

## DETERMINANTES DEL LOGRO ESCOLAR EN MÉXICO

Primeros resultados utilizando la prueba  
ENLACE media superior\*

*Rafael E. de Hoyos, Juan Manuel Espino  
y Vicente García\*\**

### RESUMEN

Aunque los años de escolaridad de los mexicanos aumentaron significativamente en los pasados 20 años, la calidad del sistema educativo —un determinante importante del crecimiento económico de largo plazo— dista mucho de ser satisfactorio. Utilizando los resultados de la primera prueba estandarizada en educación media superior (ENLACE-MS), este estudio muestra que un modelo que incluye las características del hogar, individuales, institucionales y escolares puede explicar más de 40% de las diferencias observadas en las calificaciones de matemáticas. Desde el punto de vista de la política pública, los resultados de este trabajo sugieren que, a fin de aumentar la calidad de la educación, las autoridades educativas deberían concentrarse en mejorar la infraestructura escolar, la calidad de los docentes y la relación entre los estudiantes y las autoridades escolares.

### ABSTRACT

Although Mexico had a significant increase in years of schooling in the last 20 years, the quality of its education system —an important determinant of long term growth— is far from being satisfactory. Using the results from the first standardized

\* *Palabras clave:* determinantes, logro escolar, México. *Clasificación JEL:* I21, I28. Artículo recibido el 5 de septiembre de 2010 y aceptado el 10 de diciembre de 2011.

\*\* R. E. de Hoyos, economista senior, Banco Mundial (correo electrónico: rdehoyos@worldbank.org). J. M. Espino, jefe de división, IMSS (correo electrónico: juan.espinob@imss.gob.mx). V. García, pasante de doctorado de la Universidad de Columbia, Nueva York.

test in upper middle school (ENLACE MS), the paper shows that a model including household, personal, institutional, and school characteristics can account for more than 40 percent of the total differences observed in math scores. From a policy perspective, the paper's results suggest that education authorities interested in fostering education quality should concentrate in improving school infrastructure, the quality of its teachers and the relationship between students and school authorities.

## INTRODUCCIÓN

**L**os resultados de estudios empíricos demuestran que una sociedad más escolarizada obtiene beneficios en términos de productividad e ingreso (Becker, 1962; Mincer, 1970), crecimiento económico (Denison, 1964; Hanushek *et al*, 2008), equidad en la distribución de ingresos (De Gregorio y Lee, 2002), salud (Lleras-Muney, 2005), democracia y participación ciudadana (Glaeser *et al*, 2007), entre otros. Pero, ¿a qué nos referimos por educación? La escolaridad de un país puede medirse como la escolaridad promedio o el logro cognitivo alcanzado en pruebas estandarizadas. Según Hanushek (2008), el crecimiento económico está asociado al logro cognitivo de la población y no a sus años de escolaridad. En consecuencia, para la correcta elaboración de programas e intervenciones que puedan incrementar la escolaridad de un país, resulta imprescindible identificar los factores que determinan el logro cognitivo.

En México el desempeño en términos de años de escolaridad ha sido satisfactorio; por ejemplo, la escolaridad promedio de la población de 15 años y más de edad pasó de 2.6 años en 1960 a 8.1 años en 2005. Asimismo, entre 2000 y 2005, el porcentaje de la población entre 15 y 29 años que tenía educación secundaria o más, pasó de 60.1 a 68.9%.<sup>1</sup> Sin embargo, en términos de calidad educativa, pruebas estandarizadas como PISA muestran que, en el periodo 2000-2006, no ha habido avances y los niveles actuales son insatisfactorios (Díaz *et al*, 2007). Dado el pobre desempeño que han mostrado los estudiantes mexicanos en términos de logro cognitivo, este estudio centra su análisis en los determinantes de este importante factor de desarrollo.

En general, la bibliografía identifica cuatro grandes componentes en los cuales pueden clasificarse los factores asociados al logro cognitivo: *i*) las características personales del estudiante (incluyendo los antecedentes esco-

<sup>1</sup> Cálculos propios con base en INEGI, *XII Censo General de Población y Vivienda 2000* e INEGI, *II Conteo de Población y Vivienda 2005*.

lares); *ii*) los antecedentes familiares y el ambiente del hogar del estudiante; *iii*) los factores asociados a la escuela, y *iv*) los factores institucionales (Hanushek, 1986; Fuchs y Woessmann, 2004; Glewwe y Kremer, 2006; Todd y Wolpin, 2007; Berhman, por aparecer). El objetivo del presente artículo es identificar los determinantes del logro escolar (o cognitivo) en el nivel medio superior en México. En particular, el estudio cuantifica la importancia de los cuatro componentes enumerados líneas arriba para explicar la variación total en logro escolar medido por medio del desempeño en la Evaluación Nacional del Logro Académico en Centros Escolares de Educación Media Superior de 2008 (ENLACE-MS).<sup>2</sup>

ENLACE es la primera prueba censal estandarizada que mide las capacidades lectora y matemática de cada estudiante en el último grado de educación media superior.<sup>3</sup> Los resultados de ENLACE están complementados por tres cuestionarios de contexto aplicados, respectivamente, a alumnos, docentes y directores de planteles seleccionados en una muestra.<sup>4</sup> Este es el primer estudio que identifica y cuantifica la importancia de los distintos determinantes del logro escolar en educación media superior en México.<sup>5</sup>

Este artículo no pretende contribuir metodológicamente al conocimiento de los determinantes del logro escolar; más bien utiliza métodos ortodoxos para identificar los factores más importantes asociados a éste. La contribución del estudio consiste en presentar, de manera sintetizada, la importancia de los grandes componentes relacionados con el logro académico. Esto tiene la doble finalidad de servir como guía para la toma de decisiones de política educativa e iniciar un debate académico de los determinantes del logro cognitivo en educación en México. El principal resultado de este artículo es que las características individuales del alumno (en particular los

<sup>2</sup> A lo largo de todo el documento, cuando hablamos de ENLACE nos referimos a ENLACE Media Superior.

<sup>3</sup> La prueba ENLACE no permite derivar resultados por escuela o subsistema de educación media superior. Ello porque en varias entidades federativas y subsistemas, tanto públicos como privados, existen concursos de selección de ingreso (*cream skimming*). En consecuencia, el buen desempeño en la prueba de una escuela podría ser consecuencia de que su buena reputación le permitió seleccionar a los mejores alumnos, lo cual no implica forzosamente que dicha escuela sea eficiente.

<sup>4</sup> Véase más información de ENLACE Media Superior en: <http://enlace.sep.gob.mx/ms/>. Véase los contenidos de los cuestionarios de contexto ENLACE en Serrano (2009).

<sup>5</sup> El estudio se centra en los determinantes del logro de los alumnos que llegaron al último grado de educación media superior, lo cual implica que se excluye a aquellos que abandonaron la escolaridad antes de la fecha del examen. Cabe suponer que los desertores tendrían un puntaje promedio menor a los que sí presentaron el examen. La tasa de deserción en la educación media superior durante el ciclo escolar 2007-2008 fue de 16.3 por ciento.

antecedentes escolares) explican una cuarta parte de la variación total en el desempeño de la prueba ENLACE, que constituye el principal componente. Los otros componentes son, en orden de importancia, los recursos escolares (que explica casi 8% de la variación), los factores institucionales (6%) y, por último, los antecedentes familiares (2.5 por ciento).

Desde el punto de vista de la autoridad educativa, los factores dentro del componente recursos escolares son vistos como los factores en los cuales la autoridad educativa puede incidir. Los resultados de este estudio muestran que el número de cursos y las horas impartidas de matemáticas tiene un efecto positivo en el logro en la prueba ENLACE. Encontramos que factores asociados al tipo de infraestructura como el número de laboratorios, el tamaño del plantel, las computadoras por alumno y el número de alumnos por aula tienen un efecto significativo en el resultado de ENLACE. El modelo nos dice que los alumnos en planteles con una mayor proporción de docentes titulados y cuyo director fue asignado vía un concurso de oposición tienden a obtener mejores resultados. Por último, la relación entre alumnos y entre éstos y los docentes y directivos es un determinante importante del logro escolar en el estrato medio superior.

El estudio está organizado de la siguiente manera: la sección I presenta la revisión de la bibliografía de determinantes del logro cognitivo de la cual se desprende un marco conceptual. Este marco conceptual es utilizado en las secciones II y III para desarrollar la metodología e interpretar los resultados, respectivamente. Se concluye delineando un programa de investigación para el futuro.

## I. BIBLIOGRAFÍA DE LOS DETERMINANTES DEL LOGRO

El análisis de los determinantes del logro cognitivo comenzó con el llamado Informe Coleman (*Coleman et al.*, 1966), que es un estudio precursor realizado por el gobierno de los Estados Unidos, que tuvo un alto efecto en la política educativa de ese país y cuya principal contribución fue ofrecer una explicación de las diferencias en el logro educativo entre los alumnos de distintas escuelas. En ese estudio se llegó a la conclusión de que la familia es el factor más importante en la determinación del logro, seguido por la influencia de los compañeros de escuela. Contrario a lo que esperaríamos *ex ante*, el citado estudio encuentra que los recursos escolares tenían un efecto moderado en el logro de los estudiantes.

El Informe Coleman estimuló el desarrollo de una vasta bibliografía empírica de los determinantes del desempeño escolar o logro cognitivo. Por lo que se refiere al campo de la economía de la educación, existen centenares de investigaciones acerca del tema, muchas de las cuales estiman funciones de producción educativa. La ventaja de este tipo de modelos es que permite establecer una relación estadística entre los insumos que intervienen en el desarrollo cognitivo y su resultado, de manera análoga a una función de producción que relaciona los insumos con el producto. El propósito de la investigación empírica es identificar la combinación de insumos que, dada cierta tecnología, maximiza el logro educativo.

La mayoría de las estimaciones de funciones de producción educativa enfoca su atención en un grupo relativamente reducido de insumos, lo que permite simplificar los resultados. De esta manera, los insumos que se incluyen en la función de producción educativa pueden agruparse en tres categorías generales: *i*) los que corresponden a las características personales del estudiante (incluyendo los antecedentes escolares); *ii*) los antecedentes familiares y los insumos relevantes que el estudiante recibe en el hogar, y *iii*) los recursos escolares a los que tiene acceso el estudiante. Más recientemente algunos autores han agregado una cuarta categoría que busca medir el efecto del desarrollo institucional del sistema educativo (Woessmann, 2003; Bishop y Woessmann, 2004) o de la localidad en la cual está inmersa la escuela y el hogar de los alumnos (Mayer y Serván, 2009; Urquiola y Verhoogen, 2009).

Entre las características personales del alumno, el origen étnico, el género y la edad de los alumnos son las variables que han recibido la mayor atención en la bibliografía (Goldhaber y Brewer, 1997; Fuchs y Woessmann, 2007; Todd y Wolpin, 2007). En cuanto a los antecedentes familiares del estudiante, la escolaridad de los padres y la presencia de insumos educativos en el hogar, como libros o acceso a computadoras, suelen utilizarse como variables exógenas en la función de producción educativa (Fuchs y Woessmann, 2007, Todd y Wolpin, 2007). En cuanto a los recursos escolares, los estudios suelen incluir variables de los antecedentes de los maestros (años de experiencia y escolaridad) (Card y Krueger, 1992; Hanushek, Kain y Rivkin, 1998); algún tipo de medida de lo que podría considerarse como “hacinamiento escolar” o regla de Maimónides (la proporción alumnos/maestro y el tamaño de la escuela o del grupo) (Goldhaber y Brewer, 1997; Bacolod y Tobías, 2006), así como indicadores del equipamiento de la

escuela (disponibilidad de materiales educativos o computadoras) (Hanushek, 2003; Bacolod y Tobías, 2009). Entre las variables institucionales se han incluido variables proxy que buscan medir el poder de los sindicatos de maestros (su influencia en el nombramiento de docentes) (Álvarez, García y Patrinos, 2007); indicadores del grado de descentralización en el ejercicio de los recursos y la toma de decisiones (Álvarez, García y Patrinos, 2007), o el grado de autonomía escolar y la existencia de exámenes estandarizados (Fuchs y Woessmann, 2007). Como factores institucionales atribuibles a la localidad se han incluido diversas variables que captan información del entorno socioeconómico en el que se inserta la escuela y el hogar, como el índice de marginación (Mayer y Serván, 2009) o el tamaño de la localidad (Álvarez, García y Patrinos, 2007).

A pesar del relativamente reducido número de variables que se han considerado como insumos en la función de producción educativa y el gran número de artículos que registra la bibliografía, no existe un consenso respecto a cuáles insumos son pertinentes ni la magnitud de su efecto (Todd y Wolpin, 2003). Por ejemplo, Hanushek (1996, 2003) resume en un meta análisis los hallazgos de los diferentes estudios de estimaciones de función producción y concluye que existe pocas pruebas empíricas que demuestren que los insumos escolares están asociados al logro escolar. Krueger (2003) no coincide con esta conclusión y señala que la muestra de artículos analizados por Hanushek están sesgados hacia su conclusión. Para Todd y Wolpin (2003) la falta de consenso en los resultados está determinada porque, en una buena parte de los trabajos, las ecuaciones por estimar no tienen suficiente estructura con sustento teórico.

En ese sentido, una de las críticas más relevantes a los modelos de estimación de funciones de producción es que el logro cognitivo está en función de los insumos pertinentes durante toda la vida de una persona y no sólo de los insumos actuales. Ante ello, los modelos más recientes utilizan la forma de una función de “valor agregado”, en lugar de la tradicional estimación de la producción; con la desventaja de que estos modelos imponen grandes supuestos en la función de producción subyacente (Todd y Wolpin, 2003, 2007).

Otra innovación metodológica en el estudio empírico de los determinantes del logro educativo se ha dado en el análisis de regresión, con la estimación de modelos multinivel, también conocidos como modelos jerárquicos lineales. Estos modelos consideran la naturaleza jerárquica de los datos, es decir, que los estudiantes están anidados en escuelas. La ventaja de esta me-

todología es que permite la estimación simultánea del efecto de variables en el estudiante y el de variables de la escuela.

En México existen pocos estudios de los determinantes del logro cognitivo. El trabajo precursor en este tema en nuestro país es el de Schmelkes *et al* (1997) en el cual se realizó un análisis de regresión de los resultados de una prueba de competencias en educación básica (primaria y secundaria) realizada en cinco regiones de Puebla. Posteriormente, Álvarez, García y Patriños (2007) estimaron tres funciones de producción educativa para México con los puntajes obtenidos por los jóvenes de 15 años de edad en la prueba PISA 2003 en matemáticas, ciencias y lectura. En ese estudio se incluyeron como variables independientes tres insumos: los asociados con el estudiante, que incluyen indicadores de carácter socioeconómico; los asociados con la escuela, con variables como recursos y características de la escuela, así como variables institucionales que fueron recabadas en cada estado, en las cuales se incluye la descentralización administrativa y pedagógica dentro del estado, la rendición de cuentas, la influencia del sindicato en la contratación de docentes y los conflictos entre el estado y el sindicato. La principal conclusión de ese trabajo es que la rendición de cuentas y la evaluación son necesarias para mejorar los resultados en PISA. Por su parte, Mayer y Serván (2009), estimando una función de producción estocástica con datos de la Encuesta Nacional sobre Niveles de Vida de los Hogares 2002, llegaron a la conclusión de que el logro cognitivo de los niños está en función de los antecedentes familiares y las características de la localidad.

En cuanto a los análisis multinivel en México, Backhoff *et al* (2007) realizaron un estudio para los alumnos de 6º y 9º grado de educación básica en México, con los datos de los Exámenes de la Calidad y el Logro Educativos (Excale) y sus correspondientes cuestionarios de contexto. El principal hallazgo es que la varianza de los resultados se explica más por las diferencias entre los alumnos (65% de la varianza) que por las diferencias entre las escuelas (35% de la varianza) y, en el caso de matemáticas en secundaria, los porcentajes cambian a 75 y 25%, respectivamente. Finalmente, Tristán *et al* (2008) realizaron un análisis multinivel con los resultados de ciencias en PISA 2006, tomando como marco teórico al modelo de calidad educativa desarrollado por el Instituto Nacional para la Evaluación de la Educación (INEE). Los tres estratos de anidamiento en el modelo jerárquico que estimaron son: estudiante, escuela y región. En el estudio se analizan más de 70 variables explicativas, las cuales son identificadas por medio del análisis factorial. Este

estudio concluye que, en el nivel de estudiante, existe un alto porcentaje de varianza que no es explicado por ninguna de las variables elegidas, a pesar de la abundancia de variables incluidas en el modelo. Por lo que se refiere a la escuela, los indicadores que inciden en la equidad son los que explican el mayor porcentaje de varianza. Por región, en cambio, el 99.9% de la varianza es explicada por las variables relacionadas con la eficacia.<sup>6</sup>

La revisión de la bibliografía muestra que no existe consenso respecto a cuáles son los factores que tienen una mayor importancia en la determinación del logro cognitivo. Más aún, en el caso de México, existe escasa evidencia empírica acumulada y sólo en educación básica, de la cual se infiere que las características individuales, los antecedentes familiares y las características institucionales están incidiendo en el logro cognitivo, sin que haya información concluyente de cuál de estos factores es el más relevante. Por último, la revisión de la bibliografía muestra que, en el mejor de los casos, los modelos actuales logran explicar 39% (Patrinos y García, 2009) de la variación total en logro escolar entre alumnos.

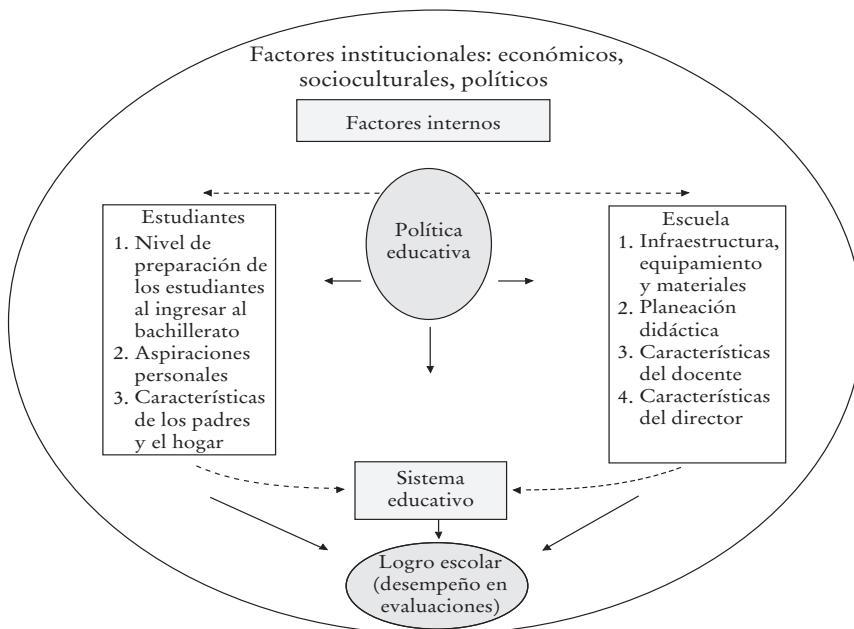
## II. MARCO CONCEPTUAL Y METODOLÓGICO

### *1. Marco conceptual*

El primer desafío del presente estudio consiste en esbozar un marco conceptual que nos permita analizar y organizar la información de los logros y sus posibles determinantes. La revisión de la bibliografía proporciona elementos que nos permiten conjeturar cómo se desarrolla el proceso de aprendizaje y por tanto qué elementos pueden incidir en éste. El diagrama 1 muestra de manera simplificada los factores que inciden en el proceso de aprendizaje. El aprendizaje es resultado de la interacción entre estudiantes —con determinadas características— y escuelas —con ciertos recursos—. La interacción entre alumnos y escuelas se desarrolla dentro de un sistema educativo, asimismo, esta relación es influida por la política educativa. Todas estas interacciones se dan dentro de un contexto institucional determinado por el desarrollo económico, político y cultural.

La prueba ENLACE, junto con los cuestionarios de contexto, nos permiten

<sup>6</sup> De acuerdo con este modelo, las variables correspondientes a eficacia en el ámbito regional son: PIB *per capita* en dólares PPP en 2004, índice de desarrollo humano, cociente del grado de escolaridad entre el analfabetismo, índice de marginación en 2005, tasa neta de cobertura en 2005, porcentaje de escuelas públicas sin computadora, porcentaje de escuelas privadas sin computadora.

DIAGRAMA 1. *Factores que inciden en el aprendizaje*

FUENTE: Elaboración propia a partir de Vegas y Petrow (2008).

identificar parte de las interacciones ilustradas en el diagrama 1. Los cuestionarios ENLACE contienen información de las características de los alumnos, antecedentes familiares y del hogar. Estos cuestionarios contienen también información de las escuelas que nos permiten conocer la relación entre los recursos escolares y el logro cognitivo. Por último, con información adicional por región, entidad federativa o municipio, es posible identificar el contexto institucional dentro del cual se da la relación entre alumnos y escuelas y por tanto estimar la importancia de éste en el proceso de aprendizaje. Identificar empíricamente los elementos esbozados en el diagrama 1 y su importancia relativa nos permite elaborar políticas educativas basadas en pruebas que pueden tener un efecto significativo en el logro escolar.

## 2. Metodología

Defínase  $Y_i$  como el logro escolar del estudiante  $i$  medido por el puntaje de matemáticas en la prueba ENLACE.<sup>7</sup> En términos muy generales podemos

<sup>7</sup> La puntuación de cada alumno se refiere a la capacidad de los sustentantes calculada con base en la

definir  $Y_i$  como resultado de *i*) las características personales del estudiante  $P_i$ ; *ii*) los antecedentes familiares del estudiante  $F_i$ ; *iii*) los recursos escolares a los que tiene acceso el estudiante  $i$  medido como las características del plantel al que asiste y los docentes que le imparten cátedra  $R_i$ , y *iv*) finalmente, el logro del estudiante  $i$  pudiera estar influido por el desarrollo institucional de la región, el estado o municipio en donde está ubicado el plantel  $I_i$ . Si definimos  $A(\cdot)$  como una función que transforma los insumos educativos  $P_i, F_i, R_i$  e  $I_i$  en logro escolar:<sup>8</sup>

$$Y_i = A(P_i, F_i, R_i, I_i) \quad (1)$$

en la que, al menos desde el punto de vista teórico,

$$\left( \frac{\partial Y_i}{\partial X_i} \right) > 0 \quad \text{y} \quad \left( \frac{\partial^2 Y_i}{\partial X_i^2} \right) < 0$$

es decir, un incremento en cualquier insumo (*ceteris paribus*), tiene un efecto positivo pero decreciente en el logro cognitivo. Sabemos tan poco de  $A(\cdot)$  que no podemos decir si es homotética de grado 1 en todos sus elementos; en otras palabras, no sabemos si hay economías a escala en el sistema educativo en México. Con base en los estudios revisados en la sección I, sabemos que no hay una sola forma funcional para la ecuación (1); por otra parte, la bibliografía identifica a una serie de factores asociados al logro, los cuales pueden ser clasificados dentro de nuestros cuatro grandes componentes  $P_i, F_i, R_i$  e  $I_i$ . Por tanto definimos  $P_i, F_i, R_i$  e  $I_i$  como vectores de tamaño  $K_1, K_2, K_3$  y  $K_4$ , respectivamente, y suponemos una relación lineal entre estos insumos y el logro cognitivo:

$$Y_i = \alpha + \kappa P_i + \gamma F_i + \beta R_i + \theta I_i + \varepsilon_i \quad (2)$$

en la que  $\alpha, \kappa, \gamma, \beta, \theta$  son parámetros por estimar y  $\varepsilon_i$  denota un término aleatorio con distribución normal, media igual a 0 y varianza conocida. En ausencia de un modelo estructural que determine la forma funcional y las

teoría de respuesta al *item*. El puntaje es obtenido al multiplicar la función de respuesta al item por cada item para obtener una función probabilística, el punto más alto de la cual es la estimación de máxima verosimilitud del parámetro de la persona.

<sup>8</sup> La función  $A(\cdot)$  depende de una variedad de factores entre los cuales podemos encontrar los procesos educativos que se establezcan en cada región, estado, plantel o inclusive los que establezca cada docente; asimismo  $A(\cdot)$  será afectada por la capacidad de gestión de cada director de plantel para utilizar los recursos escolares de manera eficiente, tomando en cuenta las características familiares e institucionales del plantel que dirige.

variables que deben ser incluidas en el modelo (2), frecuentemente el número de variables incluidas en los vectores  $P_i, F_i, R_i$  e  $I_i$  es muy alto, violándose el criterio de la parsimonia con lo cual el análisis tiende a perderse entre tanta información. El modelo ideal debe identificar un subconjunto de variables pertinentes, es decir, que estén altamente correlacionadas con  $Y_t$  y que no tengan una alta correlación entre sí.

Existen varias técnicas estadísticas para identificar las variables pertinentes y en el presente estudio utilizamos el análisis discriminante (Fisher, 1936; Tatsouka y Tiedeman, 1953). Para entender la mecánica del análisis discriminante, supóngase que separamos a toda la población de alumnos en dos grupos de acuerdo con su puntuación en la prueba ENLACE: sobresalientes y no sobresalientes. El análisis discriminante identifica un vector de variables  $x_1, x_2, \dots, x_m$ , cuya combinación lineal permite maximizar la probabilidad de que los individuos sean clasificados en la categoría a la que pertenecen. En otras palabras, las variables pertinentes identificadas por el análisis discriminante contienen información que permite clasificar a los individuos en dos (o más) categorías sin necesidad de consultar los resultados de la prueba estandarizada.

La prueba ENLACE establece no dos sino cuatro grados de logro: insuficiente, regular, bueno y excelente. Utilizando el análisis discriminante podemos identificar las variables pertinentes dentro de cada uno de los cuatro componentes ya señalados ( $P_i, F_i, R_i$  e  $I_i$ ). El análisis discriminante aplicado a cada componente tiene como resultado un subconjunto de variables que forman los vectores reducidos (parsimoniosos)  $Z'_i \in Z_i$  en que  $Z = (P, F, R, I)$  de tamaño  $k_j$  en que  $k_j \leq K_j$ , para  $j = (1, 2, 3, 4)$ . Esta estrategia reduce el número de variables independientes y facilita el análisis en torno de los determinantes del logro cognitivo:

$$Y_i = \alpha^* + \kappa^* P'_i + \gamma^* F'_i + \beta^* R'_i + \theta^* I'_i + \varepsilon_i^* \quad (3)$$

El modelo (3) puede utilizarse para medir la importancia de cada uno de los cuatro componentes en el resultado. Un modo sencillo de obtener esto es calculando la contribución en el coeficiente de determinación  $R^2$  que tiene cada uno de los cuatro componentes en el marco de una estimación de mínimos cuadrados. Sin embargo, no hay una solución analítica para calcular dicha contribución ( $dR^2/dx_k$ ) y por tanto es necesario estimar este efecto de manera empírica o numérica. La solución consiste en estimar y comparar la

$R^2$  entre dos modelos distintos en la que la única diferencia entre ambos es la presencia o ausencia del componente que se intenta evaluar. Formalmente, definimos  $X$  como un vector que contiene todos los argumentos de  $A(\cdot)$  y  $\phi$  como el conjunto vacío. La contribución del factor (o grupo de factores)  $X_k$  en el coeficiente de determinación (o cualquier otro estadístico de interés) puede ser definido como:

$$C_k = R^2(Y|X) - R^2(Y|X, X_k \in \phi) \quad (4)$$

en el que  $R^2(Y|X)$  es simplemente el coeficiente de determinación en la especificación (3) y  $R^2(Y|X, X_k \in \phi)$  es el coeficiente de determinación en otro modelo en el que uno de los componentes,  $X_k$ , es eliminado. Otra opción es definir la contribución de  $X_k$  como el incremento en  $R^2$  haciendo la siguiente comparación:

$$C_k = R^2(Y|X_k) - R^2(Y|\phi) \quad (5)$$

La ecuación (5) mide la contribución de  $X_k$  como la  $R^2$  en un modelo en el que sólo se incluye este componente.<sup>9</sup> El problema con las descomposiciones (4) y (5) es que no es exacta, es decir,  $\sum_{k=1}^K C_k \neq R^2(Y|X)$ . Una solución a este problema es eliminar todas las variables del modelo de manera secuencial hasta llegar al conjunto vacío. Sin embargo, el orden que sigamos en la secuencia de eliminación tendrá un efecto en la estimación de la contribución de cada variable o grupo de variables (componente). La descomposición de Shapley establecida por Shorrocks (1999) resuelve este problema de “dependencia secuencial” eliminando los elementos en  $X$ , siguiendo todas las secuencias de eliminación posibles, estimando la contribución ( $C_k$ ) en cada ronda y, finalmente, promediando todas las contribuciones:

$$C_k^* = \frac{1}{M!} \sum_{m=1}^{M!} C_{k,m} \quad (6)$$

en que  $M!$  es el número de posibles secuencias de eliminación y  $C_{k,m}$  representa —en nuestro caso— el cambio en  $R^2$  que se le atribuye a  $X_k$  en la secuencia de eliminación  $m$ . Esta descomposición asegura que  $\sum_{k=1}^K C_k^* = R^2(Y|X)$

<sup>9</sup> Con datos de la prueba PISA, Fuchs y Woessmann (2004) siguen las estrategias (4) y (5) para medir la importancia de factores familiares, las características del estudiante, los insumos del hogar, los recursos escolares y los factores institucionales.

y considera los diferentes grados de multicolinealidad en el modelo (3). Las contribuciones obtenidas utilizando la ecuación (6) nos dan el valor esperado del poder explicativo de cada uno de los componentes de la ecuación (3), que considera todos los escenarios o combinaciones posibles.<sup>10</sup>

### III. RESULTADOS

En esta sección se presenta los resultados de: *i*) el análisis discriminante para cada uno de los cuatro componentes en que se agrupan los factores o insumos: características individuales, recursos escolares, entorno institucional y antecedentes familiares; *ii*) la estimación de funciones de producción educativa después de identificar los factores pertinentes (ecuación 3), y *iii*) la descomposición de Shapley para cuantificar el porcentaje de la varianza en los resultados que se le puede atribuir a cada componente (ecuación 6). Todas las estimaciones que aquí se presentan consideraron que los datos provienen de una encuesta con una elaboración muestral compleja con dos estratificaciones: entidades federativas y escuelas, por lo que las varianzas fueron estimadas por el método delta y en las estimaciones se utilizaron los factores de expansión.

#### 1. Análisis discriminante

Los resultados del análisis discriminante para los cuatro componentes ( $P_i$ ,  $F_i$ ,  $R_i$  e  $I_i$ ), tomados de manera independiente, aparecen en la segunda columna del cuadro A1 del apéndice. Todas las variables incluidas en el cuadro A1 son significativas de acuerdo con el discriminante el cual utiliza el valor de la lambda de Wilks ( $\Lambda$ ) como criterio para discriminar y ordenar las variables de mayor a menor relevancia.<sup>11</sup> A mayor valor de  $\Lambda$ , mayor el poder

<sup>10</sup> McGahan y Porter (1997) realizan un ejercicio similar para obtener la importancia de factores asociados a las diferentes industrias para explicar las diferencias en utilidades en compañías que cotizan en la bolsa de los Estados Unidos.

<sup>11</sup> El estadístico  $\lambda$  de Wilks ( $\Lambda$ ) está definido como la proporción entre el determinante de la matriz de varianzas y covarianzas dentro de los grupos y el determinante de la matriz de varianzas y covarianzas total.

$$\Lambda = \frac{|VCg|}{|V Ct|}$$

Los valores de  $\Lambda$  varían entre 0 y 1: los valores cercanos a 0 indican mucha diferencia entre los grupos; por lo contrario, valores cercanos a 1 representan escasa discriminación o poca diferencia entre los grupos. El estadístico  $\Lambda$  lo utilizamos para seleccionar las variables que serán parte de la ecuación dis-

explicativo de determinada variable, lo cual implica que, para cada uno de los cuatro componentes, la primera variable en el cuadro A1 es la que tiene mayor poder discriminatorio seguida por la segunda en importancia y así sucesivamente.<sup>12</sup>

De la tercera a la quinta columna del cuadro A1 del apéndice se presentan distintos estadísticos  $R^2$ . La tercera columna muestra la  $R^2$  que se obtendría en un modelo de mínimos cuadrados ordinarios si las variables explicativas (columna 1) fueran ingresadas de manera secuencial y acumulativa en una regresión en la que el resultado de matemáticas en ENLACE 2008 es la variable dependiente. Naturalmente, al agregar más variables en un modelo de regresión, la  $R^2$  aumenta, sin embargo, como se muestra en la cuarta columna del cuadro A1, el incremento en  $R^2$  atribuido a la variable adicional es decreciente.

La quinta y última columna del cuadro A1 muestra la proporción de la  $R^2$  presentada en la tercera columna y la  $R^2$  en el modelo “completo”, es decir aquel que toma en cuenta todas las variables no descartadas por el análisis discriminante. Esta última columna muestra que, en el caso de las características de los alumnos, la gran mayoría de las variables, aunque tienen un poder discriminatorio significativo, su contribución adicional para explicar la variación en los resultados ENLACE es modesta. Con el fin de descartar variables con escaso poder explicativo y, con ello, mejorar la parsimonia del modelo, se tomó la decisión de enfocar el análisis sólo en el número de variables suficientes para explicar el 95% de la variación atribuible al modelo “completo”. Este límite, si bien es arbitrario, permite que los investigadores y los responsables de las políticas públicas enfoquen su atención en un número reducido de variables. Este límite está marcado con una pleca en el cuadro A1.

criminante, por medio del procedimiento de selección por pasos (*stepwise*). Esto es posible gracias a que el estadístico  $F$  puede calcularse a partir de  $\Lambda$ :

$$F = \left( \frac{1 - \Lambda}{\Lambda} \right) \left( \frac{n_1 + n_2 - p - 1}{p} \right)$$

Cada variable independiente candidata a ser incluida en el modelo se evalúa mediante una prueba de hipótesis de igualdad de medias entre los grupos. La variable que produce un mayor cambio en el estadístico  $F$  es la que tiene un mayor poder discriminatorio entre los grupos. En otros términos, la variable independiente que más incide en la reducción de  $\Lambda$  tiene mayor poder discriminatorio que el resto. Las variables cuya incorporación ya no se refleja en una reducción en el valor de  $\Lambda$ , no tiene sentido incorporarlas, puesto que no tienen poder discriminatorio alguno.

<sup>12</sup> El análisis discriminante sólo descartó una variable, la que se refiere al plan de estudios, en el que el alumno cursó su último año de secundaria: secundaria general, técnica, telesecundaria, abierta, para adultos o para trabajadores.

Así, al interior de las características individuales del alumno, las variables que resumen su trayectoria académica —promedio de matemáticas de secundaria, promedio general de primaria, tipo de primaria y secundaria— resultan ser las variables más importantes en la determinación del logro escolar. Ello se explica porque la producción del logro tiene una naturaleza acumulativa; es decir, el logro en educación media superior es resultado de todos los insumos que han incidido en los alumnos durante toda su vida y no sólo de los insumos que están incidiendo en el último grado escolar. En nuestro modelo, el tipo de primaria y secundaria son variables que condensan información de los insumos escolares pasados que incidieron en los alumnos. Asimismo, la introducción de las variables del promedio de matemáticas de secundaria y el promedio general de primaria están plenamente justificadas en la bibliografía del tema. De acuerdo con Todd y Wolpin (2007), la bibliografía de la función de producción educativa comúnmente adopta especificaciones de valor agregado cuando no existen o están incompletos los datos de los insumos pasados (por ejemplo, Tamura, 2001).

En su forma más básica, la especificación de valor agregado relaciona una medida de logro escolar con los insumos contemporáneos más una variable rezagada de la medición del logro, que funciona como línea base. En nuestro modelo, el promedio de matemáticas de secundaria y el promedio general de primaria son variables proxy del logro escolar rezagado. Al introducir variables que miden los insumos escolares pasados y el logro escolar rezagado en nuestro modelo estamos dándole su justa dimensión a los insumos actuales. Si no se introdujeran esas variables el modelo tendría el problema de variables omitidas y se estaría sobredimensionando la importancia de los insumos actuales en la función de producción educativa. Ahora bien, estas variables se introducen como parte de las características individuales de los alumnos, porque son elementos que trae consigo cada alumno al momento de ingresar a la educación media superior, por lo que los recursos escolares actuales ya no pueden incidir en ellas.

El hecho de ser mujer también es importante para explicar las diferencias en el desempeño. Este resultado es compatible con la evidencia empírica internacional, que muestra que las mujeres tienen mejor resultado que los hombres en las pruebas que evalúan las capacidades de comprensión lectora, mientras que los hombres obtienen un puntaje más alto en capacidad matemática (OCDE, 2009). Las expectativas de nivel máximo de estudios están incidiendo positivamente en el desempeño, aunque se reconoce que

esta variable podría tener un componente de endogeneidad importante, es decir, el buen desempeño escolar estaría también moldeando las expectativas y no sólo las expectativas estarían incidiendo en el desempeño. En cualquier caso, este problema no está resuelto en la bibliografía ni teórica ni empíricamente.

En cuanto al grupo de insumos escolares, se observa que variables que están relacionadas con la infraestructura, la escolaridad de los docentes, el ambiente escolar y el programa de estudios están incidiendo de manera relevante en el desempeño escolar. Ello implica que existen diversos ámbitos en los cuales puede incidir favorablemente la política pública educativa en el nivel medio superior, con el objetivo de mejorar el logro educativo. Por lo que se refiere al entorno institucional, las tres variables incluidas: el grado de marginación de la localidad donde se ubica la escuela, los diferentes subsistemas de educación media superior que existen en el país y las entidades federativas son importantes para explicar las diferencias en el desempeño de los alumnos.

Finalmente, en los insumos que conforman los antecedentes familiares, la escolaridad del padre, la marginación del hogar y el número de libros disponibles en el hogar explican 93.2% de la varianza dada por ese factor. Contrario a la bibliografía del tema, los años de escolaridad de la madre es la variable que menos afecta el logro escolar de los estudiantes de bachillerato. Una hipótesis que podría explicar este hecho es que la escolaridad de la madre es mucho más significativa en la infancia de los individuos, por lo que el efecto de dicho insumo estaría plenamente incorporado en las variables proxy del logro previo, que son el promedio general de primaria y el promedio de matemáticas de secundaria.

## *2. Función de producción educativa*

El modelo de la función de producción educativa fue estimado suponiendo que los estados constituyen un estrato. Asimismo, dado que los estudiantes de una misma escuela comparten características que pueden tener un efecto en su desempeño en la prueba, la estimación considera que los individuos están agrupados (*clustering*) en escuelas. Esta estrategia permitirá obtener errores estándar eficientes.

El coeficiente de determinación  $R^2$  muestra que los factores individuales son los que explican un mayor porcentaje de la varianza de la función de

producción educativa, seguido de los recursos escolares, el entorno institucional y, finalmente, los antecedentes familiares. El escaso efecto de los antecedentes familiares podría explicarse con el mismo argumento utilizado en el párrafo anterior para el caso de la escolaridad de la madre: el efecto de los insumos del hogar es mucho mayor en la infancia, lo cual queda incorporado en las variables proxie del logro escolar rezagado, mientras que para los jóvenes de 17 o más años de edad, que están por concluir su bachillerato, los antecedentes familiares son menos relevantes para explicar su logro en capacidad matemática.

En el cuadro A2 del apéndice se presenta los resultados de la estimación de cuatro modelos por mínimos cuadrados ordinarios (MCO). La metodología consistió en estimar las funciones de producción introduciendo de manera sucesiva el conjunto de insumos que conforman cada factor, omitiendo los insumos que aportaban poco a la explicación del logro, de acuerdo con la metodología establecida líneas arriba. Ello permite contar con un modelo que se caracteriza por su parsimonia, sin perder significativamente su poder explicativo. El orden en que se fueron introduciendo los factores en el modelo va de aquellos que aportan un mayor incremento en el coeficiente de determinación  $R^2$  a los que aportan menos. Así, el modelo 1 corresponde a la estimación del logro escolar en capacidad matemática como variable endógena y los factores asociados a las características individuales como variables exógenas. El modelo 2 agrega los recursos escolares como variables exógenas. En el modelo 3 se añade el entorno institucional a los dos factores anteriores. Finalmente, en el modelo 4 se estima un modelo completo, en el que se incluye todos los insumos. La mayoría de las variables introducidas en el modelo resultaron significativas. Cabe señalar que se obtiene un  $R^2$  de 0.42, que es relativamente alto en la bibliografía del tema.

Una posible crítica a los resultados presentados hasta ahora es que la variable dependiente utilizada por el análisis discriminante y la utilizada en la estimación de la función de producción educativa no es la misma. En la primera utilizamos una variable dependiente de manera discreta con cuatro grados de logro (insuficiente, elemental, bueno y excelente) para identificar los factores relevantes mientras que en la segunda explotamos la variación en la variable continua (puntaje en ENLACE). Para incrementar la confiabilidad y solidez del modelo, en el cuadro A3 del apéndice se presenta el resultado de estimar el modelo completo por medio de un modelo logístico multinomial ordenado (*ordered multinomial logit*), comparando los resul-

tados con el modelo estimado por MCO. En el modelo logístico, la variable dependiente son los niveles de logro en capacidad matemática en la prueba ENLACE.<sup>13</sup> En general, el signo de los coeficientes y el nivel de significación son compatibles con el modelo de MCO, lo cual muestra la congruencia del modelo estimado.

### *3. Importancia relativa de los cuatro componentes: Descomposición de Shapley*

La estimación de las funciones de producción educativa no dirime la cuestión de cuál de los cuatro componentes —individuales, escolares, institucionales, familiares— tienen mayor relevancia para explicar el logro escolar. Ello porque el incremento en  $R^2$  presentado en el cuadro A2 está influido por el orden en que cada componente fue agregándose en el modelo. Por lo anterior, recurrimos al cálculo de la descomposición de Shapley (ecuación 6), lo cual nos permite establecer la contribución de cada componente al  $R^2$  observado en el modelo completo.

En la primera columna del cuadro A3 del apéndice se presenta el cálculo de la  $R^2$  si se estima el logro escolar, medido como el puntaje de la prueba de capacidad matemática, como una función de cada uno de los factores o componentes. En la segunda columna se presenta el incremento en  $R^2$  como resultado de agregar cada componente al modelo completo. Ninguno de estos dos estadísticos mide la contribución real de cada componente al logro, puesto que, en el primer caso, se sobreestima el efecto de cada componente, al no controlar por el resto de los componentes. En el segundo caso, el incremento en el  $R^2$  está influido por el orden en que se incorporó a la estimación cada componente, subestimándose el efecto. El resultado de la descomposición de Shapley se presenta en la tercera columna y, en la cuarta se muestra la contribución porcentual de cada componente al logro cognitivo. Así, el componente de características individuales es el más importante para explicar el logro educativo, puesto que contribuye con 61.1% del valor del coeficiente de determinación en el modelo que incluye a los cuatro componentes, seguido por los recursos escolares (18.2%), el entorno institucional (14.5%) y, finalmente, los antecedentes familiares (6.2%).

<sup>13</sup> Los niveles de dominio fueron establecidos por un comité de expertos con base en la metodología bookmark.

## CONCLUSIONES

En los años recientes México ha logrado avances en cuanto a los años de escolaridad de su población; sin embargo, los avances en las pruebas internacionales en cuanto al logro cognitivo son muy escasos. Este hecho resulta decepcionante por sí mismo y por el efecto que tiene el logro cognitivo en el desarrollo económico. En consecuencia, y con el propósito de instrumentar políticas educativas eficaces, resulta imprescindible indagar los determinantes del logro cognitivo en México.

De la revisión de la bibliografía podemos concluir que no existe consenso en cuanto a cuáles son los determinantes del logro cognitivo, lo cual dificulta la elaboración de políticas. Por ello, esta investigación utiliza los resultados de la primera prueba estandarizada de educación media superior, conocida como ENLACE, que evalúa la capacidad matemática y lectora de los alumnos de tercer año. Esta prueba se acompaña de información captada por medio de cuestionarios de contexto aplicados a una muestra grande de alumnos, así como la aplicación de un cuestionario de contexto censal para directores. Hasta donde sabemos, es la primera vez que se utiliza estos datos para indagar los determinantes del logro cognitivo.

Ante el gran número de variables que podrían estar incidiendo en el logro cognitivo, en este documento se utilizaron diversas metodologías para obtener un modelo más parsimonioso, con el fin de que la atención de investigadores y tomadores de decisiones pueda enfocarse en un reducido número de variables. Así, primero clasificamos los posibles determinantes del logro en cuatro componentes: características individuales, recursos escolares, entorno institucional y antecedentes familiares, de acuerdo con la bibliografía del tema. Después, utilizamos un análisis discriminante que nos permitió ordenar las variables de acuerdo con su poder para clasificar a los alumnos en cuatro grados de logro: elemental, insuficiente, bueno y excelente. Con ello, estimamos funciones de producción educativa y fijamos un criterio *ad hoc* para descartar las variables con menor poder discriminatorio y que explicaban menos de 5% de la variabilidad considerada en cada componente. Con la finalidad de determinar qué tipo de componente era el que más contribuía a explicar el logro cognitivo utilizamos la descomposición de Shapley, con la cual podemos obtener el porcentaje de contribución de cada componente a la variabilidad del modelo, sin que dicho cálculo dependa del orden en que cada componente haya sido introducido en el modelo.

Los resultados muestran que las características individuales de los alumnos (en particular sus antecedentes escolares) es el componente más importante para evaluar el logro cognitivo, seguido de los recursos escolares, el entorno institucional y, por último, los antecedentes familiares. Este hallazgo resulta de la mayor importancia para la política pública, puesto que implica que la autoridad educativa tiene margen de maniobra para aumentar el logro educativo por medio de los recursos escolares. Otro resultado interesante es la escasa contribución de los antecedentes familiares al logro cognitivo. Ello podría explicarse porque el entorno familiar podría ser más importante en edades más tempranas del desarrollo, por lo que su efecto estaría incorporado en el logro cognitivo alcanzado en la educación básica, pero sería menos importantes en los jóvenes de 17 años y más de edad que cursan el tercer año de bachillerato.

El siguiente paso en la investigación de los determinantes del logro cognitivo en México consistirá en utilizar los datos de las siguientes aplicaciones de la prueba ENLACE, con el objetivo de analizar los determinantes de los cambios en el logro cognitivo. Ello con el fin de brindar elementos que mejoren la toma de decisiones en la política educativa.

## APÉNDICE

CUADRO A1.  $\lambda$  de Wilks y cambio en  $R^2$  resultado de agregar insumos, por componente

Variables (insumo o factor)	$\Lambda$	$R^2$	$\Delta R^2$	$R^2$ porcentaje
<i>Componente 1: Característica de los alumnos</i>				
Promedio de matemáticas de secundaria	.823	0.2108	0.2108	60.1
Promedio general de primaria	.790	0.2502	0.0394	71.4
Género	.756	0.2852	0.0350	81.3
Sostenimiento de escuela secundaria	.731	0.3064	0.0212	87.4
Tipo de escuela del último año de la primaria	.721	0.3199	0.0135	91.2
Escolaridad máxima que te gustaría alcanzar	.714	0.3303	0.0104	94.2
Año y mes de nacimiento	.711	0.3341	0.0038	95.3
Capacidad para buscar información en internet	.709	0.3374	0.0033	96.2
Beca de alto rendimiento en EMS	.707	0.3381	0.0007	96.4
Faltar a clases estando en la escuela	.706	0.3389	0.0008	96.7
Faltar a la escuela	.704	0.3403	0.0014	97.1
Discapacidad física	.703	0.3420	0.0017	97.5
Tiene trabajo	.702	0.3429	0.0009	97.8
Beca por desventaja económica en EMS	.702	0.3434	0.0005	97.9

Variables (insumo o factor)	$\Lambda$	$R^2$	$\Delta R^2$	$R^2$ porcentaje
Ingreso mensual promedio esperado si estudia una carrera universitaria	.701	0.3464	0.0030	98.8
Ingreso mensual promedio esperado si no estudia una carrera universitaria	.701	0.3494	0.0030	100
Beca de alto rendimiento en secundaria	.700	0.3496	0.0002	100
Número de veces que ha cambiado de escuela en EMS	.700	0.3499	0.0003	100
Horas al día dedicadas a estudiar o hacer tareas fuera del horario escolar	.700	0.3500	0.0001	100
Estudiantes que dejaron de estudiar un año en educación media superior	.699	0.3502	0.0002	100
Casado	.699	0.3504	0.0002	100
Beca por desventaja económica en primaria	.699	0.3505	0.0001	100
Beca por desventaja económica en secundaria	.699	0.3506	0.0001	100
Beca de alto rendimiento en primaria	.699	0.3506	0.0000	100
<i>Componente 2: Recursos escolares</i>				
Número de laboratorios en el ciclo 2008	.960	0.0416	0.0416	25.5
Número de materias de matemáticas en el grado que cursa	.932	0.0706	0.0290	43.3
Proporción de profesores con posgrado	.910	0.0942	0.0236	57.8
Horas de clase de matemáticas	.894	0.1133	0.0191	69.5
Relación entre alumnos	.883	0.1279	0.0146	78.4
Tasa de alumnos por maestro	.878	0.1338	0.0059	82.0
Tasa de alumnos por maestro	.878	0.1338	0.0059	82.0
Tamaño de la escuela	.872	0.1414	0.0076	86.7
Computadora por alumno	.871	0.1460	0.0046	89.5
Alumnos por aula	.869	0.1472	0.0012	90.3
Proporción de profesores sin licenciatura	.868	0.1484	0.0012	91.0
Relación entre alumnos y autoridades de la escuela	.867	0.1496	0.0012	91.7
Director por concurso de oposición	.866	0.1565	0.0069	96.0
Gravedad del ausentismo de los docentes	.865	0.1595	0.0030	97.8
Tasa de alumnos por maestro (cuadrado)	.865	0.1602	0.0007	98.2
Alumnos por aula (cuadrado)	.864	0.1613	0.0011	98.9
Proporción de profesores de tiempo completo	.863	0.1616	0.0003	99.1
Relación entre alumnos y maestros	.862	0.1630	0.0014	99.9
Proporción de profesores de tecnologías	.862	0.1631	0.0001	100
<i>Componente 3: Entorno institucional</i>				
Marginación de la localidad de la escuela	.962	0.0530	0.0530	43.2
Subsistema de educación media superior	.961	0.1003	0.0473	81.7
Entidad federativa	.961	0.1228	0.0225	100
<i>Componente 4: Antecedentes familiares</i>				
Escolaridad del padre	.956	0.0530	0.0530	61.3
Marginación del hogar	.941	0.0719	0.0189	83.1
Número de libros en el hogar	.932	0.0806	0.0087	93.2
Escolaridad de la madre	.928	0.0865	0.0059	100

**CUADRO A2. Determinantes del logro escolar en educación media superior**

Variable	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
<i>Características individuales</i>								
Mujer	-0.38	(0.009)***	-0.36	(0.008)***	-0.37	(0.008)***	-0.36	(0.008)***
Promedio de primaria 6-6.4	-0.46	(0.052)***	-0.39	(0.051)***	-0.37	(0.049)***	-0.36	(0.051)***
Promedio de primaria 6.5-6.9	-0.34	(0.039)***	-0.27	(0.040)***	-0.25	(0.038)***	-0.24	(0.039)***
Promedio de primaria 7-7.4	-0.28	(0.023)***	-0.23	(0.024)***	-0.22	(0.022)***	-0.20	(0.022)***
Promedio de primaria 7.5-7.9	-0.18	(0.014)***	-0.15	(0.016)***	-0.16	(0.015)***	-0.15	(0.015)***
Promedio de primaria 8-8.4	-0.12	(0.011)***	-0.10	(0.010)***	-0.09	(0.009)***	-0.08	(0.009)***
Promedio de primaria 9-9.4	0.16	(0.012)***	0.15	(0.011)***	0.15	(0.010)***	0.15	(0.010)***
Promedio de primaria 9.5-9.9	0.34	(0.013)***	0.30	(0.013)***	0.31	(0.012)***	0.30	(0.012)***
Promedio de primaria 10	0.44	(0.031)***	0.40	(0.030)***	0.43	(0.030)***	0.41	(0.030)***
Promedio en último año de matemáticas en secundaria 6-6.4	-0.22	(0.023)***	-0.16	(0.024)***	-0.20	(0.023)***	-0.20	(0.024)***
Promedio en último año de matemáticas en secundaria 6.5-6.9	-0.27	(0.024)***	-0.21	(0.023)***	-0.25	(0.022)***	-0.25	(0.023)***
Promedio en último año de matemáticas en secundaria 7-7.4	-0.23	(0.011)***	-0.19	(0.012)***	-0.21	(0.013)***	-0.21	(0.013)***
Promedio en último año de matemáticas en secundaria 7.5-7.9	-0.16	(0.010)***	-0.14	(0.010)***	-0.15	(0.010)***	-0.16	(0.010)***
Promedio en último año de matemáticas en secundaria 8-8.5	0.18	(0.011)***	0.16	(0.011)***	0.16	(0.010)***	0.16	(0.010)***
Promedio en último año de matemáticas en secundaria 9-9.4	0.45	(0.013)***	0.42	(0.012)***	0.42	(0.011)***	0.42	(0.011)***
Promedio en último año de matemáticas en secundaria 9.5-9.9	0.71	(0.017)***	0.67	(0.017)***	0.68	(0.015)***	0.68	(0.015)***
Promedio en último año de matemáticas en secundaria 10	0.89	(0.027)***	0.83	(0.025)***	0.86	(0.024)***	0.86	(0.023)***
Primaria pública rural	-0.11	(0.019)***	-0.06	(0.017)***	-0.03	(0.013)***	-0.02	(0.012)***

Secundaria privada	0.28	(0.036)***	0.30	(0.038)***	0.27	(0.028)***	0.26	(0.028)***
Concluir el bachillerato	-0.41	(0.029)***	-0.32	(0.024)***	-0.28	(0.018)***	-0.26	(0.018)***
Estudiar técnico superior universitario	-0.24	(0.015)***	-0.20	(0.015)***	-0.17	(0.013)***	-0.15	(0.013)***
Estudiar carrera profesional	-0.14	(0.011)***	-0.12	(0.010)***	-0.11	(0.008)***	-0.10	(0.008)***
<i>Recursos escolares</i>								
Número de laboratorios en el ciclo 2008	0.02	(0.004)***	0.02	(0.004)***	0.02	(0.004)***	0.02	(0.004)***
Tasa de alumnos por maestro	-0.01	(0.002)***	-0.01	(0.002)***	-0.01	(0.002)***	-0.01	(0.002)***
Proporción de profesores con licenciatura o posgrado	0.23	(0.072)***	0.11	(0.061)*	0.12	(0.060)**		
Alumnos por aula	0.00	(0.001)***	0.00	(0.001)*	0.00	(0.001)*		
Número de materias asignadas a matemáticas en grado que cursa	0.06	(0.004)***	0.06	(0.004)***	0.06	(0.004)***	0.06	(0.004)***
Horas de clase de matemáticas	0.03	(0.004)***	0.04	(0.003)***	0.04	(0.003)***	0.04	(0.003)***
Malas relaciones entre alumnos	-0.25	(0.023)***	-0.24	(0.021)***	-0.23	(0.022)***	-0.23	(0.022)***
Buenas relaciones entre alumnos	-0.12	(0.010)***	-0.11	(0.008)***	-0.11	(0.008)***	-0.11	(0.008)***
Excelentes relaciones entre alumnos	0.01	(0.013)	0.01	(0.012)	0.00	(0.012)		
Malas relaciones entre alumnos y directores	0.03	(0.014)***	0.03	(0.012)***	0.02	(0.012)***		
Buenas relaciones entre alumnos y directores	0.04	(0.008)***	0.04	(0.007)***	0.04	(0.007)***	0.04	(0.007)***
Excelentes relaciones entre alumnos y directores	-0.09	(0.018)***	-0.08	(0.015)***	-0.08	(0.015)***	-0.08	(0.015)***
Director presentó examen de oposición	-0.06	(0.031)***	-0.01	(0.025)	-0.01	(0.025)	-0.01	(0.025)
Computadora por alumnos	0.08	(0.054)	0.17	(0.056)***	0.17	(0.056)***	0.17	(0.055)***
Menos de 50 alumnos	-0.18	(0.111)*	-0.35	(0.096)***	-0.34	(0.096)***	-0.34	(0.094)***
Entre 50 y menos de 100 alumnos	0.04	(0.062)	-0.11	(0.059)*	-0.09	(0.058)		
Entre 100 y menos de 200 alumnos	0.02	(0.049)	-0.08	(0.047)*	-0.07	(0.046)		
Entre 200 y menos de 300 alumnos	-0.06	(0.044)	-0.12	(0.038)***	-0.11	(0.037)***		
Entre 300 y menos de 400 alumnos	0.02	(0.047)	-0.04	(0.039)	-0.03	(0.038)		

CUADRO A2 (*conclusión*)

Variable	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
<i>Características de las instituciones</i>								
Marginación muy alta	-0.71	(0.108)***	-0.71	(0.110)***	-0.65	(0.110)***		
Marginación alta	-0.14	(0.066)**	-0.14	(0.064)*	-0.11	(0.064)*		
Marginación media	-0.13	(0.047)***	-0.13	(0.046)***	-0.12	(0.046)***		
Marginación baja	-0.04	(0.028)	-0.04	(0.027)	-0.03	(0.027)		
Bachillerato general	0.13	(0.030)***	0.13	(0.030)***	0.12	(0.030)***		
Profesional técnico	-0.10	(0.030)***	-0.10	(0.030)***	-0.10	(0.030)***		
Descentralizados de los estados	0.12	(0.037)***	0.12	(0.036)***	0.13	(0.036)***		
Centralizados de los estados	0.14	(0.056)***	0.14	(0.056)***	0.15	(0.056)***		
Autónomos	0.08	(0.058)	0.08	(0.058)	0.08	(0.058)		
Privados	0.07	(0.056)						
Controlando con una variable por entidad federativa								
<i>Características de la familia</i>								
Padre sin escolaridad							0.04	(0.015)***
Padre con primaria completa							0.01	(0.008)
Padre con nivel medio superior							0.01	(0.011)
Padre con carrera técnica							0.03	(0.011)***
Padre con licenciatura							0.10	(0.012)***
Padre con posgrado							0.04	(0.022)*
Marginación de la casa							-0.04	(0.005)***
Ninguno							-0.07	(0.021)***
11 a 25 libros							0.01	(0.010)
26 a 50 libros							0.01	(0.009)
51 a 100 libros							0.02	(0.012)***
101 a 200 libros							0.02	(0.014)
201 a 500 libros							0.04	(0.018)**
Más de 500 libros							0.04	(0.027)***
Constante	0.201	(0.018)***	-0.457	(0.092)***	-0.431	(0.088)***	-0.50	(0.087)***
Observaciones	193 720	169 593	169 593	165 451	165 451			
Observaciones ponderadas	791 036	703 197	703 197	688 493	688 493			
R <sup>2</sup>	0.34	0.39	0.42	0.43				

\*\*\* 99%, \*\* 95%, \* 90 por ciento.

CUADRO A3. *Estimación del modelo logístico multinomial ordenado*

Variable	Mínimos cuadrados ordinarios		Modelo logístico multinomial ordenado	
	Coef.	S.E.	Coef.	S.E.
<i>Características individuales</i>				
Mujer	-0.36	(0.008)*	-0.90	(0.022)*
Promedio de primaria 6-6.4	-0.36	(0.051)*	-0.97	(0.177)*
Promedio de primaria 6.5-6.9	-0.24	(0.039)*	-0.71	(0.123)*
Promedio de primaria 7-7.4	-0.20	(0.022)*	-0.50	(0.069)*
Promedio de primaria 7.5-7.9	-0.15	(0.015)*	-0.37	(0.043)*
Promedio de primaria 8-8.4	-0.08	(0.009)*	-0.21	(0.027)*
Promedio de primaria 9-9.4	0.15	(0.010)*	0.38	(0.026)*
Promedio de primaria 9.5-9.9	0.30	(0.012)*	0.71	(0.030)*
Promedio de primaria 10	0.41	(0.030)*	0.98	(0.071)*
Promedio en último año de matemáticas en secundaria 6-6.4	-0.20	(0.024)*	-0.45	(0.077)*
Promedio en último año de matemáticas en secundaria 6.5-6.9	-0.25	(0.023)*	-0.65	(0.060)*
Promedio en último año de matemáticas en secundaria 7-7.4	-0.21	(0.013)*	-0.52	(0.037)*
Promedio en último año de matemáticas en secundaria 7.5-7.9	-0.16	(0.010)*	-0.38	(0.029)*
Promedio en último año de matemáticas en secundaria 8-8.5	0.16	(0.010)*	0.39	(0.028)*
Promedio en último año de matemáticas en secundaria 9-9.4	0.42	(0.011)*	0.98	(0.029)*
Promedio en último año de matemáticas en secundaria 9.5-9.9	0.68	(0.015)*	1.57	(0.036)*
Promedio en último año de matemáticas en secundaria 10	0.86	(0.023)*	2.02	(0.052)*
Primaria pública rural	-0.02	(0.012)	-0.03	(0.033)
Estudió en primaria privada	0.10	(0.018)*	0.22	(0.041)*
Estudió en primaria indígena	-0.15	(0.065)**	-0.43	(0.167)*
Estudió en primaria comunitaria	-0.21	(0.016)*	-0.66	(0.053)*
Secundaria privada	0.26	(0.028)*	0.62	(0.063)*
Concluir el bachillerato	-0.26	(0.018)*	-0.68	(0.059)*
Estudiar técnico superior universitario	-0.15	(0.013)*	-0.32	(0.037)*
Estudiar carrera profesional	-0.10	(0.008)*	-0.24	(0.021)*
<i>Recursos escolares</i>				
Número de laboratorios en el ciclo 2008	0.02	(0.004)*	0.05	(0.010)*
Tasa de alumnos por maestro	-0.01	(0.002)*	-0.01	(0.004)*
Proporción de profesores con licenciatura o posgrado	0.12	(0.060)**	0.31	(0.159)***
Alumnos por aula	0.00	(0.001)***	0.01	(0.003)***
Número de materias asignadas a matemáticas en grado que cursa	0.06	(0.004)*	0.15	(0.011)*
Horas de clase de matemáticas	0.04	(0.003)	0.09	(0.009)*
Mala relaciones entre alumnos	-0.23	(0.022)*	-0.53	(0.061)*
Buena relaciones entre alumnos	-0.11	(0.008)*	-0.24	(0.026)*
Excelente relaciones entre alumnos	0.00	(0.012)	-0.01	(0.030)

CUADRO A3 (*conclusión*)

Variable	Mínimos cuadrados ordinarios		Modelo logístico multinomial ordenado	
	Coef.	S.E.	Coef.	S.E.
Mala relaciones entre alumnos y directores	0.02	(0.012)***	0.08	(0.034)**
Buena relaciones entre alumnos y directores	0.04	(0.007)*	0.10	(0.019)*
Excelente relaciones entre alumnos y directores	-0.08	(0.015)*	-0.18	(0.043)*
Director presentó examen de oposición	-0.01	(0.025)	0.00	(0.062)
Computadora por alumnos	0.17	(0.055)*	0.43	(0.141)*
Menos de 50 alumnos	-0.34	(0.094)*	-0.93	(0.239)*
Entre 50 y menos de 100 alumnos	-0.09	(0.058)	-0.25	(0.142)***
Entre 100 y menos de 200 alumnos	-0.07	(0.046)	-0.22	(0.111)**
Entre 200 y menos de 300 alumnos	-0.11	(0.037)*	-0.29	(0.090)*
Entre 300 y menos de 400 alumnos	-0.03	(0.038)	-0.09	(0.096)
<i>Características de las instituciones</i>				
Marginación muy alta	-0.65	(0.110)*	-1.86	(0.523)*
Marginación alta	-0.11	(0.064)***	-0.18	(0.163)
Marginación media	-0.12	(0.046)**	-0.27	(0.120)**
Marginación baja	-0.03	(0.027)	-0.08	(0.068)
Bachillerato general	0.12	(0.030)*	0.26	(0.075)*
Profesional técnico	-0.10	(0.030)*	-0.26	(0.076)*
Descentralizados de los estados	0.13	(0.036)*	0.32	(0.091)*
Centralizados de los estados	0.15	(0.056)*	0.38	(0.138)*
Autónomos	0.08	(0.058)	0.24	(0.143)
Privados	0.06	(0.055)	0.20	(0.140)
Controlando con una variable por entidad federativa				
<i>Características de la familia</i>				
Padre sin escolaridad	0.04	(0.015)**	0.11	(0.045)**
Padre con primaria completa	0.01	(0.008)	0.03	(0.025)
Padre con nivel medio superior	0.01	(0.011)	0.03	(0.027)
Padre con carrera técnica	0.03	(0.011)*	0.08	(0.029)*
Padre con licenciatura	0.10	(0.012)*	0.23	(0.031)*
Padre con posgrado	0.04	(0.022)***	0.12	(0.052)**
Marginación de la casa	-0.04	(0.005)*	-0.08	(0.013)*
Ninguno	-0.07	(0.021)*	-0.15	(0.059)**
11 a 25 libros	0.01	(0.010)	0.05	(0.029)***
26 a 50 libros	0.01	(0.009)	0.03	(0.027)
51 a 100 libros	0.02	(0.012)**	0.08	(0.031)**
101 a 200 libros	0.02	(0.014)	0.10	(0.040)**
201 a 500 libros	0.04	(0.018)**	0.14	(0.046)*
más de 500 libros	0.04	(0.027)	0.14	(0.070)**
Constante	-0.50	(0.087)*		
cut1			1.26	(0.220)*
cut2			3.87	(0.220)*
cut3			5.94	(0.222)*

\*\*\* 99%, \*\* 95%, \* 90 por ciento.

**CUADRO A4. Descomposición de Shapley. Contribución de cada componente en el logro escolar**

Componente	R <sup>2</sup> (individual)	ΔR <sup>2</sup> (completo)	Shapley	Contribución (porcentaje)
Individual	33.03	19.43	25.59	61.08
Escolar	15.65	2.91	7.63	18.22
Institucional	12.26	3.35	6.09	14.53
Familiar	8.07	0.32	2.58	6.16
Total	69.01	26.01	41.88	100

## REFERENCIAS BIBLIOGRÁFICAS

- Álvarez, J., V. García y H. A. Patrinos (2007), “Institutional Effects as Determinants of Learning Outcomes. Exploring State Variations in Mexico”, Policy Research Working Papers 4286, Banco Mundial.
- Backhoff, E., et al (2007), *Factores escolares y aprendizaje en México. El caso de la educación básica*, México, Instituto Nacional para la Evaluación de la Educación.
- Bacolod, M. P., y J. L. Tobias (2006), “Schools, School Quality and Achievement Growth: Evidence from the Philippines”, *Economics of Education Review* 25, pp. 619-632.
- Becker, Gary (1962), “Investment in Human Capital: A Theoretical Analysis”, *Journal of Political Economy*, 70, pp. 9-49.
- Behrman, Jere R. (por aparecer), “Investment in education—Inputs and Incentives”, D. Rodrik y M. R. Rosenzweig (comps.), *Handbook of Development Economics: The Economics of Development Policy*, vol. 5, Amsterdam, North-Holland.
- Bishop, John J., y Ludger Woessmann (2004), “Institutional Effects in a Simple Model of Educational Production”, *Education Economics* 12(1), pp. 17-38.
- Card, David, y Alan B. Krueger (1992), “Does School Quality Matter? Returns to Education and the Characteristics of public Schools in the United States”, *The Journal of Political Economy*, vol. 100, núm. 1, febrero, pp. 1-40.
- Coleman, James S., et al (1966), *Equality of Educational Opportunity*, Washington, U. S. Government Printing Office.
- De Gregorio, José, y Jong-Wha Lee (2002), “Education and Income Inequality: New Evidence from Cross-Country Data”, *Review of Income and Wealth* 48, pp. 395-416.
- Denison, Edward (1964), “Measuring the Contribution of Education (and the Residual) to the Economic Growth”, *The Residual Factor of Economic Growth*, París, OCDE.
- Díaz, M. A, G. Flores y F. Martínez (2007), *PISA 2006 en México*, México, INEE.
- Fisher, R. A. (1936), “The Use of Multiple Measurements in Taxonomic Problems”, *Annals of Eugenics* 7, pp. 179-188.

- Fuchs, Thomas, y Ludger Woessman (2007), "What Accounts for International Differences in Student Performance? A Re-examination Using PISA data", *Empirical Economics*, vol. 32, núm. 2-3, mayo.
- Glaeser, E. L., G. A. M. Ponzatto y A. Sheleifer (2007), "Why Does Democracy Need Education?", *Journal of Economic Growth*, vol. 12 (2), pp. 77-99.
- Glewwe, Paul, y Michael Kremer (2006), "Schools, Teachers, and Education Outcomes in Developing Countries", E. Hanushek y F. Welch (comps.), *Handbook on the Economics of Education*, vol. 2, Amsterdam, North-Holland.
- Goldhaber, D., y D. Brewer (1997), "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity", *Journal of Human Resources*, vol. 32, núm. 3, pp. 505-523.
- Hanushek, Eric A. (1986), "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature*, vol. 24 (3), septiembre, pp. 1141-1177.
- \_\_\_\_\_, (1995), "Interpreting Recent Research on Schooling in Developing Countries", *The World Bank Research Observer* 10.2, agosto, pp. 227-253.
- \_\_\_\_\_, (2003), "The Failure of Input-Based Schooling Policies", *Economic Journal* 113, núm. 485, febrero, pp. 64-98.
- \_\_\_\_\_, (2008), "Education Production Functions", Palgrave Encyclopedia, Steven N. Durlauf y Lawrence E. Blume (comps.), *The New Palgrave Dictionary of Economics*, Basingstoke, Palgrave Macmillan.
- \_\_\_\_\_, J. F. Kain y S. G. Rivkin (1998), "Teachers, Schools, and Academic Achievement", NBER Working Paper W6691.
- \_\_\_\_\_, et al (2008), "Education and Economic Growth" *Education Next*, primavera.
- Lleras-Muney, Adriana (2005), "The Relationship Between Education and Adult Mortality in the United States", *Review of Economic Studies* 72, pp. 189-221.
- Mayer, David, y Edson E. Serván (2009), "Formación de la capacidad cognitiva en México: Impactos económicos y de políticas públicas", *Estudios Económicos*, número extraordinario, pp. 83-122.
- Mincer, Jacob (1970), "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach", *The Journal of Economic Literature* 8(1), 1-26.
- OCDE (2009), *Education at a Glance*, París, OCDE.
- Schmelkes, S., et al (1997), *La calidad de la educación primaria. Un estudio de casos*, México, Fondo de Cultura Económica.
- Serrano, Valeria (2009), *Factores asociados a la habilidad matemática en la prueba ENLACE Media Superior*, México, Subsecretaría de Educación Media Superior, SEP.
- Tamura, Robert (2001), "Teachers, Growth and Convergence", *Journal of Political Economy*, vol. 109 (5), octubre, pp. 1021-1059.
- Tatsouka, M., y D. Tiedeman (1954), "Discriminant Analysis", *Review of Educational Research*, vol. 24, núm. 5, pp. 402-420.

- Todd, Petra E., y Kenneth I. Wolpin (2003), "On the Specification and Estimation of the Production Function for Cognitive Achievement", *The Economic Journal*, 113, febrero, F3-F33.
- \_\_\_\_\_, y \_\_\_\_ (2007), "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps", *Journal of Human Capital*, vol. 1, núm. 1.
- Tristán, A., et al (2008), *Análisis multínivel de la calidad educativa en México ante los datos de PISA 2006*, México, Instituto Nacional para la Evaluación de la Educación.
- Urquiola, Miguel, y Erick Verhoogen (2009), "Class-Size Caps, Sorting, and the Regression-Discontinuity Design", *American Economic Review*, vol. 99 (1), marzo, pp. 179-215.
- Vegas, Emiliana, y Jenny Petrow (2008), *Incrementar el aprendizaje estudiantil en América Latina. El desafío para el siglo XXI*, Washington, Banco Mundial y Ediciones Mayo.
- Woessmann, Ludger (2003), "Schooling Resources, Educational Institutions, and Student Performance: The International Evidence", *Oxford Bulletin of Economics and Statistics* 65 (2), pp.117-170.