



FACULTAD
DE CIENCIAS
ECONÓMICAS



Universidad
Nacional
de Córdoba

REPOSITORIO DIGITAL UNIVERSITARIO (RDU-UNC)

El impacto de la escuela media sobre la desigualdad de resultados entre estudiantes argentinos en PISA 2009. Una aplicación multinivel en presencia de heterogeneidad

Guillermo A. Coseani, Natalia S. Tolosa, Héctor R. Gertel

Ponencia presentada en XLVIII Reunión Anual de la Asociación Argentina de Economía Política
realizado en 2013 en Rosario. Santa Fe, Argentina



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual
4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

El impacto de la escuela media sobre la desigualdad de resultados entre estudiantes argentinos en PISA 2009. Una aplicación multinivel en presencia de heterogeneidad¹.

Guillermo A. Coseani¹, Natalia S. Tolosa² y Héctor R. Gertel³

1 *Instituto de Economía y Finanzas, FCE, UNC.* Correo: gcoseani@gmail.com

2 *Instituto de Economía y Finanzas, FCE, UNC.* Correo: nstolosa@gmail.com

3 *Instituto de Economía y Finanzas, FCE, UNC.* Correo: hgertel@gmail.com

Resumen

El estudio de la contribución relativa de las características escolares, personales y familiares a la explicación del rendimiento del alumno presenta dificultades metodológicas relacionadas al carácter anidado de los datos y a la presencia de heterogeneidad en las relaciones existentes entre los factores observables y el puntaje obtenido. Este trabajo utiliza los datos de la prueba de lectura de PISA 2009 para Argentina y aplica para su estudio el método multinivel por cuantiles recientemente propuesto en Geraci y Bottai. Los resultados confirman la diferente importancia que asume la escuela en la explicación de los resultados de la prueba de lectura a lo largo de la distribución de puntajes.

Abstract

The analysis of the relative contribution of school, personal and family characteristics in explaining student performance has some methodological difficulties related to the nested nature of the data and the presence of heterogeneity in the relationships between the observable factors and the score obtained. This paper uses data from the 2009 PISA reading test for Argentina and applies a multilevel quantile method, proposed recently by Geraci and Bottai, for its study. The results confirm the different importance the school has in explaining the reading test results along the distribution of scores.

Código JEL: I21, I28

Palabras clave: datos anidados, heterogeneidad, método multinivel por cuantiles, PISA 2009.

¹ Para ser presentado en XLVIII Reunión Anual de la Asociación Argentina de Economía Política, 13, 14 y 15 de noviembre de 2013. Rosario, Santa Fe. Este trabajo fue financiado mediante proyecto Secyt-UNC, periodo 2012-2013, resolución 162/12 y proyecto Secyt-PICT 2007, 803.

1. Introducción

Los resultados obtenidos por Argentina en la edición 2009 de las evaluaciones internacionales PISA situaron al país en una pobre posición. Por ejemplo, en la prueba de lectura el país ocupó la posición 58 entre 65 participantes al alcanzar un puntaje medio de 398 puntos. En la escala de PISA dicho puntaje se ubica por debajo del nivel 2, que representa el dominio de un conjunto básico de habilidades y capacidades para resolver situaciones que se le presentarán al estudiante en su vida adulta. El nivel alcanzado por el promedio de los estudiantes sugiere la debilidad de la escuela secundaria argentina para proporcionar un nivel razonable de habilidades para la vida. Por otro lado, la desigualdad de puntajes entre los estudiantes también es alta. El Cuadro N° 1 presenta la proporción de alumnos de Argentina y otros países de América Latina según el nivel de habilidades alcanzado en la prueba de lectura de PISA 2009. Allí puede observarse la alta desigualdad que caracteriza a la Argentina dentro de la región. Utilizando los seis niveles de competencias en lectura definidos por PISA² surge que el porcentaje de alumnos argentinos que alcanzó niveles superiores de competencias (nivel 4 o más) es similar al registrado en países como Brasil, Uruguay y Chile; sin embargo, existe una mayor concentración de alumnos en el nivel inferior (1b y 1a) e incluso por debajo de éste. La desigualdad en el puntaje argentino sólo puede ser comparada en la región con la existente en Panamá y Perú. En suma, no sólo resulta preocupante el bajo nivel de aprovechamiento escolar mostrado por los alumnos argentinos; también es necesario resaltar que la distribución de puntajes es una de las más desiguales de la región tal como lo apuntan Marchionni, Vazquez, y Pinto (2012).

Cuadro N° 1: Puntaje promedio y porcentaje de alumnos en cada nivel de competencia lectora

País	Puntaje promedio	Nivel del alumno promedio	Niveles de competencia lectora							
			Menor a 1b	1b	1a	2	3	4	5	6
Chile	449	2	1,3	7,4	21,9	33,2	25,6	9,3	1,3	0,0
Uruguay	426	2	5,5	12,5	23,9	28,0	20,3	8,1	1,7	0,1
México	425	2	3,2	11,4	25,5	33,0	21,2	5,3	0,4	0,0
Colombia	413	2	4,2	13,9	29,0	30,6	17,1	4,6	0,5	0,0
Brasil	412	2	5,0	16,0	28,6	27,1	15,9	6,1	1,2	0,1
Argentina	398	1a	10,8	15,8	25,0	25,4	16,0	6,0	0,9	0,1
Panamá	371	1a	13,3	23,1	28,9	20,7	10,1	3,4	0,5	0,0
Perú	370	1a	14,1	22,0	28,7	22,1	10,1	2,6	0,4	0,0

Fuente: Elaboración propia en base a PISA 2009

La desigualdad entre estudiantes puede descomponerse para su estudio en la diferencia del puntaje de cada alumno comparado con el puntaje promedio de la escuela a la cual asisten (variación intra-escolar), y en la diferencia entre los puntajes promedios de las escuelas (variación inter-escolar)³. Esta última se explica por ciertas características que comparten todos los alumnos de una misma escuela y difieren entre establecimientos (características grupales), mientras que la dispersión de los rendimientos de los alumnos respecto del rendimiento promedio de su propia escuela se explica por diferencias particulares entre ellos (características individuales). Cervini encuentra que las características de la escuela

² comenzando por los niveles 1b, 1a y 2 que refieren a competencias sencillas de comprensión de texto, hasta los niveles 5 y 6 que implican un trabajo de interpretación complejo basado en habilidades en lectura y conocimientos elevados. Los alumnos ascienden en el ranking a medida que aumenta la dificultad de las tareas que pueden resolver y se supone que en cada nivel pueden responder exitosamente las tareas propias del mismo, como de los inferiores.

³ Algunos trabajos analizan niveles superiores de variación: provinciales, nacionales, etc. Sin embargo dadas las características del muestreo de PISA sólo es posible analizar la varianza de puntajes en dos niveles.

explican alrededor del 31% de la varianza de los puntajes obtenidos en las evaluaciones censales realizadas por el Ministerio de Cultura y Educación de Argentina (Cervini, 2008). Krüger y Formichella trabajando con las pruebas PISA 2009 encuentran que el factor escuela explicaría en la media hasta un 57% de la varianza de los puntajes entre alumnos. Estos resultados sugieren que las diferencias entre escuelas es un componente importante en la explicación de la varianza total.

Resulta interesante preguntarse si el efecto de las diferencias entre el puntaje obtenido por los centros es igualmente importante en ambas colas de la distribución de puntajes. Esta tarea presenta simultáneamente múltiples desafíos relacionados al carácter anidado de los datos y a la posible existencia de efectos heterogéneos entre cuantiles. La literatura que analiza el efecto independiente de la escuela sobre los resultados escolares se apoya en la aplicación de métodos lineales jerárquicos o HLM tradicional (Cervini, 2002) (Marchionni, Vazquez, & Pinto, 2012) (Decándido, 2011). Aunque respetan el carácter jerárquico de los datos, estos trabajos resuelven el problema considerando el puntaje del alumno "promedio". Esto supone homogeneidad en el comportamiento de la población de alumnos bajo estudio, de manera que los resultados obtenidos son representativos del comportamiento esperado de toda la muestra. Sin embargo, la literatura ha demostrado que en presencia de heterogeneidad en las poblaciones educativas disminuye la representatividad del puntaje medio (Gertel, Giulodori, Vera, & Bustos, 2010) (Sosa Escudero, Giovagnoli, & Porto, 2009). En estos casos, la regresión por cuantiles propuesta por Koenker y Bassett (Koenker & Bassett, 1982; Koenker & Bassett, 1978) es una técnica de gran utilidad cuando se busca describir el comportamiento de toda la distribución condicional de puntajes. Diversos autores han ensayado estimaciones aplicando ecuación cuantílica para observar la posible presencia de distinta intensidad de efectos para una misma variable explicativa a lo largo de la distribución de la variable dependiente. Estos avances permitieron realizar análisis más precisos de los factores que influyen sobre el rendimiento escolar. No obstante, no existe hasta el momento un estudio que utilice ambas técnicas en conjunto. Más recientemente, Geraci y Bottai (Geraci & Bottai, 2007; Geraci & Bottai, 2011) propusieron una solución para el tratamiento conjunto del problema que supone la presencia de datos jerárquicos y la heterogeneidad de los efectos, desarrollando un algoritmo que permite estimar modelos HLM por cuantiles. Este avance ofrece una herramienta para resolver el interrogante planteado, dado que el mismo requiere realizar un análisis por tramo de la distribución de puntajes atendiendo simultáneamente a la jerarquía de los datos para identificar la contribución de las características escolares a la dispersión de resultados obtenidos por los alumnos.

El presente trabajo propone introducir el supuesto de heterogeneidad de las poblaciones educativas en el análisis del peso relativo de los factores escolares e individuales en la determinación del rendimiento en lectura de Argentina en PISA 2009. Con este propósito, se aplica el método propuesto por Geraci y Bottai (Geraci & Bottai, 2011). Esta técnica basada en la probabilidad incorpora interceptos aleatorios a las regresiones condicionales por cuantiles y permite reflejar simultáneamente la dependencia inherente al carácter anidado de los datos educativos y la presencia de efectos heterogéneos de un conjunto de factores seleccionados a lo largo de la distribución de puntajes condicionada. A través de la aplicación de este método se indaga la importancia relativa de las características institucionales e individuales sobre el rendimiento académico, según la posición del estudiante en la distribución de puntajes. En particular, se busca confirmar la hipótesis de que las características institucionales poseen un menor poder explicativo de la variabilidad total del puntaje cuando se analizan estudiantes posicionados en los extremos de las distribuciones de puntajes de sus escuelas. Se espera que estos alumnos se encuentren más influenciados por sus características personales y familiares que por las características de las escuelas a las cuales asisten en comparación al efecto que éstas tienen sobre el rendimiento de sus pares posicionados alrededor del puntaje promedio. Por el contrario, se espera que los estudiantes con desempeño cercano al centro de la distribución sean más sensibles a los factores escolares. Esta información resulta de gran importancia para

comprender las dimensiones que afectan a la desigualdad educativa en Argentina al momento de diseñar políticas públicas en educación.

Con el propósito de indagar la importancia relativa que asume la institución escolar a lo largo de la distribución de puntajes, el trabajo se estructura de la siguiente manera: en la próxima sección se expone una revisión de la literatura referida al análisis de los factores individuales y escolares que afectan el desempeño académico de los alumnos y a la heterogeneidad en el efecto de dichos factores entre cuantiles. En la sección 3 se presentan los datos utilizados en esta investigación junto con la metodología desarrollada por Geraci y Bottai y los interrogantes de investigación formulados para una primera aplicación de modelos multinivel por cuantiles a muestras educativas. En la sección 4 se presenta el análisis de los resultados empíricos obtenidos. Finalmente en la sección 5 se concluye con una discusión de los principales resultados obtenidos.

2. Antecedentes en la investigación

A continuación se revisan algunos antecedentes disponibles en la literatura referidos al estudio de los factores que afectan el rendimiento académico. En primer lugar, se repasan estudios que analizan el efecto y el poder explicativo de un conjunto de factores seleccionados sobre el rendimiento de los alumnos, respetando el carácter anidado de los datos educativos. En segundo lugar, se reseñan antecedentes académicos que comprueban la existencia de heterogeneidad en las relaciones entre el puntaje obtenido y el conjunto de factores seleccionados. Finalmente, se presentan estudios que proporcionan avances relevantes en técnicas de análisis para el tratamiento conjunto de datos anidados en presencia de heterogeneidad.

La Función de Producción Educativa (Hanushek, 1979) (Hanushek, 2007) es una herramienta de amplio uso en Economía de la Educación para el estudio de las relaciones existentes entre el rendimiento de los alumnos en pruebas estandarizadas y sus características personales, familiares y de la escuela a la cual asisten. El carácter anidado de los datos educativos (alumnos agrupados en escuelas, y escuelas agrupadas en sistemas) requiere de la aplicación de un método de estimación que tome en cuenta esta característica. Caso contrario, la aplicación de técnicas convencionales de análisis de regresión lineal arrojan estimadores sesgados e inconsistentes (Rowe, Hill, & Holmes-Smith, 1995) (Brik & Raudenbush, 1992). (Monette, Qing, & Kwan, 2001) (Gelman & Hill, 2009). Una manera de evitar estas consecuencias es la utilización de técnicas de análisis jerárquico de datos, como la regresión multinivel, recomendada por la OCDE (OECD, 2004). Los modelos multinivel respetan el hecho de que los estudiantes están anidados dentro de escuelas, siendo posible diferenciar y dimensionar apropiadamente los efectos de las características grupales e individuales.

En general, los estudios que utilizan análisis de regresión multinivel para tratar datos educativos en América Latina y Argentina obtienen conclusiones similares en cuanto a los factores más influyentes en el rendimiento académico. Entre los trabajos más relevantes pueden citarse las investigaciones realizadas por Duarte, Bos y Moreno con información del Segundo Estudio Regional Comparativo y Explicativo (SERCE) referido a América Latina (Duarte, Bos, & Moreno, 2009), y Cervini (Cervini, 2002) (Cervini, 2008) y Kruger y Formichella (Krüeger & Formichella, 2013) para el estudio del sistema educativo argentino. Entre los principales resultados de estos trabajos, se pueden mencionar las siguientes coincidencias: (i) las escuelas tienen un peso relativo significativo en la variación del rendimiento entre alumnos⁴, (ii) existe una relación positiva y significativa entre la condición socioeconómica de los estudiantes y sus resultados académicos; adicionalmente, el perfil socioeconómico promedio de la escuela constituye un importante predictor del puntaje alcanzado y contribuye a explicar una porción relevante de la dispersión en los aprendizajes

⁴ Cervini encuentra que alrededor del 30% de la varianza en el rendimiento de lengua se debe a diferencias entre escuelas. Kruger y Formichella por su parte, encuentran valores cercanos al 57%

y (iii) los resultados anteriores sugieren la existencia de una fuerte segmentación institucional asociada a un alto grado de segregación socioeconómica entre centros.

Los trabajos citados hasta aquí confirmaron el efecto directo de las variables que captan el nivel socioeconómico del alumno y del alumnado escolar, como así también el significativo peso relativo de la escuela en el rendimiento académico. Es necesario remarcar que estas investigaciones han supuesto homogeneidad en el comportamiento de la población educativa, por lo tanto los resultados obtenidos en estimaciones a la media se consideran adecuadamente representativos del comportamiento de toda la población. Sin embargo, en los últimos años, se ha demostrado la presencia de heterogeneidad en las relaciones existentes entre el rendimiento escolar (variable dependiente) y diversas co-variables que buscan explicarlo (Decándido, Gigena, Giuliadori, & Gertel, 2011) (Sosa Escudero, Giovagnoli, & Porto, 2009). Por lo tanto, analizar los resultados académicos basándose únicamente en el comportamiento medio no resulta representativo de la totalidad de la población educativa. El impacto y la contribución relativa de los factores que afectan el rendimiento académico pueden presentar importantes diferencias según cuál sea la posición del estudiante en la distribución de puntajes. Esta información resulta de especial interés para el diseño de políticas educativas orientadas a impulsar a los alumnos de peores desempeños, no representados por los estudios basados en el alumno "promedio". A los fines de describir el comportamiento de toda la distribución condicional de puntajes, la regresión por cuantiles propuesta por Koenker y Bassett (Koenker & Bassett, 1978) (Koenker & Bassett, 1982) es una técnica de gran utilidad que ha sido frecuentemente utilizada en la literatura. Para el caso del sistema educativo argentino existen investigaciones que han justificado la necesidad de utilizar técnicas de regresión por cuantiles en el estudio de los factores relacionados al rendimiento académico. Decándido et al. demostraron que el impacto de un conjunto de variables seleccionadas sobre los resultados obtenidos por los estudiantes evaluados en SERCE, difiere para valores bajos, medios y altos de la distribución condicional de puntajes (Decándido, Gigena, Giuliadori, & Gertel, 2011). En la misma línea se encuentran los resultados obtenidos por Sosa Escudero, Giovagnoli y Porto, en su estudio del efecto de las características individuales sobre la distribución del rendimiento universitario. Los resultados empíricos sugieren la presencia de efectos heterogéneos en la relación entre los factores seleccionados (características familiares, posibilidad de no trabajar durante la carrera, entre otros) y la performance de los alumnos. El impacto de una mejora en estas condiciones resulta más fuerte en términos de incremento en el rendimiento para aquellos alumnos ubicados en la cola inferior de la distribución (Sosa Escudero, Giovagnoli, & Porto, 2009).

Sin embargo, la técnica de regresiones por cuantiles originalmente desarrollada por Koenker y Bassett supone que los datos son independientes entre sí, es decir, no considera posibles casos de anidamiento. Para un análisis más acabado, resulta importante contar con técnicas de estimación que respeten simultáneamente ambas características de los datos de las poblaciones educativas. Recientemente, Geraci y Bottai (Geraci & Bottai, 2007)(Geraci & Bottai, 2011) han propuesto un método de estimación basado en la probabilidad, que extiende la regresión por cuantiles de datos independientes a datos agrupados. Esta técnica está siendo actualmente utilizada en investigaciones referidas a bioestadística y tratamientos biológicos (Kjetil H. Løland) (Fenske, Fahrmeir, Hothorn, Rzehak, & Höhle, 2013) (Geraci & Bottai, Linear quantile mixed models, 2011), no obstante tiene un alto potencial en el estudio de la Economía de la Educación. Mediante la incorporación de interceptos aleatorios a las regresiones condicionales por cuantiles, el desarrollo de los autores permite reflejar simultáneamente la dependencia inherente al carácter anidado de los datos educativos y la presencia de efectos heterogéneos sobre la variable respuesta.

3. Metodología y tratamiento de los datos

1) Datos

En la presente investigación se utiliza la base de datos proporcionada por el proyecto PISA en su edición 2009. PISA es un operativo de evaluaciones estandarizadas a nivel internacional desarrollado conjuntamente por la UNESCO y la OECD desde el año 2000 con el objetivo de medir la preparación que poseen los alumnos de 15 años para enfrentar los desafíos que les presenta el mundo globalizado. Adicionalmente a los resultados referidos a las competencias de los alumnos en Lectura, Ciencias y Matemática, PISA desarrolla y provee información de cuestionarios complementarios respondidos por los alumnos y los directivos o responsables de las escuelas, por lo que es de gran utilidad para evaluar los factores personales e institucionales que intervienen en el proceso educativo. El operativo 2009 hizo hincapié en el área de Lectura⁵, por lo que se toma como variable dependiente, o producto del proceso educativo, el desempeño alcanzado en esta prueba. El resultado reportador por PISA surge de un complejo procedimiento que consiste en calcular cinco valores plausibles para cada estudiante con el propósito de reflejar impactos de variable ambientales no controlados. Por lo que el método recomendado por la OCDE para estimar el efecto de una variable seleccionada sobre los resultados consiste en obtener el estimador de la variable seleccionada para cada valor plausible y luego promediarlos de forma de poder extrapolar la predicción del efecto sobre el resultado muestral al resultado poblacional (Adams & Wu, 2002).

El diseño de muestreo utilizado en las evaluaciones PISA es de tipo estratificado en dos etapas donde las unidades de muestreo de la primera etapa son las escuelas con estudiantes de 15 años de edad y las unidades de muestreo de la segunda etapa son los estudiantes de las escuelas seleccionadas con anterioridad. En otras palabras, una vez que las escuelas son seleccionadas para estar en la muestra, se prepara una lista de los estudiantes de 15 años de edad para cada escuela. En caso en los que la lista contenga más de 35 estudiantes, solamente se seleccionan 35 estudiantes asignando igual probabilidad a cada uno. Para las listas que contienen menos de 35 estudiantes, se toma el total (Adams & Wu, 2002). La complejidad del procedimiento lleva a respetar este diseño muestral en las estimaciones para lo cual es necesario tomar en cuenta el uso de los pesos muestrales provistos por PISA⁶.

2) Estrategia metodológica

A los fines de estudiar los factores que se relacionan con el rendimiento educativo argentino a partir de la información provista por el operativo PISA, se estiman modelos multinivel en dos niveles: (i) alumno o nivel 1; y (ii) escuela o nivel 2. Estos modelos reconocen el hecho de que los alumnos están anidados dentro de la escuela por lo que es posible evaluar el porcentaje de la variación relativa del rendimiento académico que se explica por

⁵ El operativo se realiza cada tres años evaluando las competencias de los estudiantes en Matemática, Ciencias y Lectura, haciendo hincapié en una de estas áreas en cada edición, en forma rotativa. La primera edición, correspondiente al año 2000 tuvo su énfasis en Lectura; la edición del año 2003 en Matemáticas y la del año 2006 en Ciencias.

⁶ “Si las unidades muestrales no tienen las mismas oportunidades de ser seleccionadas y si los parámetros poblacionales se calculan sin tener en cuenta estas probabilidades diferenciales, quizá los resultados también queden sesgados. Para compensar estas distintas posibilidades, es necesario ponderar o asignar pesos a los datos. La ponderación consiste en reconocer que algunas unidades de la muestra son más importantes que otras y deben contribuir más que las restantes al cálculo de cualquier estimación poblacional. Una unidad de muestra con probabilidad de selección muy pequeña se considerará más importante que una unidad con gran probabilidad de selección. Por tanto, los pesos son inversamente proporcionales a la probabilidad de selección” (OECD, 2004b).

características de los estudiantes y el porcentaje explicado por las características de las escuelas⁷.

La estructura algebraica del modelo de regresión de dos niveles, es la siguiente:

$$y_{ij} = \alpha_j + \beta_w W_{ij} + \varepsilon_{ij} \quad (1)$$

$$\alpha_j = \gamma_{00} + \gamma_{0z} Z_j + u_{0j} \quad (2)$$

Donde y_{ij} representa la competencia lectora de los estudiantes evaluados en el operativo, el subíndice i hace referencia al alumno (nivel 1) y el subíndice j a la escuela a la cual asiste (nivel 2). El intercepto del modelo multinivel (α_j) siempre es un efecto aleatorio y se divide, tal como lo expresa la ecuación (2) en una parte fija (γ_{00}) que indica el intercepto general – media de los interceptos de las escuelas-, y una parte aleatoria ($\gamma_{0z} Z_j + u_{0j}$), que expresa la diferencia entre el puntaje de la escuela y la media de los interceptos de las mismas. W_{ij} representa el vector de variables independientes del nivel alumno y Z_j el vector de variables independientes del nivel escuela.

Combinando las ecuaciones (1) y (2), se obtiene la ecuación simple del modelo, que provee la estructura algebraica clásica del modelo de intercepto aleatorio a estimar:

$$y_{ij} = \gamma_{00} + \beta_w W_{ij} + \gamma_{0z} Z_j + u_{0j} + \varepsilon_{ij} \quad (3)$$

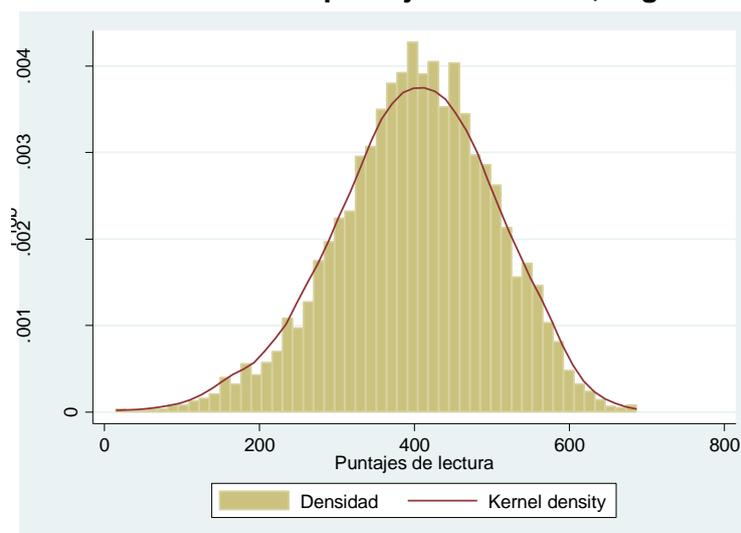
Esta especificación permite identificar, a través del intercepto aleatorio (u_{0j}), el efecto que tiene asistir a una escuela determinada sobre el rendimiento académico luego de haber controlado por características personales y familiares del alumno. En este modelo ε_{ij} refleja la diferencia de puntajes no explicada entre alumnos de una misma escuela.

Como se comentó previamente este tipo de modelos permiten dimensionar correctamente el impacto de las variables correspondientes a ambos niveles sobre el puntaje obtenido. Sin embargo, debido a que el modelo focaliza su análisis en la media deja de lado la cuestión de la heterogeneidad en las relaciones entre los factores observables y la variable dependiente. Una manera de determinar la presencia de heterogeneidad es a través del análisis de la simetría de la distribución de puntajes, dado que distribuciones asimétricas sugieren la presencia de este tipo de relaciones. La Figura 1 muestra gráficamente la distribución del puntaje obtenido en las pruebas de lectura por los estudiantes argentinos evaluados en PISA 2009. Tomando la mediana de la distribución de 403 puntos, se observa que en la cola izquierda (observaciones menores a 200 puntos) existe mayor dispersión que en la cola derecha (observaciones mayores a 600 puntos). Adicionalmente, el valor del coeficiente de asimetría de -0,2749 confirma el sesgo negativo de esta distribución⁸.

⁷ “Puesto que PISA selecciona por cada centro participante una muestra aleatoria de una población de la misma edad de entre todos los cursos y clases, permite la descomposición de la varianza sólo en dos niveles: una varianza entre centros y una varianza dentro de los centros” (OECD, 2004)

⁸ Se realizó la prueba de Shapiro-Wilks para probar la hipótesis de distribución normal de los puntajes para Argentina. La misma arrojó el rechazo de la hipótesis nula, concluyéndose que los puntajes de lectura no provienen de una distribución normal.

Figura N° 1: Distribución de los puntajes en lectura, Argentina PISA 2009



Fuente: Elaboración propia en base a datos PISA 2009

La presencia de asimetría en la distribución de puntajes indica que la estimación de modelos multinivel en la media no constituye una medida representativa del comportamiento de toda la población, motivando la aplicación de técnicas de estimación por cuantiles. Precisamente, la alternativa desarrollada por Geraci y Bottai tiene la ventaja de poder estimar parámetros de funciones cuantílicas sujetas a múltiples efectos aleatorios tal como es el caso de las funciones lineales multinivel que nos ocupa. La técnica se basa en la conexión existente entre el método de mínimos cuadrados penalizados tradicionalmente utilizado en las regresiones por cuantiles y la maximización de la función de probabilidad asimétrica de Laplace. A partir de esta conexión, los autores extienden el método de regresión por cuantiles tradicional, para aplicarlo a datos dependientes. Las ventajas del método lo hacen útil para su aplicación a diversos campos de investigación asociados a muestras con estructuras complejas de dependencia. Entre ellos se encuentra el estudio de los efectos de ciertas drogas en pruebas clínicas con grupos de control o placebo basados en datos de medidas repetidas (Geraci & Bottai, 2007) (Geraci & Bottai, 2011), paneles de estudio (Velez Reyes 2011, Salvioni 2013, Damassio 2013), pruebas epidemiológicas, etc. La aplicación de estos modelos en datos anidados no ha sido muy explorada a pesar de su gran potencial para el estudio de pruebas educativas donde las unidades muestrales se encuentran anidadas y existe la presencia de heterogeneidad.

Esta investigación aplica por primera vez el modelo propuesto por Geraci y Bottai a muestras educativas, con el propósito de estudiar el peso relativo que tienen los factores escolares e individuales en la determinación del rendimiento en lectura en presencia de heterogeneidad. Este enfoque permite realizar estimaciones considerando simultáneamente la heterogeneidad existente en la población y la estructura jerárquica de los datos, mediante la incorporación de efectos aleatorios a modelos de regresión por cuantiles.

Partiendo de la ecuación (3), si expresamos convenientemente los coeficientes de regresión fijos indicados por el intercepto general (γ_{00}), los coeficientes asociados a cada variable de nivel uno (β_w) y los coeficientes (γ_{0z}) que expresan el efecto global de cada escuela sobre sus alumnos en un vector $\theta_x = (\gamma_{00}, \beta_w, \gamma_{0z})$ conjuntamente con las variables de ambos niveles expresadas como $X_{ij}^T = (1; W_{ij}; Z_j)$ y renombrado al error asociado a las escuelas (u_{0j}) como u_j podemos re-exresar la ecuación (3) como un modelo multinivel específico para cada cuantil de la distribución de puntajes de la manera siguiente:

$$G_{y_{ij} | u_j}(\tau | x_{ij}, u_j) = X_{ij}^T \theta_x^{(\tau)} + u_j \quad (4)$$

Donde $G_{y_{ij} | u_j}(\tau | x_{ij}, u_j) = F_{y_{ij} | u_j}^{-1}(\cdot)$ es la inversa de la función acumulada de puntajes individuales condicionada a los efectos fijos u_j correspondientes al τ –ésimo cuantil. En la ecuación (4) el problema a resolver es la estimación de los coeficientes de regresión contenidos en el vector $\theta_x^{(\tau)}$ y los efectos fijos u_j propios del cuantil τ . Como se observa en la ecuación (4) Las observaciones y_{ij} , se encuentran condicionadas a u_j , donde u_j es un error asociado a cada escuela. Los efectos fijos (u_j) inducen a una estructura de correlación entre las observaciones dentro de una misma escuela. Por lo tanto, es conveniente asumir que:

- i) El efecto fijo (u_j) se encuentra idénticamente distribuido de acuerdo a alguna función de densidad caracterizada por el parámetro τ dependiente de una matriz de varianza-covarianza Ψ y que (u_j) es independiente de τ .
- ii) Los errores aleatorios para los alumnos de la escuela j en el cuantil τ ($\varepsilon_{ij}^{(\tau)}$) son independientes y se distribuyen Asimétrica Laplace con $AL(0, \sigma, \tau)$.
- iii) Los u_j y $\varepsilon_{ij}^{(\tau)}$ son independientes uno de otro.

Nuestro modelo, que considera únicamente el intercepto aleatorio u_{0j} para capturar el efecto de asistir a una escuela específica y efectos fijos para todas las variables de nivel estudiante y escuela, queda expresado de la siguiente manera:

$$y_{ij}^{(\tau)} = \gamma_{00}^{(\tau)} + \beta_w^{(\tau)} W_{ij} + \gamma_{0z}^{(\tau)} Z_j + u_{0j} + \varepsilon_{ij} \quad (5)$$

La estructura algebraica resultante es similar a un modelo de intercepto aleatorio como el planteado en la ecuación (3). Geraci y Bottai han propuesto recientemente un algoritmo que incluye técnicas de integración numérica y de optimización no estándar necesarios para estimar eficientemente los parámetros de interés de ecuaciones multinivel por cuantiles como la expuesta en la ecuación (5) (Geraci & Bottai, 2011).

La aplicación de este novedoso método, utilizando los resultados obtenidos por los alumnos argentinos evaluados en PISA 2009, nos permite determinar la importancia relativa de las características institucionales e individuales sobre el rendimiento académico en distintos tramos de la distribución de puntajes. En el próximo apartado se plantean los interrogantes que guían esta investigación y que se abordan haciendo uso del modelo de estimación planteado en (5).

3. Principales interrogantes

Este trabajo busca confirmar la hipótesis de que las características escolares tienen un menor poder explicativo sobre la variabilidad total del puntaje cuando se analizan estudiantes posicionados en los extremos de las distribuciones de puntajes de sus escuelas. Se espera que estos alumnos se encuentren más influenciados por sus características personales y familiares que por las características de las escuelas a las cuales asisten en comparación con sus compañeros de rendimiento cercano al promedio. Adicionalmente se proponen los siguientes interrogantes de investigación motivados en la gran desigualdad de puntajes de Argentina en PISA 2009 advertida al comienzo del trabajo en el Cuadro N° 1:

1. Los trabajos que estudian los efectos para la media de la población coinciden en la asociación positiva entre el nivel socioeconómico familiar y el rendimiento académico del alumno. Sin embargo, se puede intuir que la intensidad de la relación difiere según la posición ocupada por el estudiante en la distribución de puntajes. En particular, se espera que ésta sea menor para los alumnos con rendimiento cercano a la media de la distribución
2. Los alumnos que provienen de sectores desfavorecidos padecen un doble castigo, a saber: a la desventaja económica familiar inicial se suma un efecto contextual negativo dada la mayor probabilidad de asistir a escuelas de bajo nivel socioeconómico promedio. El modelo multinivel por cuantiles permite examinar el

efecto del contexto socioeconómico sobre el rendimiento educativo y analizar si ésta asociación varía entre alumnos ubicados en distintos tramos de la distribución de puntajes.

3. Finalmente, se espera que el efecto del contexto socioeconómico escolar sobre el puntaje del alumno tenga un impacto superior al del nivel socioeconómico de su familia.

Para la estimación del modelo de función de producción educativa representado por la ecuación (5), se seleccionaron las siguientes variables:

- (i) A nivel alumno, se consideraron las variables “genero”, “repitió”, “joyread”, “memor” y “cstrat”. La primera de ellas fue incluida bajo la forma de una variable dummy que asume el valor de 1 si el alumno es de sexo masculino. Se espera que ésta tenga un efecto negativo sobre el puntaje dada la evidencia empírica a favor de que las mujeres obtienen mejores resultados en las pruebas de lectura. La segunda variable “repitió” es una variable dummy que asume el valor 0 si el alumno se encuentra en el décimo grado o superior y 1 en caso de que se encuentre en noveno grado o inferior, e intenta reflejar la historia académica previa de los alumnos y su signo esperado es negativo. La tercera variable busca capturar el gusto de los alumnos por la actividad evaluada, por lo que se espera un efecto positivo sobre el puntaje en esta prueba. Por último, las variables “memor” y “cstrat” fueron incorporadas para controlar por las estrategias de estudio utilizadas por el alumno (la frecuencia con que estudia memorizando la información y la frecuencia con que estudia resumiendo la información esencial, controlando que todo esté bajo control). A nivel familiar se decidió incorporar el Índice de status económico, social y cultural elaborado por PISA (“escs”) que resume tres dimensiones principales del nivel socioeconómico del alumno: el estatus ocupacional de sus padres, el clima educativo y la riqueza o posesiones del hogar.
- (ii) A nivel escuela se incluyeron “scmatedu”, “screpito” y “gestión” como variables de control para analizar el efecto de “scscs” sobre la varianza en el rendimiento medio alcanzado por los centros. La primera de las variables capta la calidad de los materiales educativos disponibles por los centros, se espera que mientras mayor sea esta disponibilidad mejor será el rendimiento promedio alcanzado por las escuelas. La variable “screpito” busca reflejar el “efecto de pares repitentes” por el cual una alta concentración de repitentes en la escuela puede afectar negativamente el desempeño de los estudiantes no rezagados. La variable dummy “gestion” asume el valor de 1 si el centro es de gestión privada y se espera un signo positivo debido a la mayor autonomía en la toma de decisiones y la capacidad de selección que poseen las escuelas privadas. Finalmente se incorpora como variable independiente el promedio del Índice de status económico, social y cultural para cada centro, con el objetivo de analizar el efecto de la composición del alumnado sobre el rendimiento de los estudiantes y la desigualdad en el puntaje promedio de los centros.

Una descripción detallada de la construcción de las variables utilizadas en el trabajo se presenta en la Tabla A.1 del anexo, mientras que la Tabla A.2 describe los principales estadísticos correspondientes a dichas variables.

El tamaño de la muestra referida a la prueba de lectura de Argentina en PISA 2009 es de 4774 estudiantes, anidados en 199 escuelas; siendo el número ponderado de alumnos de 472.106, un 74% de la población de 15 años de edad. A los fines de analizar el efecto composicional de las escuelas sobre el rendimiento académico se optó por eliminar aquellas observaciones correspondientes a las escuelas para las cuales no pudo construirse el índice de status socioeconómico de gran parte de su alumnado o que contaban con menos de seis alumnos, quedando la muestra reducida a 4680 alumnos agrupados en 180 escuelas.

4. Análisis empírico

El ordenamiento secuencial adoptado para el análisis responde a las hipótesis e interrogantes planteados y al proceso lógico de la técnica de análisis utilizada. Los modelos propuestos en el trabajo se estiman mediante técnicas de optimización no lineal haciendo uso del paquete “lqmm” disponible para el software estadístico R (Geraci, Package ‘lqmm’, 2012).

Para el análisis empírico se centraron las variables independientes con el propósito de permitir la correcta interpretación de los coeficientes estimados. Se estableció como alumno benchmark a aquel que posee características individuales promedio para la muestra del país, de sexo femenino, que no ha repetido de curso y asiste a una escuela pública con características promedio.

Primer resultado: partición inicial de la varianza en los dos niveles de agregación por cuantiles (modelo “vacío”)

Se comienza el análisis empírico con la partición de la varianza del rendimiento por niveles. La determinación del peso relativo de las características individuales y escolares sobre la varianza total permitirá contrastar la primera hipótesis planteada.

La descomposición de la varianza total se obtiene de manera sencilla estimando un modelo multinivel vacío, es decir sin variables predictoras. El mismo proporciona estimaciones sin sesgo de la varianza entre escuelas y dentro de las mismas. Al no poseer variables predictoras, $\alpha_j^{(\tau)}$ representa la media de las escuelas para cada cuantil, mientras que la varianza de u_{0j} será la varianza del rendimiento entre escuelas. A cada alumno se le asigna como puntaje predicho la media de su centro correspondiente al cuantil al cual pertenece, por lo que la varianza de ε_{ij} es igual a la varianza en el rendimiento de los alumnos dentro de las escuelas.

A partir de estas estimaciones se puede calcular el coeficiente de correlación intraclase o rho (ρ), definido como el ratio entre la varianza de los interceptos aleatorios y la varianza total. En otras palabras, este coeficiente indica el porcentaje de la varianza total del que son responsables las escuelas.

$$\rho = \frac{\psi_u^2}{\psi_u^2 + \psi^2}$$

Esta medida aporta una señal de las diferencias y similitudes entre centros escolares en un país determinado. Más específicamente, una correlación intraclase alta indica grandes diferencias entre centros escolares (baja inclusión académica), lo cual significa que los estudiantes tenderán a agruparse de manera sistemática en escuelas donde la mayoría de ellos tienen características similares. (OECD, 2010)

Los resultados del modelo no condicionado para la muestra de estudiantes argentinos en las evaluaciones PISA 2009 se resumen en el Cuadro N°2.

Cuadro N° 2. Resultados modelo no condicionado.

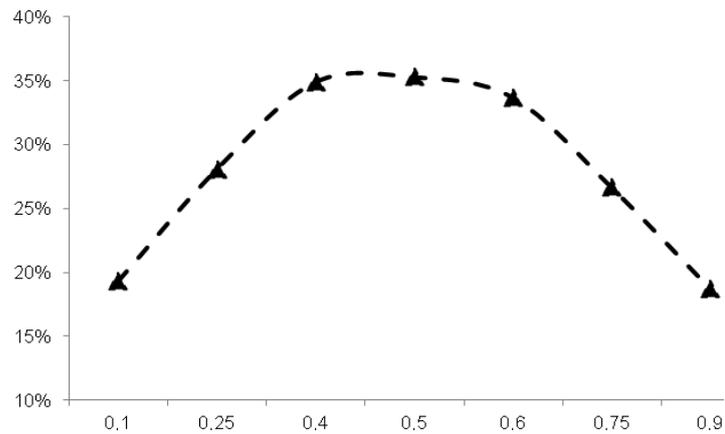
		Cuantiles						
		Media	0,1	0,25	0,4	0,5	0,6	0,75
Intercepto general	393,60	286,2	344,98	380,38	398,28	416,7	441,84	483,08
Intercepto Aleatorio	5702,81	4.197,60	4.018,80	4.091,80	3.875	3.768	3.506,80	3.740,20
Log		-28.087	-26.867	-22.189	-27.117	-21.700	-26.512	-27.882
Scale		13,15	24,05	29,06	29,79	28,6	23,06	12,65
AIC		56.217	54.940	54385	54.243	54.258	54.630	55.771
Varianza Total	12263,86	21.711	14.305	11.720	10.978	11.183	13.145	19.945
Varianza Intra-Escuelas	6561,05	17.514	10.286	7.629	7.103	7.415	9.638	16.205
Varianza Entre-Escuelas	5702,81	4.198	4.019	4.092	3.875	3.768	3.507	3.740
Correlación intraclase	47%	19%	28%	35%	35%	34%	27%	19%

Fuente: Elaboración propia en base a PISA 2009

El primer punto a remarcar es la diferente importancia relativa de la escuela sobre la varianza total asignada en el análisis por cuantiles en comparación con la atribuida por el análisis basado en el alumno “promedio”. Como puede observarse en los valores de la correlación intraclase reportados en el Cuadro anterior, el peso sobre la varianza total atribuible a las características de la escuela por el análisis en cuantiles en ningún caso alcanza el 47% de la estimación realizada en la media.

Del análisis de la descomposición de la varianza por cuantiles surge el comportamiento en forma de “U” invertida de la correlación intraclase, reflejada en el Gráfico N°2. Esto implica que la diferencia de rendimientos entre escuelas tiene un peso relativo mayor sobre la varianza total de los cuantiles centrales ($\tau = 0,40$, $\tau = 0,60$ y la mediana) que la importancia atribuible para la varianza de los cuantiles extremos ($\tau = 0,10$ y $\tau = 0,90$). Más específicamente, las características escolares poseen un mayor poder explicativo sobre la desigualdad de puntajes entre aquellos alumnos que se ubican cerca del centro de la distribución de resultados de su escuela. A su vez la dispersión entre alumnos con desempeños muy inferiores o superiores al promedio es más sensible a las características individuales y al contexto extra-escolar en comparación con el poder explicativo de estos factores sobre la desigualdad entre alumnos con desempeño promedio. Este resultado confirma la hipótesis inicial y sugiere que las políticas educativas enfocadas en las escuelas pueden ser ineficientes para tratar a los estudiantes de bajo rendimiento debido a que estos se encuentran más influenciados por sus características personales y el ambiente extra-escolar.

Figura N°2. Índice de Correlación Intraclase por cuantiles.



Fuente: Elaboración propia en base a PISA 2009

Adicionalmente, existe evidencia de una mayor variabilidad entre el puntaje promedio de las escuelas en los cuantiles inferiores dado que el intercepto aleatorio presenta un patrón decreciente a medida que nos desplazamos a lo largo de la distribución de puntajes.

Segundo resultado: análisis de la relación entre características individuales y rendimiento del alumno por cuantiles.

Con el propósito de responder al interrogante que plantea la posible existencia de efectos heterogéneos en la relación entre el nivel socioeconómico del alumno y el rendimiento alcanzado, se introduce la siguiente especificación de la ecuación (5):

$$y_{ij}^{(\tau)} = \gamma_{00} + \beta_{w1}^{(\tau)}(\text{genero})_{ij} + \beta_{w2}^{(\tau)}(\text{repetio})_{ij} + \beta_{w3}^{(\tau)}(\text{joyread})_{ij} + \beta_{w4}^{(\tau)}(\text{cstrat})_{ij} + \beta_{w5}^{(\tau)}(\text{memor})_{ij}$$

En la misma, el supra-índice τ especifica el cuantil para el cual se estima el modelo multinivel. El Cuadro N° 3 presenta los resultados obtenidos para cada uno de los cuantiles considerados.

Cuadro N°3– Resultados modelo estimado N°1

	Media	Cuantiles							
		0,10	0,25	0,40	0,50	0,60	0,75	0,90	
(Intercept)	422,84 ***	350,60 ***	387,86 ***	409,73 ***	422,66 ***	439,80 ***	465,90 ***	513,39 *	
Genero	-15,79 ***	-26,74 ***	-20,27 **	-17,40 ***	-17,05 *	-15,78 **	-12,77 (*)	-11,60	
Repitió	-46,12 ***	-53,86 ***	-54,45 ***	-51,99 ***	-51,79 ***	-51,90 ***	-53,72 ***	-46,42 *	
Escs	0,89 ***	1,22 **	0,94 *	0,87 ***	0,86 ***	0,94 ***	0,98 ***	0,89 *	
Joyread	1,20 ***	0,92 (*)	0,93 **	0,98 ***	1,01 ***	1,11 ***	1,19 ***	1,29 ***	
Menor	-0,47 ***	-0,32	-0,38	-0,46 **	-0,45 *	-0,49 ***	-0,53 ***	-0,58 **	
Cstrat	0,65 ***	0,63 *	0,64 *	0,75 **	0,69 ***	0,62 ***	0,64 ***	0,61 **	
Scale		11,81 ***	21,51 ***	25,95 ***	26,67 ***	25,69 ***	20,85 ***	11,15 ***	
Efecto aleatorio	3007,32	2,466	2,269	2,159	2,147	1,994	1,779	1,986	
Log		25,955	25,363	25,104	25,039	25,05	25,214	25,688	
AIC		51,928	50,745	50,226	50,097	50,119	50,447	51,395	
Varianza Total Explicada	32%	<u>24%</u>	<u>27%</u>	<u>30%</u>	<u>29%</u>	<u>29%</u>	<u>28%</u>	<u>27%</u>	
Varianza Intra-escuela Explicada	47%	19%	20%	20%	20%	20%	20%	22%	
Varianza Entre escuela Explicada	18%	41%	44%	47%	45%	47%	49%	47%	

Nota: (*) p<.10; *p<.05; **p<.025;***p<.01
Fuente: Elaboración propia en base a PISA 2009

La incorporación de variables de nivel uno al modelo no condicionado contribuyó a explicar 24% de la varianza total en el primer cuantil, y porcentajes algo mayores (entre 27% y 30%) para los cuantiles superiores.

Con respecto a la varianza intra-escuela total indicada en el Cuadro N°1, el modelo condicionado a las variables del nivel uno explica entre el 19 % para los cuantiles inferiores y el 22% para los cuantiles superiores. Para el caso de la varianza entre-escuela total presentada en el Cuadro N°1, el modelo condicionado a las variables del nivel uno contribuye a explicar entre el 41 % para los cuantiles inferiores y el 49% para los cuantiles superiores.

Los coeficientes de regresión obtenidos confirman el efecto positivo y significativo del nivel socioeconómico de la familia para todos los cuantiles estimados. Como se observa en el Cuadro N°1, el impacto de esta variable no es homogéneo entre cuantiles. Mientras que para el cuantil 0,10 el coeficiente correspondiente a la variable “escs” es 1,22; en la región central es de 0,86 y en el extremo superior de la distribución alcanza valores entorno a 0,90⁹.

Con respecto a las demás variables consideradas en el modelo, los coeficientes de regresión obtenidos para la variable “genero” tienen un efecto negativo sobre la competencia lectora en todos los cuantiles, indicando que el rendimiento en lectura de las mujeres es

⁹ Estimaciones adicionales no presentadas en esta versión del trabajo indican que mientras “escs” contribuye a explicar 2% de la varianza de nivel uno, su contribución es entre cinco a diez veces más fuerte en relación a la varianza del nivel dos. Este hecho refleja la influencia del hogar sobre la escuela a través del efecto pares que en este caso sugiere la presencia de una fuerte segregación escolar. Resultados similares han sido encontrado para Argentina en otros trabajos (Jaume, 2013) (Gasparini, Jaume, Serio, & Vazquez, 2011).

superior al de los hombres. Nuevamente, el efecto no es homogéneo ya que la brecha entre el puntaje de mujeres y varones es superior para los cuantiles inferiores, disminuye a lo largo de la distribución de puntajes y deja de ser estadísticamente significativa para los alumnos que poseen logros educativos superiores al promedio ($\tau = 0,75$ y $\tau = 0,90$). Esto implica que para los alumnos sobresalientes el género deja de ser influyente en términos de rendimiento.

Los resultados confirman el efecto negativo de ser repitente, el cual es aproximadamente similar entre cuantiles. Los alumnos repitentes obtienen aproximadamente 52 puntos menos que sus pares no repitentes, lo que equivale a más de medio nivel de competencia en la escala de PISA. Es necesario remarcar la alta incidencia de la repitencia en el alumnado argentino, en el operativo PISA del año 2009 representaban el 38% la muestra.

Los coeficientes de regresión para la variable que indica la utilización de estrategias de estudios muestra el efecto positivo de su aplicación sobre la competencia lectora, similar para todos los cuantiles. Esto marca la importancia de que los profesores persuadan a sus alumnos para que incorporen elementos de este tipo de métodos de estudio. Por otro lado, utilizar la memorización de contenidos como herramienta de estudio posee un efecto negativo, significativo y creciente para los cuantiles superiores.

La variable referida a los hábitos de lectura del alumno ("joyread") muestra efectos positivos y estadísticamente significativos para todos los cuantiles, siendo los coeficientes de regresión mayores para los alumnos de alta competencia lectora. Por lo tanto, estimular el hábito de la lectura mejora la competencia lectora de los alumnos, sobre todo en aquellos que sobresalen en su curso por obtener los mejores rendimientos.

Al correr un modelo sencillo en la media se obtienen coeficientes similares a los que surgen de una regresión cuantílica estimada para $\tau = 0,5$. Como se observa en el Cuadro N°2 en las secciones extremas de la distribución los coeficientes asumen valores que en algunos casos difieren sensiblemente de los obtenidos para la región central. Por lo tanto analizar el efecto de un conjunto de variables seleccionadas utilizando como única medida los coeficientes correspondientes a la región central omite información valiosa.

Tercer resultado: análisis de la relación entre características individuales, escolares y rendimiento del alumno por cuantiles.

Finalmente se adicionan las variables del nivel dos que permiten determinar el poder explicativo y la relación existente entre la composición socioeconómica escolar y el puntaje alcanzado por los alumnos, según la posición ocupada por los mismos en la distribución de puntaje (interrogantes 2 y 3).

El modelo empírico a estimar incluyendo ahora las variables del nivel uno y dos, siguiendo la especificación de la ecuación (5) es:

$$y_{ij}^{(\tau)} = \gamma_{00} + \beta_{w1}^{(\tau)}(genero)_{ij} + \beta_{w2}^{(\tau)}(repitio)_{ij} + \beta_{w3}^{(\tau)}(joyread)_{ij} + \beta_{w4}^{(\tau)}(cstrat)_{ij} \\ + \beta_{w5}^{(\tau)}(memor)_{ij} \\ + \gamma_{0z1}^{(\tau)}(scscs)_j + \gamma_{0z2}^{(\tau)}(screpitio)_j + \gamma_{0z3}^{(\tau)}(scmatedu)_j + \gamma_{0z4}^{(\tau)}(gestion)_j + u_{0j} + \varepsilon_{ij}$$

En el Cuadro N° 4 se exponen los resultados del presente modelo y en el Gráfico A.1 del Anexo se muestran los gráficos para las variables consideradas en el modelo y los coeficientes de regresión obtenidos a partir de las estimaciones. En el eje horizontal se encuentran el cuantil estimado (τ) y en el eje vertical los valores de los coeficientes de regresión. Los gráficos muestran las estimaciones por cuantiles (puntos) y sus intervalos de confianza al 95 % (barras de error). Al mismo tiempo se incluyeron para comparación el valor del coeficiente de regresión obtenido para un modelo multinivel en la media (línea sólida) y sus intervalos de confianza al 95 % (línea punteada).

Como se observa en el Cuadro N°4, la incorporación de variables que reflejan las características de la escuela aumenta el poder explicativo del modelo en una proporción no

despreciable. La varianza total explicada por el modelo completo, que incluye variables de nivel uno y dos, es de 29% en el primer cuantil y porcentajes algo mayores (entre 34% y 38%) para los cuantiles superiores.

Con respecto a la varianza intra-escuela, el modelo completo explica entre un 22% para los cuantiles inferiores y un 29% para los cuantiles superiores. Para el caso de la varianza entre-escuela, el modelo completo contribuye a explicar entre un 59% para los cuantiles inferiores y un 72% para los cuantiles superiores. Comparando la explicación lograda por el modelo condicionado a las variables del nivel uno (expuesta en el Cuadro N°2) versus la lograda por el modelo completo, la incorporación de variables del nivel escuela no aporta gran poder explicativo sobre la varianza intra-escuela, pero sí para la dispersión de puntajes entre escuela.

A pesar de esta mejora en el poder explicativo, de las variables referidas a las características de la escuela sólo el nivel socioeconómico y cultural promedio de la escuela ("scescs") mostró efectos estadísticamente significativos. Los coeficientes obtenidos difieren entre cuantiles de puntaje, destacándose la asociación existente con el rendimiento en el cuantil superior, para el cual un aumento en un 10% en el nivel socioeconómico promedio de la escuela se asocia con un aumento de casi 20 puntos en el puntaje obtenido por los alumnos que concurren a la misma. Es importante marcar, que el impacto del nivel socioeconómico promedio de la escuela (entre 1,54 y 1,95 puntos de rendimiento educativo por punto porcentual de nivel socioeconómico promedio de la escuela) es sensiblemente superior al efecto del nivel socioeconómico propio del alumno (entre 0,66 y 0,84 puntos según el cuantil), avalando la eficacia de una mayor inclusión socioeconómica dentro de las escuelas argentinas para disminuir las brechas de puntajes entre alumnos.

Por otro lado, en el gráfico A2 del Anexo se observa que los coeficientes obtenidos para el modelo multinivel en la media resultan similares a los surgidos de la regresión cuantifica para la mediana $\tau = 0,5$. Como se observa estos valores difieren de los valores obtenidos en las secciones extremas de la distribución ($\tau = 0,10, \tau = 0,90$). Por ende la utilización de la media o de la mediana como única medida de análisis omite información valiosa del efecto de estas variables sobre la distribución de puntajes.

Finalmente, el intercepto aleatorio que capta el efecto de asistir a una escuela determinada sobre el rendimiento académico ha disminuido luego de haber controlado por el efecto de características individuales y del contexto socioeconómico de la escuela, pero continúa siendo significativo y superior para el cuantil inferior. Es posible afirmar entonces que subsisten diferencias significativas entre el rendimiento promedio alcanzado por los centros, atribuibles a prácticas o métodos de enseñanza aplicadas dentro de las aulas que no llegan a ser captadas por el modelo propuesto.

Cuadro N°4 – Resultado modelo estimado N°2

			Cuantiles													
	Media		0,1	0,25	0,4	0,5	0,6	0,75	0,9							
(Intercept)	432,33	***	352,31	***	394,54	***	418,7	***	433,99	***	450,39	***	476,2	***	522,34	***
Genero	-15,27	***	-26,95	***	-20,4	**	-18,6	***	-17,01	***	-14,99	*	-11,76		-11,26	
Repitió	-42,31	***	-48,83	***	-46,72	***	-43,03	***	-44,16	***	-42,87	***	-43,72	***	-40,5	***
Escs	0,72	***	0,84	*	0,7	*	0,71	**	0,7	***	0,74	***	0,66	**	0,74	*
Joyread	1,21	***	0,94	*	0,95	***	1,01	***	1,1	***	1,08	***	1,19	***	1,34	***
Menor	-0,45	***	-0,28		-0,4		-0,44	*	-0,44	*	-0,5	***	-0,49	***	-0,57	***
Cstrat	0,66	***	0,56	(*)	0,69	***	0,74	***	0,69	***	0,67	***	0,67	***	0,65	**
Scscs	1,53	***	1,57	(*)	1,59	**	1,78	***	1,69	***	1,54	***	1,72	***	1,95	***
Gestión	8,62		19,77		17,19		12,97		11,99		10,66		8,89		-3,83	***
Scmatedu	0,51		0,69		0,84	(*)	0,37		0,47		0,4		0,36		0,03	
repitio.sc	-23,65		-18,84		-22,76		-32,86		-28,96		-23,13		-24,46		-35,75	(*)
Scale			11,61		21,24		25,7		26,4		25,37		20,52		10,98	
Efecto aleatorio			1.736		1.428		1.301		1.261		1.235		1.028		1.054	
Log			25.787		25.207		14.970		24.881		24.884		25.036		25.516	
AIC			51.599		50.441		49.920		49.787		49.793		50.097		51.059	
Varianza Total Explicada	47%		29%		34%		38%		38%		37%		35%		34%	
Varianza Intra-escuela Explicada	79%		22%		22%		22%		21%		22%		22%		25%	
Varianza Entre escuela Explicada	18%		59%		64%		68%		67%		67%		71%		72%	

Nota: (*) p<.10; *p<.05; **p<.025;***p<.01
Fuente: Elaboración propia en base a PISA 2009

5. Discusión de resultados

Los resultados obtenidos en esta primera aplicación del método HLM por cuantiles a muestras educativas confirman la importancia de estimar el impacto de las características escolares sobre la desigualdad en el rendimiento educativo considerando la presencia de heterogeneidad.

En primer lugar se observó que las características de la escuela media argentina poseen diferente importancia relativa sobre la dispersión de puntajes entre alumnos según el cuantil analizado. Las escuelas asumen mayor responsabilidad sobre la desigualdad de puntajes entre alumnos con rendimientos cercanos al promedio de la misma, disminuyendo para los alumnos con rendimientos extremos. Adicionalmente, el comportamiento decreciente del intercepto aleatorio a lo largo de la distribución de puntajes confirma que las escuelas argentinas tienen una brecha de puntajes mayor entre sus peores alumnos que la existente entre sus mejores alumnos.

Con respecto al análisis de las características individuales del alumno, se confirma la existencia de relaciones heterogéneas con el puntaje alcanzado. El impacto de variables sensibles para el desempeño académico como el nivel socioeconómico y cultural (“escs”) y la repitencia del año escolar (“repitió”) es más fuerte para los alumnos de peores rendimientos en comparación con los de rendimiento cercanos al promedio. Esta

información resulta valiosa para el diseño de políticas educativas que busquen impulsar el desempeño de los alumnos con menores puntajes y sólo puede ser captada por estudios que consideren las particularidades de cada tramo de la distribución.

Finalmente, la incorporación de variables del nivel escolar al modelo multinivel por cuantiles confirmó la existencia de un efecto de pares relacionado al contexto socioeconómico de la escuela, que supera ampliamente al impacto del nivel socioeconómico propio de la familia del alumno.

En suma, el análisis multinivel por cuantiles permitió revelar información referida a factores influyentes sobre el rendimiento educativo no disponible en estudios focalizados en la media. Este trabajo presenta evidencia sobre la necesidad de un tratamiento diferenciado de la desigualdad educativa, dado que políticas enfocadas en indicadores a nivel escuela tendrán mayor eficiencia en disminuir la desigualdad de puntajes entre alumnos de rendimiento promedio, perdiendo efectividad para acortar brechas entre alumnos de menores rendimientos.

Bibliografía

- Adams, R., & Wu, M. (2002). *PISA 2000 Technical Report*.
- Brik, A., & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, California .
- Cervini, R. (2008). Comparando la inequidad en los logros escolares de la educación primaria y secundaria en Argentina: un estudio multinivel.
- Cervini, R. (2002). Desigualdades en el logro académico y reproducción cultural en Argentina. Un modelo de tres niveles. .
- Coleman, J., Campbell, E., Hobson, C., Mcpartland, J., Mood, A., & Weinfeld, F. (1966). *Equality of Educational Opportunity*. Washington, D.C : U.S Government Printing Office.
- Decándido, G. (2011). Factores que afectan las competencias de los alumnos argentinos en PISA 2009. Un estudio empírico de dos niveles con efecto interacción.
- Decándido, G., Gigena, M., Giuliodori, R. F., & Gertel, H. R. (2011). Comportamiento de poblaciones escolares heterogeneas en las pruebas SERCE 2006 de matemáticas.
- Duarte, J., Bos, M. S., & Moreno, M. (2009). Inequidad en los aprendizajes escolares en América Latina. Análisis multinivel del SERCE según la condición socioeconómica de los estudiantes.
- Fenske, N. (2012). *Structured additive quantile regression with applications to modelling undernutrition and obesity of children*. Munich.
- Fenske, N., Fahrmeir, L., Hothorn, T., Rzehak, P., & Höhle, M. (2013). Boosting Structured Additive Quantile Regression for Longitudinal Childhood Obesity Data. *The International Journal of Biostatistics*. Volume 9, Issue 1 .
- Fielding, A., Yang, M., & Goldstein, H. (2003). Multilevel ordinal models for examination grades. *Statistical Modelling* , 127-153.
- Gasparini, L., Jaume, D., Serio, M., & Vazquez, E. (2011). *La segregación escolar en Argentina*.
- Gelman, A., & Hill, J. (2009). *Data analysis using regression and Multilevel/Hierarchical Models*. New York, Cambridge University Press .
- Geraci, M. (15 de 09 de 2012). Package 'lqmm'.
- Geraci, M., & Bottai, M. (2011). *Linear quantile mixed models*.
- Geraci, M., & Bottai, M. (2013). *Linear quantile mixed models*.
- Geraci, M., & Bottai, M. (2007). Quantile regression for logitudinal data using the asymmetric Laplace distribution. *Biostatistics* , 8(1), 140-154.
- Geraci, M., & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, 8, 1 , 140-154.

- Geraci, M., & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* , 140-154.
- Gertel, H., Giulodori, R., Vera, M. L., & Bustos, G. (2010). El efecto fijo de las regiones en la regresión por cuantiles del desempeño académico de los estudiantes al término de la educación secundaria de Argentina.
- Gigena, M., Vera, M. L., Giulodori, R. F., & Gertel, H. R. (2011). The potential of quantile regression to comparative analysis of heterogeneity and achievement in schools across countries. *CIES*.
- Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources* , nº 14 (3), p.p. 351-388.
- Hanushek, E. A. (2007). Education Production Functions. En *Palgrave Encyclopedia*.
- Hanushek, E. A. (1971). *Teacher Characteristics and gains in student achievement: estimation using micro-data*. *American Economic Review* 61, 280-288.
- Hindriks, J., Verschelde, M., & Rayp, G. (2010). Ability tracking, social segregation and educational opportunity: evidence from Belgium.
- Hindriks, J., Verschelde, M., Rayp, G., & Schoors, K. (2010). Ability tracking, social segregation and educational opportunity: evidence from Belgium.
- Jaume, J. D. (2013). *Un estudio sobre el incremento de la segregación escolar en Argentina*.
- Kjetil H. Løland, E. S. (s.f.). The Association between Progression of Atherosclerosis and the Methylated Amino Acids Asymmetric Dimethylarginine and Trimethyllysine.
- Koenker. (2004).
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* (91), 74-89.
- Koenker, R. (1978). *Regression quantiles* (Vol. 46).
- Koenker, R., & Bassett, G. (1982). An Empirical Quantile Function for Linear Models with iid Errors. *Journal of the American Statistical Association* , 407-15.
- Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica* , 33-50.
- Krueger, N. S., & Formichella, M. M. (2013). El Impacto de la Segregación Social del Alumnado sobre los Aprendizajes en la Escuela Secundaria Argentina. *Congreso 2013 de la Asociación de Estudios Latinamericanos* .
- Krueger, N., & Formichella, M. M. (2013). El Impacto de la Segregación Social del Alumnado sobre los Aprendizajes en la Escuela Secundaria Argentina.
- Krueger, N. S., & Formichella, M. M. (2013). El impacto de la Segregación Social del Alumnado sobre los Aprendizajes en la Escuela Secundaria Argentina.

- Lamarche, C. (2006). *Robust penalized quantile regression estimation for panel data*. Manuscrito.
- Liu, Y., & Bottai, M. (2009). Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics* , 5(1), 28.
- Manski, C. F. (1993). Identification of endogenous social effects - the reflection problem. *Review Of Economic Studies* 60 (3) , 531-542.
- Marchionni, M., Vazquez, E., & Pinto, F. (2012). *Desigualdad educativa en la Argentina. Análisis en base a los datos PISA 2009*.
- Monette, G., Qing, S., & Kwan, E. (2001). A First Look at Multilevel Models. *Institute for Social Research* , 96.
- OCDE. (2004). *PISA 2003: Manual de análisis de datos: Usuarios de SPSS*.
- OCDE. (2010). *PISA 2009 Results: What makes a school successful? Resources, Policies and Practices*.
- OCDE. (2010). *PISA 2009 Results: Overcoming Social Background-Equity in Learning Opportunities and Outcomes (Volume II)* .
- OCDE. (2009). *PISA Data Analysis Manual SPSS*.
- OCDE. (2009). *PISA Data Analysis Manual SPSS*.
- OECD. (2004a). *PISA 2003: Manual de análisis de datos: Usuario de SPSS*. París: OECD.
- OECD. (2004b). *PISA 2003: Manual de análisis de datos: Usuarios de SPSS*. Paris: OECD.
- OECD. (2010). *PISA 2009 Results: What makes a school successful? - Resources, Policies and Practices (Volumen IV)*. OECD.
- OECD. (2010a). *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volum IV)*. OECD, Publishing.
- Rangvid Schindler, B. (2003). *Educational Peer Effects Quantile Regression Evidence form Denmark with PISA 2000 data*. Caponague, Dinamarca: Institute of Local Government Studies.
- Raudenbus, S., & Willms, D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics* , 20(4), 307-335.
- Rivas, A., Vera, A., & Bezem, P. (2010). *Radiografía de la educación argentina*.
- Rowe, K., Hill, P., & Holmes-Smith, P. (1995). Methodological Issues in educational Performance and School Effectiveness Research: A Discussion with Worked Examples. *Australian Journal of Education* V. 39 Nº 3 , 217 - 248.
- Sosa Escudero, W., Giovagnoli, P. I., & Porto, A. (2009). The effects of individual characteristics on the distribution of college performance. *Económica* .

Steenbergen, M. R., & Jones, B. S. (2002). *Modeling multinivel data structures*. American Journal of Political Science, Vol. 46, N° 1. Pag 218-237.

Velez Reyes, F. A. (2011). *Robustez a Distribución Para Regresión Cuantílica en Datos Longitudinales*. Universidad Nacional de Colombia.

Western, B. (1998). *Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach*. American Journal of Political Science 42: 1233-1259.

Tabla A.1. Descripción de las variables utilizadas

Etiqueta	Variable	Signo Esperado	Significado
PV1READ PV2READ PV3READ PV4READ PV5READ	Valores Plausibles en lectura		Números aleatorios extraídos de una distribución de puntajes en lectura que pueden ser asignados razonablemente a cada individuo.
Nivel Alumno			
genero	Género	(-)	Variable dummy que asume el valor 0 si el niño es de sexo femenino, y 1 en caso de que sea de sexo masculino.
reptio	Repitió de grado	(-)	Variable dummy que asume el valor 0 si el estudiante se encuentra en décimo grado o más arriba, y 1 en caso de que el estudiante esté en noveno grado o más abajo.
memor	Memorización	(-)	Índice derivado de la frecuencia con la cual los estudiantes han hecho las siguientes acciones al momento de estudiar: i) intentar memorizar cada tema cubierto en el texto, ii) intentar memorizar tantos detalles como sea posible, iii) leer el texto tantas veces que puede ser recitado, y iv) leer el texto una y otra vez. Valores más altos en el índice indican mayor importancia hacia esta estrategia de estudio. La variable fue re-escalada en un rango de 0/100 y centrada para el alumno promedio de la muestra para el país.
cstrat	Estrategias de control	(+)	Índice derivado de la frecuencia con la cual los estudiantes han hecho las siguientes acciones al momento de estudiar: i) comenzar por comprender exactamente lo que tiene que aprender, ii) chequear si entiende lo que está leyendo, iii) intentar entender cada concepto, aún cuando realmente no lo tiene que entender, iv) asegurarse de recordar la mayoría de los puntos importantes del texto, y v) cuando no entiende algo, buscar información adicional para clarificar esto. Valores más altos en el índice revelan mayor importancia hacia esta estrategia de estudio. La variable fue re-escalada en un rango de 0/100 y centrada para el alumno promedio de la muestra para el país.
joyread	Gusto por la lectura	(+)	El índice mide el gusto del estudiante por la lectura. Se deriva de las respuestas de los estudiantes en ocho preguntas sobre sus hábitos de lectura, el disfrute y actitud hacia la actividad. La variable fue re-escalada en un rango de 0/100 y centrada para el alumno promedio de la muestra para el país.
escs	Índice de status socio-económico y cultural	(+)	Índice que captura el status socio-económico y cultural de la familia del alumno. Incluye el status ocupacional de los padres, su nivel educativo, y diferentes posesiones en el hogar: riqueza, posesiones culturales y recursos educativos. La variable fue re-escalada en un rango de 0/100 y centrada para el alumno promedio de la muestra para el país.
Nivel Escuela			
scscs	Nivel socio-económico de la Escuela	(+)	Índice calculado como el promedio del "escs" por escuela, captura el entorno socio-económico y cultural de la escuela a la que asiste el alumno, donde valores más altos en el índice indican un mejor entorno socio-económico y cultural del colegio. La variable fue re-escalada en un rango de 0/100 y centrada en el promedio de la muestra para el país.
scmatedu	Materiales Educativos	(+)	Índice derivado de siete ítems que miden la percepción del director sobre algunos factores que pueden dificultar la instrucción en su escuela: i) escasez o inadecuado equipamiento del laboratorio de ciencias, ii) escasez o inadecuados materiales de instrucción, iii) escasez o inadecuadas computadoras para la instrucción, iv) lenta o inadecuada conexión a internet, v) escasez o inadecuado software de computación, vi) escasez o inadecuados materiales en la biblioteca, y vii) escasez o inadecuados recursos audio-visuales. Todos los ítems fueron re-escalados por PISA, de modo tal que un mayor valor del índice indica mejor calidad de recursos escolares. La variable fue re-escalada en un rango de 0/100 y centrada en el promedio de la muestra para el país.

screpitio	Proporción de repitentes en la escuela	(-)	Variable que captura el promedio de repitentes por escuela
gestion	Tipo de gestión de la escuela	(+)	Variable dummy que clasifica a las escuelas como públicas o privadas de acuerdo a si la gestión está a cargo de una agencia pública o privada.

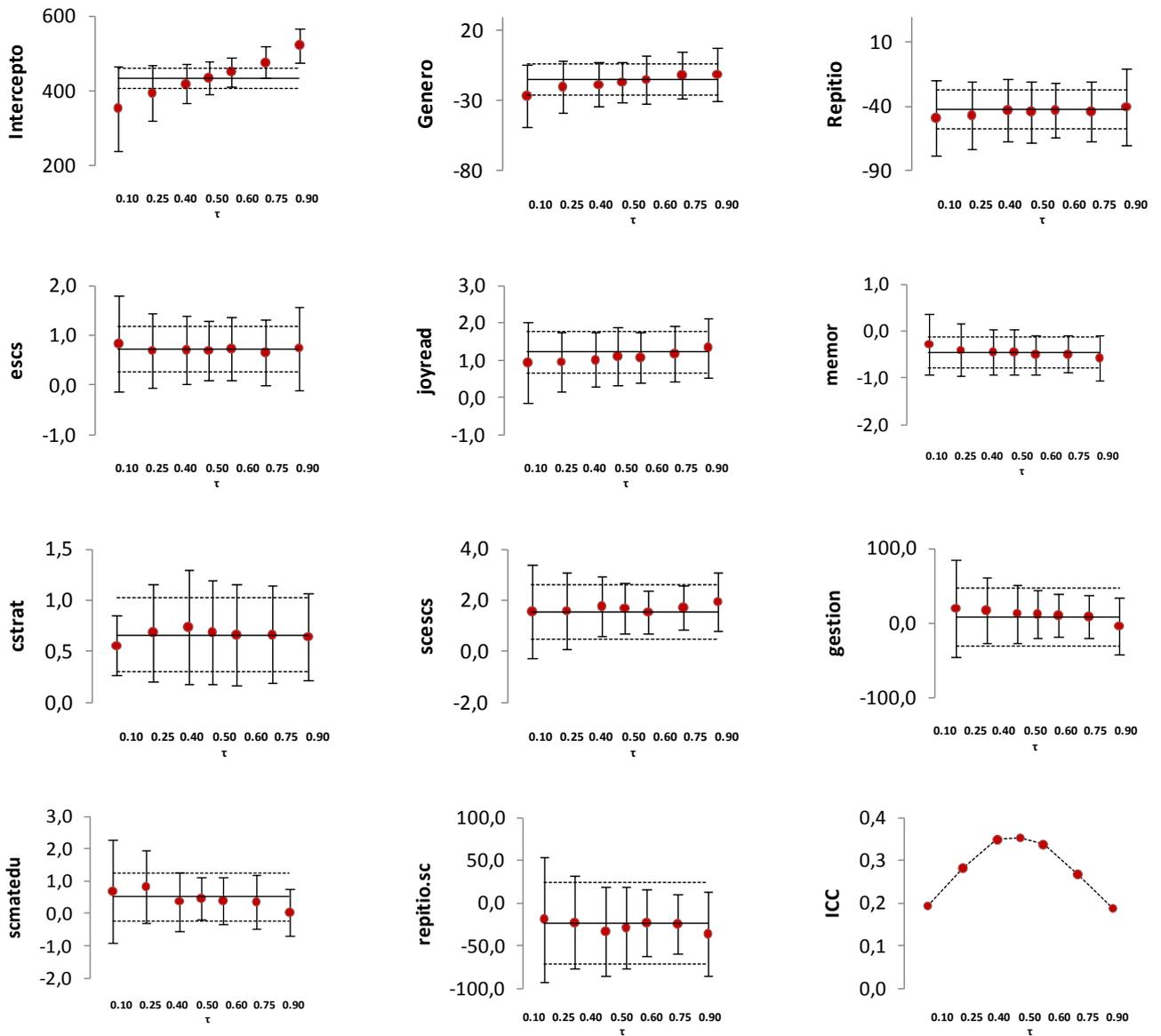
Fuente: Elaboración propia en base a PISA 2009

Tabla A.2. Estadísticos Descriptivos de las variables utilizada

Variable	N	Media	Desv. Estándar	Mínimo	Máximo
genero	4680	0.46	0.5	0	1
repitio	4641	0.37	0.48	0	1
joyread	4542	45.59	10.92	0	100
escs	4612	55.84	14.94	0	100
memor	4590	58.38	17.1	0	100
cstrat	4582	62.5	16.15	0	100
scscs	4680	44.69	19.5	0	100
screpitio	4641	0.37	0.35	0	1
scmatedu	4658	52.15	24.15	0	100
gestion	4680	0.36	0.48	0	1

Fuente: Elaboración propia sobre la base de PISA 2009

Gráfico A.1. Coeficientes para el modelo completo, por cuantiles y en la media.



Fuente: Elaboración propia sobre la base de PISA 2009