE 0	- Suma de los bienes - Máximo nivel educativo familiar - Máximo nivel profesional familiar	12,5	10,6	64	8 477 566
NSE 1	- Ingresos familiares - Máximo nivel educativo familiar - Máximo nivel profesional familiar	12,0	10,8	60	8 481 460
NSE 2	- Ingresos familiares - Promedio de nivel educativo familiar - Suma de los bienes	13,4	10,6	67	8 470 428
NSE 3	- Máximo nivel profesional familiar - Promedio de nivel educativo familiar - Suma de los bienes	13,2	11	66	8 472 236

Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009.

<sup>a</sup> Regresiones simples con las variables en el nivel individual.

<sup>b</sup> El porcentaje de datos que faltan se calcula sobre el total de registros con cuestionarios respondidos y el promedio general calculado (811.406).

<sup>c</sup> Para poder comparar la verosimilitud, se retiraron los datos que faltaban de todas esas variables, de manera que quedaron 704.481 registros con el promedio por colegio y los 4 componentes NSE en el nivel individual, y con ellos se rehicieron los modelos y se calcularon los criterios de información de Akaike (AIC) de cada uno.

NSE: indicador de nivel socioeconómico.

PISA: Programa de Evaluación Internacional de Alumnos.

CUADRO 3

**Brasil: datos longitudinales disponibles sobre los colegios en el Examen Nacional de Enseñanza Media (ENEM), 2009-2010**

	ENEM de 2009	ENEM de 2010
Datos con identificación del colegio	1 536 023	1 379 447
Datos con identificación del colegio (en porcentajes)	37,0	29,8
Número de colegios identificados	32 006	32 318
Número de colegios participantes en los dos años		28 010
Número de colegios válidos en los dos años <sup>a</sup>		17 359

Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009 y 2010.

<sup>a</sup> Según los criterios del INEP para el cálculo y la divulgación de los promedios por colegios.

# V

## Resultados

La primera cuestión relevante, sobre la base de las observaciones anteriores, consiste en si los colegios presentes en el ENEM constituyen una muestra representativa de los colegios brasileños. Más específicamente, habría que preguntarse si los últimos colegios de la clasificación del ENEM se podrían considerar los peores del país a la hora de lograr el ingreso de sus alumnos en la enseñanza superior. Observando la comparación con los microdatos del Censo Escolar de 2009 (véase el cuadro A.1 del anexo), los colegios que participan en el ENEM suelen estar en mejores condiciones que el promedio nacional, de manera que no constituyen una muestra particularmente representativa del total de colegios<sup>10</sup>.

<sup>10</sup> Este resultado y algunos otros presentados aquí se encuentran también en la tesis doctoral de Rodrigo Travitzki.

Es decir, la clasificación del ENEM abarca una muestra distorsionada de la totalidad de la enseñanza media brasileña, en la que se preseleccionan los colegios en mejores condiciones. Al parecer existe un problema de selección de la muestra también en relación con los colegios, probablemente relacionado con el carácter voluntario del examen. Este sería un argumento contrario al uso de la clasificación del ENEM como indicador de eficacia escolar en el Brasil en las políticas de responsabilidad. La misma divulgación de los promedios es una política de este tipo (aunque moderada), de manera que sería deseable investigar más a fondo otros argumentos, contrarios o favorables a esta divulgación.

Al comparar longitudinalmente los dos años, se percibe que hay cierta estabilidad en los resultados (véanse los cuadros 4 y 5), a pesar del carácter voluntario del examen. Tal estabilidad puede estar relacionada con

CUADRO 4

### Brasil: análisis descriptivo de los datos longitudinales de los colegios en el ENEM, 2009-2010<sup>a</sup>

	2009		2010	
	Promedio	Desviación estándar	Promedio	Desviación estándar
Colegios: nota general	534	56	537	54
Colegios: nota de la prueba objetiva	494	55	505	53
Colegios: nota de la redacción	575	65	570	61
Colegios: número de participantes en la prueba objetiva	45	45	55	53
Colegios: número de participantes en la redacción	44	44	54	53

*Fuente:* Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009 y 2010.

<sup>a</sup> Solo en los 17.359 colegios válidos en los dos años. Las cinco diferencias entre los años fueron estadísticamente significativas ( $p < 0,001$ ).

CUADRO 5

### Brasil: correlaciones entre los promedios de las notas correspondientes a los colegios en el ENEM y los valores agregados por los colegios, 2009-2010

Variable	Mismo método <sup>a</sup> , año diferente	Método diferente, año diferente	Método diferente, mismo año
Promedio bruto	0,84	0,84	1
Valor añadido	0,46	0,43	0,96
Pendiente del gradiente socioeconómico (GSE)	0,18	0,17	0,83

*Fuente:* Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009 y 2010.

<sup>a</sup> El método se refiere, en realidad, a un pequeño aspecto del método: el trío de variables introducido en el análisis de componentes principales para elaborar el componente de nivel socioeconómico (NSE).

su consolidación en la sociedad y en los colegios, así como con su uso para acceder a la enseñanza superior y, además, con la teoría de respuesta al ítem que comenzó a utilizarse en este ámbito en 2009<sup>11</sup>. También es posible que exista una relación entre la estabilidad de los promedios y factores externos al colegio o al examen, como se verá más adelante.

Los resultados de 2010 fueron ligeramente mejores que los de 2009 gracias a las notas de la prueba objetiva, ya que en la redacción se produjo una caída. ¿Qué significa esta diferencia: un valor añadido, una diferencia entre generaciones o una variación normal en indicadores de este tipo?

Por una parte, se podría considerar que se trata de un avance real en el resultado por colegios, y que la disparidad en las notas no se debe atribuir a una diferencia en el nivel de dificultad de los exámenes, ya que las preguntas se analizan previamente y las notas se calculan (en realidad, se estiman) según la función logística de tres parámetros de la teoría de respuesta al ítem, desarrollada por Birbaum en 1968 (INEP/MEC, 2011b). Sin embargo, la igualización de las escalas de habilidad solo se puede realizar adecuadamente si hay ítems comunes a las dos pruebas (Andrade, Tavares y Da Cunha Valle, 2000), algo que es inviable en un examen estandarizado e impreso que sirve para entrar en universidades buenas y gratuitas.

Por otra parte, se puede partir del principio de que es difícil que el colegio haya mejorado de hecho en solo un año. En ese sentido, sería fundamental que las medidas de la eficacia escolar basadas en exámenes fueran múltiples, es decir, que abarcasen más de un año. En los modelos de valor añadido longitudinal (relativos a cuánto mejora el alumno con el tiempo), “la mayoría de los autores recomiendan el uso de por lo menos tres medidas” (Martínez Arias, 2009, pág. 225).

## 1. Análisis multinivel del ENEM de 2009

El número total de registros en los microdatos del ENEM de 2009 es de 4.148.721 alumnos, de los cuales 2.218.191 respondieron al cuestionario socioeconómico y solo 1.339.445 estaban en el último curso de la enseñanza media, según sus propias declaraciones. Aplicando los criterios de validez del INEP, quedaron 811.406 registros

individuales para el análisis multinivel de la clasificación de colegios.

La concordancia en el nivel 1 (estados) fue de 0,77 y en el nivel 2 (colegios) fue de 0,82, lo que muestra que hay más consistencia entre las notas de alumnos del mismo colegio que entre las de alumnos del mismo estado, como era de esperar. A su vez, estos números sugieren que ambos niveles son significativos en el análisis, suposición que fue confirmada comparando la verosimilitud de modelos con y sin estas variables.

El coeficiente de correlación intraclase —para cada uno de los niveles separadamente— fue de 0,25 para los colegios y de 0,03 para los estados. Sin embargo, al calcular el coeficiente de correlación intraclase en el modelo de 3 niveles, la contribución del estado permaneció en un 3%, en tanto que la del colegio disminuyó al 22%, quedando aún el 75% de la varianza total para el nivel individual (véase el cuadro A.4 del anexo). Con esto se puede decir que un 3% del 25% inicialmente atribuido a las diferencias entre las escuelas se puede atribuir, en efecto, a la diferencia entre estados. En un estudio de regresión multinivel con resultados de otro examen brasileño, el Sistema de Evaluación de la Educación Básica (SAEB), se estimó que el porcentaje de variación de los resultados individuales que puede explicarse por el colegio es de un 39%, valor considerablemente mayor de lo que normalmente se encuentra en los países desarrollados (en torno del 20%) y que podría deberse a las grandes diferencias entre los colegios en el Brasil (Franco y otros, 2007). En ese sentido, el grado de diferencias entre los colegios incluidos en la clasificación del ENEM se parece más al de los países desarrollados que al general del Brasil, probablemente debido a las distintas características de las dos pruebas, ya que el SAEB se realiza por muestreo y su objetivo es representar a todos los colegios brasileños, mientras que el ENEM es voluntario y funciona como examen de entrada para la enseñanza superior.

A fin de investigar la influencia de las características de los alumnos y de los colegios en las notas del ENEM, se ajustaron diversos modelos multinivel utilizando el componente NSE y otras variables disponibles en los microdatos de la prueba y del cuestionario (véase el cuadro A.2 del anexo).

En el cuadro A.3 del anexo se presentan los coeficientes de esos modelos, permitiendo determinar en qué medida el contexto altera las características de las notas del alumno. Se observa que el efecto de la mayoría de las variables disminuye al introducir las dos variables de contexto NSE y apenas se altera con la introducción de las demás, lo que sugiere que la variable elaborada

<sup>11</sup> En algunos estudios se señala que la teoría de respuesta al ítem tiende a producir resultados más estables a lo largo del tiempo, en comparación con la teoría clásica de las pruebas o contrastes, utilizada hasta 2008 (Andrade, Tavares y Da Cunha Valle, 2000).

con la metodología del programa PISA de 2009 tiene un gran poder explicativo.

Entre las variables que más cambian en el ámbito de los colegios, se puede destacar la dependencia administrativa del colegio. En un primer análisis, los colegios privados parecen mucho mejores que los estatales, por ejemplo, pero la diferencia entre ellos disminuye bastante en la segunda columna. Los colegios federales, por su parte, siguen estando a bastante distancia de los demás una vez que se tiene en cuenta el contexto, lo que indica que son instituciones muy eficientes y capaces de producir buenos resultados incluso en circunstancias desfavorables.

Algo semejante ocurre con la proporción de personas de raza blanca, que es cinco veces menos importante después de tener en cuenta las condiciones de contexto. Sin embargo, si se considera que el color de la piel es una de las variables introducidas en el nivel individual, el hecho de que la proporción de estudiantes blancos en el colegio siga siendo significativa en el modelo completo, sugiere que hay una gran influencia de esta característica en cuanto a los resultados generados por la educación brasileña, conclusión que es corroborada por la relativa estabilidad de la influencia del color de la piel en el nivel individual, como puede verse en el cuadro A.3 del anexo.

A diferencia de la dependencia administrativa, el tipo de colegio parece ejercer una influencia menos vinculada al contexto. Los colegios para adultos, por ejemplo, presentan resultados inferiores a todos los demás en las tres columnas, con pequeñas variaciones. Esto sugiere que las diferencias entre los tipos de colegios son estructurales y el grado de comparación es pequeño, hecho que debería tenerse en cuenta en la forma de publicación de los resultados del ENEM por colegios.

En el cuadro A.3 del anexo se indica que el contexto influye más respecto del colegio que del individuo. Para verificar esta hipótesis se ajustó un modelo con las dos variables estandarizadas (*z-score*) y se obtuvo un coeficiente de 39 para el NSE del colegio y de 10 para el NSE individual.

En consecuencia, se puede decir que a una familia en condiciones menos favorables le valdría la pena llevar a sus hijos a un colegio al que asisten alumnos de contextos más favorables. Teniendo en cuenta que el presente análisis se basa en los resultados de un examen de selección para universidades gratuitas (y para acceder a becas en el caso de universidades privadas), esta conclusión resulta aún más consistente.

Dado el objetivo de investigar el efecto de las variables de contexto en la clasificación de colegios del

ENEM, resulta especialmente relevante la magnitud del poder explicativo del contexto en relación con la nota del alumno o, de forma más específica, el porcentaje de varianza explicada por los modelos. En tal sentido, los resultados de este estudio son significativos tanto para la discusión de las metodologías de valor añadido, como para la clasificación de colegios del ENEM propiamente dicha.

En el ámbito metodológico, corroborando conclusiones anteriores, se observa que el componente NSE, inspirado en el programa PISA, tiene un gran poder explicativo del resultado en los colegios (75%) y que la introducción de las demás variables aumentó la varianza explicada, pero relativamente poco, hasta alcanzar el 79% (véase el cuadro A.4 del anexo). Por lo tanto, parece razonable utilizar únicamente el componente NSE como control del contexto, si se desea realizar un análisis más práctico y con un mínimo de datos ausentes. Sin embargo, para el objetivo del presente estudio resulta más adecuado emplear el modelo completo, que se ajustó mejor a los datos según el análisis de verosimilitud.

En relación con el uso del ENEM como indicador de calidad escolar, en el cuadro A.4 del anexo se revelan algunas limitaciones significativas. Si bien el resultado respecto de los individuos sigue sin estar suficientemente explicado por las variables introducidas en el modelo, no ocurre lo mismo en relación con los colegios y de los estados. Eso significa que, como máximo, se puede atribuir al esfuerzo y al mérito de los colegios el 21% de la variación de las notas en el ámbito de estas instituciones, ya que ese es el porcentaje que no está determinado por factores más allá de su control. Si se realiza el mismo análisis separadamente para los dos “subpromedios”, se constata que este número disminuye al 13% en las notas de la prueba objetiva y aumenta al 38% en las notas de la redacción (véase el cuadro A.5 del anexo).

Aunque este resultado (inédito con los datos del ENEM) no constituya una novedad en el escenario de la investigación internacional, no deja de ser preocupante, en la medida en que este examen se ha ido consolidando año tras año como un relevante indicador de calidad escolar en el Brasil.

En otras palabras, el análisis de la varianza de los residuos en los diferentes niveles de los modelos jerárquicos sugiere que este examen individual podría evaluar los méritos del alumno, pero contiene poca información sobre el mérito de los colegios y los estados cuando se toman en cuenta las condiciones de contexto.

Cabe preguntar, finalmente, en qué medida los “resultados brutos” por colegio (los promedios divulgados anualmente) son diferentes de los “resultados netos”,

es decir, del valor añadido después de considerar las condiciones de contexto. Existe cierta correlación entre las dos notas ( $r = 0,51$   $p < 0,001$ ) que puede visualizarse en el gráfico 3, donde cada punto representa un colegio situado en el plano según dos medidas de calidad.

Simplificando la evaluación de los colegios como “buenos” o “malos” (sobre y bajo el promedio, respectivamente), en la mayoría de los casos (66%) los dos indicadores tuvieron resultados similares. Sin embargo, conviene resaltar que unos 7 mil colegios fueron evaluados de forma opuesta según cada uno de los criterios, cifra bastante significativa. Más específicamente, el 14% de los colegios se consideraron buenos según el promedio bruto, pero malos en términos de valor añadido, mientras que en el 20% de los casos sucedió lo contrario (véase el gráfico 3).

Esta última observación es especialmente problemática, ya que significa que una quinta parte de los colegios fueron evaluados como malos en la clasificación del ENEM, pero obtuvieron resultados por sobre lo esperado cuando se consideraron sus condiciones de contexto. Por

lo tanto, se trataba de colegios meritorios, probablemente en un contexto desfavorable, que podrían estar siendo subvalorados por la publicación anual de los promedios por colegios, divulgada en forma de clasificación ordenada por los grandes medios de comunicación. Este efecto probablemente se agrava debido a las diferencias encontradas entre los colegios incluidos en el ENEM y el promedio del Brasil. En ese sentido, la forma actual de esta política, dirigida a fomentar la calidad de la enseñanza escolar, podría estar provocando en muchos casos un efecto contrario al deseado.

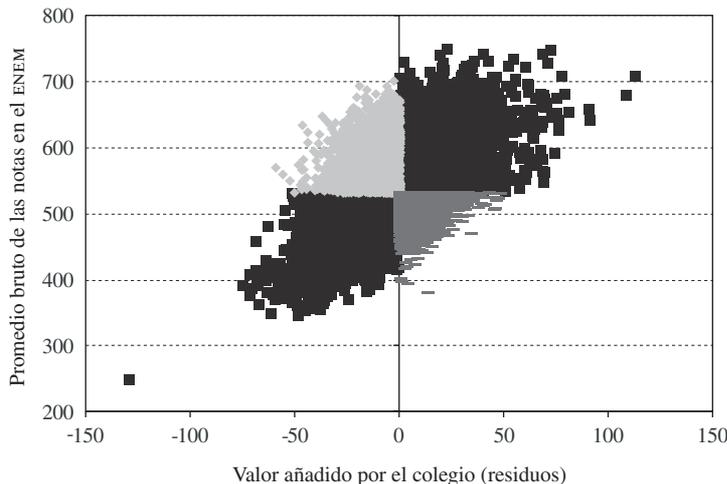
**2. Análisis longitudinal: 2009 y 2010**

Hasta aquí, se ha considerado el valor que los colegios supuestamente han añadido a los alumnos (en términos de los resultados en el ENEM brasileño), utilizando datos transversales. Para finalizar, se aplicará el otro concepto de valor añadido (véase el gráfico 2) a los microdatos del ENEM mediante información longitudinal, es decir, se averiguará cuánto valor se ha añadido a los colegios a lo

GRÁFICO 3

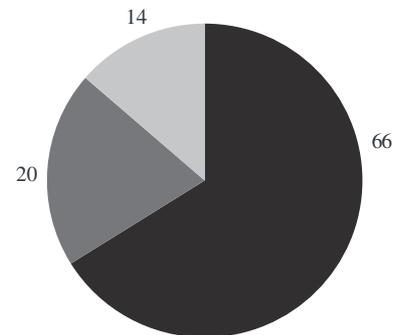
**Brasil: comparación del promedio bruto de las notas correspondientes a los colegios participantes en el ENEM y el valor añadido por esos colegios, 2009**

A. Comparación de la calidad de los colegios según el promedio bruto y el valor añadido



- Colegios con resultado similar según el promedio bruto y el valor añadido
- ◆ Colegios buenos según el promedio bruto, pero malos en términos de valor añadido
- Colegios malos según el promedio bruto, pero buenos en términos de valor añadido

B. Proporción de colegios con resultados similares o diferentes según el promedio bruto y el valor añadido (en porcentajes)



Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009.

largo de un cierto período. Dadas las limitaciones de los datos actuales (véase el cuadro 1), esta parte probablemente resulte más útil para el análisis metodológico que para la evaluación propiamente dicha de los colegios brasileños.

¿Cómo afecta la introducción de las variables de contexto a la estabilidad de los datos longitudinales? Se calculan las correlaciones entre los promedios brutos y “netos” (valor añadido) de los colegios en 2009 y 2010, utilizando dos métodos diferentes en 2009. Los resultados indican que hay una considerable estabilidad en los promedios brutos, la que se pierde cuando se controlan las condiciones de contexto (véase el cuadro 5). Esto puede tomarse como un argumento en favor de la publicación de los promedios brutos, tal como se hace actualmente, ya que esos promedios constituirían una medida razonablemente estable y robusta de la eficacia escolar. Sin embargo, teniendo en cuenta que el 79% de la variación de los resultados por colegios se explica por el contexto (véase el cuadro A.4 del anexo) y que la correlación entre el componente NSE de los colegios en 2009 y en 2010 es de 0,95, parece plausible concluir que la estabilidad de los promedios del ENEM se relaciona más con las condiciones del contexto que con el mérito de los colegios.

El cuadro 5 también permite ver que el uso de otras variables de contexto no arrojó resultados muy diferentes en los colegios, en contraste con lo que sucedió en otros estudios. Una posible explicación de esta estabilidad del valor añadido partiendo de diferentes variables consistiría en la inclusión de muchas variables en los modelos (véase el cuadro A.3 del anexo), lo que podría causar un efecto de grupo razonablemente resistente a la modificación de algunas de sus partes. A su vez, con respecto a los estados, el uso de diferentes métodos redujo la estabilidad del valor añadido (véase el cuadro 6), lo que demuestra la complejidad de este tipo de metodología.

En el cuadro 5 también se puede observar que la pendiente del GSE, es decir, la magnitud del cambio de la nota en función del componente NSE, varía mucho de un año a otro. Esto plantea algunas cuestiones sobre la fiabilidad de este tipo de medida en el ámbito de los colegios. Entre los estados, por otra parte, tanto la pendiente como el valor añadido presentan una gran estabilidad entre los dos años cuando se utiliza el mismo método (véase el cuadro 6).

Otra cuestión pertinente acerca de la estabilidad longitudinal de la clasificación del ENEM consiste en si esa estabilidad es homogénea en los diferentes estratos, es decir, entre colegios considerados buenos, medianos o malos. Al analizar separadamente los tres estratos, se percibe que la posición en la clasificación varía más entre los colegios considerados malos que entre los buenos (véase el cuadro A.6 del anexo), lo que podría constituir una segunda prueba<sup>12</sup> de que la clasificación del ENEM no es una fuente de información muy fiable para comparar los peores colegios. Por otra parte, es posible que sea una buena referencia para comparar los mejores colegios, pero más en lo atinente a la eficacia que al mérito de los colegios.

Entre los colegios medianos, se encontró una gran variedad en la posición en la clasificación y poca variación en el promedio, lo que a su vez apoya la idea de que podría existir una precisión excesiva en los promedios. Es decir, pequeñas variaciones en los promedios causan grandes cambios en la posición en la clasificación, sobre todo en el caso de los colegios medianos.

<sup>12</sup> La primera se deriva del cuadro A.1 del anexo.

CUADRO 6

**Brasil: correlaciones entre los promedios de las notas correspondientes a los estados en el ENEM y los valores añadidos por los estados, 2009-2010**

Variable	Mismo método <sup>a</sup> , año diferente	Método diferente, año diferente	Método diferente, mismo año
Promedio bruto	1	1	1
Valor añadido	0,93	0,60	0,66
Pendiente del gradiente socioeconómico (GSE)	0,91	0,85	0,95

Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009 y 2010.

<sup>a</sup> El método se refiere, en realidad, a un pequeño aspecto del método: el trío de variables introducido en el análisis de componentes principales para elaborar el componente de nivel socioeconómico (NSE).

# VI

## Conclusiones

Los resultados aquí obtenidos se pueden dividir en dos grupos, uno relativo a las metodologías de valor añadido y otro referente al examen propiamente dicho. En el primer caso, se investigaron los efectos producidos por la introducción de variables de contexto. En el segundo, se intentó identificar el tipo de información que proporciona la clasificación del ENEM a la sociedad brasileña.

En términos metodológicos, el análisis multinivel del ENEM de 2009 mostró que, aisladamente, la variable de contexto basada en el programa PISA tiene un gran poder explicativo de la nota de los alumnos, especialmente en lo que se refiere a los colegios (75%), mientras que la introducción de las demás variables solo aumentó la varianza explicada al 79%. Se descubrió que el contexto era cuatro veces más importante respecto de los colegios que de los individuos, mientras que entre los estados no resultó significativo en comparación con los otros dos. Al comparar el desempeño de los colegios en términos del promedio bruto y del valor añadido, el 34% de los resultados fueron contradictorios, es decir, el mérito de un tercio de las instituciones no se evaluó adecuadamente teniendo en cuenta los diferentes contextos.

El análisis longitudinal reveló que existe una estabilidad razonable en el promedio bruto entre los dos años ( $r = 0,8$ ), que se reduce al introducir las variables de contexto ( $r = 0,5$ ). La pendiente del GSE, a su vez, resultó bastante inestable ( $r = 0,2$ ), lo que plantea algunas cuestiones relativas a la fiabilidad de esta variable, especialmente cuando se realiza una única medida. En el ámbito de los estados, todos los indicadores se comportaron de manera estable entre 2009 y 2010.

Por otra parte, las diferencias de desempeño entre los colegios públicos y privados se redujeron bastante al considerar las condiciones de contexto, algo que no sucedió al comparar los colegios normales y los colegios para adultos. Las diferencias de desempeño entre diferentes “razas” (el color de la piel) también se mantuvieron después de considerar el efecto del contexto socioeconómico, si bien se necesitan más estudios a este respecto.

La clasificación del ENEM de 2009 abarca el 35% de los colegios de enseñanza media en el Brasil, con unas condiciones de infraestructura superiores al promedio nacional, debido a la finalidad y al carácter voluntario del examen. El análisis multinivel puso de manifiesto que el 3% de la variación en las notas puede

atribuirse al estado, el 22% a los colegios y el 75% a los alumnos. Las notas de la prueba objetiva sufrieron una mayor influencia del contexto (87%) que las notas de la redacción (62%), lo que puede significar que la redacción es más justa (en términos de méritos) o menos fiable que la primera. Es necesario realizar más estudios para investigar estas dos hipótesis. Al comparar los dos años, se percibe que la estabilidad de los promedios es mayor en los “mejores” colegios y que en las instituciones “medias” hay grandes variaciones en cuanto a la posición en la clasificación y pequeñas variaciones en cuanto al promedio.

Estos resultados indican que la clasificación del ENEM presenta serias limitaciones como indicador de calidad escolar a nivel nacional. Los peores colegios en esta clasificación no deberían considerarse los peores del Brasil y, entre los colegios “medios”, la diferencia en los promedios es muy pequeña. Los mejores colegios, por otra parte, presentan cierta estabilidad en el promedio bruto. Esta estabilidad puede significar que el examen es más informativo para los estratos superiores de la escala de habilidades, lo que sería comprensible dada su finalidad de selección para la enseñanza superior. Es necesario realizar otros estudios para verificar esta hipótesis.

En definitiva, ¿qué tipo de información proporciona a la sociedad brasileña la clasificación de colegios del ENEM? Los resultados de este estudio indican que, al limitarse al desempeño bruto, la clasificación representa más las condiciones socioeconómicas de los colegios que su mérito, teniendo en cuenta las diferencias de contexto. Esto se debe a que: i) el contexto puede explicar cuatro quintas partes de la varianza de las notas entre los colegios; ii) los promedios brutos son estables y el valor añadido es inestable, y iii) las condiciones de contexto de los colegios son aún más estables ( $r = 0,95$ ) en los dos años analizados.

En consecuencia, este indicador de calidad escolar podría ser utilizado apropiadamente por familias que deseen elegir una universidad para sus hijos y disfruten de buenas condiciones económicas. Sin embargo, para las familias menos favorecidas y para el Estado (en lo relativo a las políticas de responsabilidad y rendición de cuentas), la clasificación de colegios del ENEM es poco informativa e incluso engañosa. Otorgarle un valor excesivo podría provocar un incremento de las

desigualdades entre los colegios, ya que contribuiría a subvalorar instituciones que hacen un buen trabajo en condiciones precarias y a favorecer a aquellas que se dedican a estratos socioeconómicos altos de la sociedad

brasileña. Estas conclusiones confirman otros resultados de la literatura internacional, lo que apunta a la necesidad de crear otros indicadores de calidad escolar en los países democráticos.

## ANEXO

## CUADRO A.1

**Brasil: comparación de los colegios presentes en el ENEM con el total de los colegios, 2009**

(En número de colegios y porcentajes)

Característica	Colegios en la clasificación del ENEM	Todos los colegios de enseñanza secundaria <sup>a</sup>	Todos los colegios
Número de colegios <sup>b</sup>	18 605	30 554	255 445
Urbanos <sup>c</sup>	97,2	92,1	53,0
Privados (en porcentajes) <sup>c</sup>	24,6	30,9	19,8
Red pública de agua <sup>c</sup>	93,5	90,3	64,1
Red pública de alcantarillado (en porcentajes) <sup>c</sup>	69,0	66,2	40,7
Sala de informática <sup>c</sup>	90,3	81,8	23,2
Laboratorio de ciencias <sup>c</sup>	53,2	43,4	7,2
Cancha de deportes <sup>c</sup>	80,1	68,6	21,5
Biblioteca (en porcentajes) <sup>c</sup>	75,1	71,3	25,3
Fotocopiadora <sup>c</sup>	67,6	63,5	30,7
Internet de banda ancha <sup>c</sup>	76,5	71,4	32,2

Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) y del Censo Escolar de 2009.

<sup>a</sup> Solo de enseñanza regular y de enseñanza de jóvenes y adultos.

<sup>b</sup> Número de colegios.

<sup>c</sup> En porcentajes.

## CUADRO A.2

**Brasil: variables incluidas en el modelo completo utilizando los datos del ENEM de 2009**

Variable	Nivel <sup>a</sup>	Tipo
Promedio	0	Numérica
Color de la piel	0	Catégorica
Sexo	0	Catégorica
Religión	0	Catégorica
Componente de nivel socioeconómico (NSE) individual	0	Numérica
Componente de nivel socioeconómico (NSE) del colegio	2	Numérica
Dependencia administrativa	2	Catégorica
Modalidad	2	Catégorica
Proporción de alumnos blancos	2	Numérica
Proporción de alumnos que realizaron el curso preparatorio	2	Numérica

Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009.

<sup>a</sup> Los niveles están numerados de lo más general a lo más particular (1 = estado, 2 = colegio) y el nivel 0 equivale al individuo.

CUADRO A.3

**Brasil: efectos de la introducción de las variables de contexto utilizando los datos del ENEM de 2009**

	Modelos univariados	Modelos con componentes de nivel socioeconómico (NSE) en los dos niveles	Modelo completo
<b>Componente de nivel socioeconómico (NSE)</b>			
NSE (individual)	10	8	8
NSE (promedio del colegio)	41	36	22
<b>Nivel del individuo</b>			
Sexo [femenino]	0	0	0
Sexo [masculino]	-15	-17	-17
Color [blanco]	0	0	0
Color [pardo]	-9	-7	-6
Color [negro]	-16	-13	-11
Color [amarillo]	-11	-10	-10
Color [indígena]	-34	-31	-29
Religión [católica]	0	0	0
Religión [protestante/evangélica]	7	7	9
Religión [espiritismo]	13	9	10
Religión [umbanda/candomblé]	-9	-13	-8
Religión [otra]	13	11	14
Religión [sin religión]	17	14	18
<b>Nivel del colegio</b>			
Dependencia administrativa [federal]	0	0	0
Dependencia administrativa [estatal]	-108	-65	-66
Dependencia administrativa [municipal]	-97	-57	-57
Dependencia administrativa [privado]	-17	-39	-41
Tipo [regular]	0	0	0
Tipo [jóvenes y adultos]	-48	-42	-42
Proporción de alumnos blancos	162	41	30
Proporción de alumnos que realizaron el curso preparatorio	109	8	8

*Fuente:* Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009.

Nota: los ceros en estas categorías son, en realidad, referencias de las variables categóricas, mientras que los coeficientes de las otras categorías son relativos a las primeras.

Coefficientes de diversos modelos ( $p < 0,01$ ), teniendo como variable de respuesta la nota del alumno. En la primera columna, las variables explicativas están solas, en la segunda están acompañadas del componente NSE respecto de los individuos y de los colegios y en la tercera columna se muestra el modelo compuesto por todas las variables juntas. Como las variables no están estandarizadas, las comparaciones deben realizarse horizontalmente. Las únicas comparaciones verticales que tienen sentido son las efectuadas entre factores de la misma variable categórica (identificable mediante los corchetes).

CUADRO A.4

**Brasil: varianza de los residuos y varianza explicada en los resultados del ENEM de 2009**

	Intercepto del nivel 1 (estado)	Intercepto del nivel 2 (colegio)	Residuos individuales
Varianza en el modelo 0	356	2 507	8 482
Varianza en el modelo 1 (solo variables del nivel individual)	197	1 658	8 191
Varianza en el modelo 2 (NSE en los dos niveles)	121	637	8 305
Varianza en el modelo 3 (completo)	74	529	8 129
Varianza dentro del modelo 0	0,03	0,22	0,75
Varianza explicada del modelo 1	0,45	0,34	0,03
Varianza explicada del modelo 2	0,66	0,75	0,02
Varianza explicada del modelo 3	0,79	0,79	0,04

*Fuente:* Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009.

CUADRO A.5

**Brasil: varianza de los residuos de la prueba objetiva y de redacción del ENEM de 2009**

(En valores absolutos y porcentajes)

	Intercepto del nivel 1 (estado)	Intercepto del nivel 2 (colegio)	Residuos individuales
Nota de la prueba objetivo (modelo nulo)	453	2 500	4 133
Nota de la redacción (modelo nulo)	348	3 064	23 497
Nota de la prueba objetiva (modelo completo)	88	312	3 929
Nota de la redacción (modelo completo)	240	1 160	22 413
Varianza explicada de la prueba objetiva (en porcentajes)	80,6	87,5	4,9
Varianza explicada de la redacción (en porcentajes)	31,0	62,1	4,6

Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009.

CUADRO A.6

**Brasil: desviaciones estándar de las diferencias de los resultados en el ENEM, por estratos, 2009-2010**

	Desviación estándar de la diferencia entre los promedios brutos	Desviación estándar de la diferencia entre las posiciones en la clasificación
Dos mil mejores en 2009	26,6	1 055
Dos mil medianos en 2009	25,8	3 300
Dos mil peores en 2009	32,4	2 803

Fuente: Instituto Nacional de Estudios e Investigaciones Educativas (INEP), microdatos del Examen Nacional de Enseñanza Media (ENEM) de 2009 y 2010.

*Bibliografía*

- Andrade, E.C. (2008), "'School accountability' no Brasil: experiências e dificuldades", *Revista de Economia Política*, vol. 28, N° 3, São Paulo, Centro de Economía Política.
- Andrade, D.F., H.R. Tavares y R. da Cunha Valle (2000), *Teoria da resposta ao item: conceitos e aplicações*, São Paulo, Associação Brasileira de Estadística (ABE).
- Apple, M.W. (2001), "Comparing neo-liberal projects and inequality in education", *Comparative Education*, vol. 37, N° 4, Taylor & Francis.
- Bliese, P. (2012), "Multilevel: Multilevel Functions. R package version 2.4" [en línea] <http://cran.r-project.org/package=multilevel>.
- Brandão, Z. (2000), "Fluxos escolares e efeitos agregados pelas escolas", *Em Aberto*, vol. 17, N° 71, Brasília, Instituto Nacional de Estudios e Investigaciones Educativas (INEP).
- Dubet, F. (2004), "O que é uma escola justa?", *Cadernos de Pesquisa*, vol. 34, N° 123.
- Ferrão, M.E. (2009), "Sensibilidad de las especificaciones de los modelos de valor añadido: Midiendo el estatus socioeconómico", *Revista de Educación*, N° 348, Madrid, Ministerio de Educación, Cultura y Deporte.
- Franco, C. y otros (2007), "Qualidade e equidade em educação: reconsiderando o significado de 'fatores intra-escolares'", *Ensaio*, vol. 15, N° 55, Río de Janeiro.
- Gobierno de España (2006), "Ley orgánica 2/2006, de 3 de mayo, de educación" [en línea] [http://www.boe.es/boe/consultas/bases\\_datos/doc.php?id=BOE-A-2006-7899](http://www.boe.es/boe/consultas/bases_datos/doc.php?id=BOE-A-2006-7899).
- Heck, R.H. (2000), "Examining the impact of school quality on school outcomes and improvement: a value-added approach", *Educational Administration Quarterly*, vol. 36, N° 4, SAGE.
- Heckman, J.J. (1976), "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models", *Annals of Economic and Social Measurement*, vol. 5, N° 4, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).
- INEP/MEC (Instituto Nacional de Estudios e Investigaciones Educativas/Ministerio de Educación) (2011a), "Nota técnica. Médias do ENEM 2010 por escola" [en línea] <http://inep.gov.br/>.
- \_\_\_\_\_ (2011b), "Nota técnica. Procedimento de cálculo das notas do Enem" [en línea] <http://inep.gov.br/>.
- \_\_\_\_\_ (2010), "Nota técnica. Médias do ENEM 2009 por escola" [en línea] <http://inep.gov.br/>.
- \_\_\_\_\_ (2007), "Nota técnica. Notas médias do Enem 2006 por município e por escola dos alunos concluintes do ensino médio em 2006" [en línea] <http://inep.gov.br/>.
- Karsten, S., A. Visscher y T. de Jong (2001), "Another side to the coin: the unintended effects of the publication of school performance data in England and France", *Comparative Education*, vol. 37, N° 2, Taylor & Francis.
- Keeves, J.P., N. Hungi y T. Afrassa (2005), "Measuring value added effects across schools: should schools be compared in performance?", *Studies in Educational Evaluation*, vol. 31, N° 2-3, Amsterdam, Elsevier.
- Ladd, H.F. y R.P. Walsh (2002), "Implementing value-added measures of school effectiveness: getting the incentives right", *Economics of Education Review*, vol. 21, N° 1, Amsterdam, Elsevier.
- Linn, R.L. y C. Haug (2002), "Stability of school-building accountability scores and gains", *Educational Evaluation and Policy Analysis*, vol. 24, N° 1, Washington, D.C., American Educational Research Association.
- Luyten, H. (1994), "Stability of school effects in Dutch secondary education: the impact of variance across subjects and years", *International Journal of Educational Research*, vol. 21, N° 2.

- Mandeville, G.K. (1988), "School effectiveness indices revisited: cross-year stability", *Journal of Educational Measurement*, vol. 25, N° 4, Wiley.
- Martínez Arias, R. (2009), "Usos, aplicaciones y problemas de los modelos de valor añadido en educación", *Revista de Educación*, N° 348, Madrid, Ministerio de Educación, Cultura y Deporte.
- Martínez Arias, R., J.L. Gaviria Soto y M. Castro Morera (2009), "Concepto y evolución de los modelos de valor añadido en educación", *Revista de Educación*, N° 348, Madrid, Ministerio de Educación, Cultura y Deporte.
- McCaffrey, D.F. y otros (2004), "Models for value-added modeling of teacher effects", *Journal of Educational and Behavioral Statistics*, vol. 29, N° 1, SAGE.
- MEC (Ministerio de Educación) (2002), *Examen Nacional de Ensimó Médio (ENEM): documento básico 2002*, Brasilia.
- Meyer, R.H. (1997), "Value-added indicators of school performance: a primer", *Economics of Education Review*, vol. 16, N° 3, Amsterdam, Elsevier.
- Murillo Torrecilla, F.J. (2005), *La investigación sobre eficacia escolar*, Barcelona, Octaedro.
- OCDE (Organización de Cooperación y Desarrollo Económicos) (2010), *PISA 2009 Results. Overcoming Social Background: Equity in Learning Opportunities and Outcomes (Volume II)*, París, OECD Publishing.
- Oliveira, R.P. de y G.C. Araujo (2005), "Qualidade do ensino: uma nova dimensão da luta pelo direito à educação", *Revista Brasileira de Educação*, N° 28, Río de Janeiro.
- Passarinho, N. (2012), "MEC vai substituir a Prova Brasil pelo Enem em cálculo do Ideb, diz ministro" [en línea] <http://g1.globo.com/vestibular-e-educacao/noticia/2012/08/mec-vai-substituir-prova-brasil-pelo-enem-em-calculo-do-ideb-diz-ministro.html>.
- Raudenbush, S.W. y A.S. Bryk (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Thousand Oaks, SAGE Publications.
- Raudenbush, S.W. y J.D. Willms (1995), "The estimation of school effects", *Journal of Educational and Behavioral Statistics*, vol. 20, N° 4, Washington, D.C., American Educational Research Association.
- Ray, A., H. Evans y T. McCormack (2009), "El uso de los modelos nacionales de valor añadido para la mejora de las escuelas británicas", *Revista de Educación*, N° 348, Madrid, Ministerio de Educación, Cultura y Deporte.
- Reckase, M.D. (2004), "The real world is more complicated than we would like", *Journal of Educational and Behavioral Statistics*, vol. 29, N° 1, Washington, D.C., American Educational Research Association.
- Saunders, L. (1999), "A brief history of educational 'value added': how did we get to where we are?", *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, vol. 10, N° 2, Taylor & Francis.
- Tekwe, C.D. y otros (2004), "An empirical comparison of statistical models for value added assessment of school performance", *Journal of Educational and Behavioral Statistics*, vol. 29, N° 1, Washington, D.C., American Educational Research Association.
- Thomas, S. (1998), "Value-added measures of school effectiveness in the United Kingdom", *Prospects*, vol. 28, N° 1, Springer.
- Thomas, S. y otros (1997), "Stability and consistency in secondary schools' effects on students' GCSE outcomes over three years", *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, vol. 8, N° 2, Taylor & Francis.
- Van Petegem, P. y otros (2005), "Publishing information on individual schools?", *Educational Research and Evaluation: An International Journal on Theory and Practice*, vol. 11, N° 1, Taylor & Francis.
- West, A. y H. Pennell (2000), "Publishing school examination results in England: incentives and consequences", *Educational Studies*, vol. 26, N° 4, Taylor & Francis.
- Willms, J.D. (2006), *Learning Divides: Ten Policy Questions About the Performance and Equity of Schools and Schooling Systems*, Montreal, Instituto de Estadística de la UNESCO.
- Zhang, X. e Y. Zha (2010), "On 'abandoning examination' phenomenon of Gaokao", [en línea] [http://en.cnki.com.cn/Article\\_en/CJFDTOTAL-HJKS201012007.htm](http://en.cnki.com.cn/Article_en/CJFDTOTAL-HJKS201012007.htm).