

# DIFERENCIAS EN RESULTADO EDUCATIVO DE LAS REGIONES ESPAÑOLA SEGÚN PISA-2006

J. Ignacio García Pérez  
*Universidad Pablo de Olavide, FCEA y FEDEA*

Marisa Hidalgo Hidalgo  
*Universidad Pablo de Olavide*

J. Antonio Robles Zurita  
*Fundación Centro de Estudios Andaluces*

## Resumen

El objetivo de este trabajo es analizar en qué medida las diferencias de resultado educativo detectadas entre las regiones españolas participantes en PISA-2006 son explicadas por las variables individuales, familiares y de colegio disponibles. Cuando se estima el resultado educativo de los estudiantes teniendo en cuenta los distintos tipos de variables de forma conjunta se obtiene como resultado que las variables familiares socioeconómicas y culturales, el interés de los estudiantes, y el porcentaje de repetidores explican gran parte de estas diferencias. En este trabajo se estima un modelo de Regresiones Alternantes que permite controlar por la selección endógena de los alumnos *No Repetidores* respecto a los *Repetidores*. Tomando como referencia a la Comunidad de Andalucía, tanto para el grupo de *Repetidores* como para el de *No Repetidores*, se estima que las regiones de Aragón, Asturias, Cantabria, Castilla y León, Galicia y La Rioja obtienen mejor resultado educativo una vez que se controla por características observables e inobservables. También para ambos grupos, los alumnos catalanes que hablan catalán y del resto de España no son diferentes a los alumnos andaluces. Las regiones que para el grupo de *No Repetidores* obtienen peor resultado a Andalucía son Cataluña si hablan castellano y Navarra y País Vasco si hablan eusquera. Mientras que para los *Repetidores* las regiones de peor rendimiento educativo son Navarra si hablan eusquera y País Vasco si hablan castellano

Palabras clave: PISA, Rendimiento educativo, Comunidades Autónomas, Sistema educativo

## 1 INTRODUCCIÓN Y MOTIVACIÓN

El objetivo de este trabajo es analizar en qué medida las diferencias de nivel educativo detectadas entre regiones españolas participantes en PISA-2006 son explicadas por las variables de dotaciones iniciales (variables que les viene dadas al sistema educativo y las políticas públicas no pueden hacer mucho para cambiarlas), individuales y de colegio. En España algunas Comunidades<sup>1</sup> ampliaron la muestra de forma voluntaria para tener datos representativos de su región, lo que permite la realización de un análisis regional. El resto de CCAA sin muestra ampliada forman un grupo denominado por Resto de España en lo que sigue.

---

<sup>1</sup> Andalucía, Asturias, Aragón, Cantabria, Castilla y León, Cataluña, Galicia, La Rioja, Navarra y País Vasco.

Como se puede observar en la Tabla 1 del apéndice la media española en rendimiento educativo se sitúa en 488 puntos frente a una media para el conjunto de la OCDE de 500 puntos. Sin embargo, existen diferencias importantes entre regiones que se van a señalar para cada una de ellas. Andalucía es, entre las regiones con muestra ampliada, la región española con el rendimiento educativo más bajo con 474 puntos de nota media. En el otro extremo tenemos a La Rioja y Castilla y León que presentan la característica inversa: el rendimiento es 46 puntos superior a Andalucía y en 20 puntos superior a la media de la OCDE. Así mismo, hay comunidades que se sitúan en lugares de bajo rendimiento educativo como es el caso de Cataluña y las regiones sin muestra ampliada que aparecen agrupadas bajo la denominación “Resto de España”. El resto de comunidades se sitúan en una zona intermedia que va desde los 495 puntos de País Vasco y los 513 de Aragón.

Desde el sector académico ha habido otros trabajos en los que se han utilizado los datos de PISA de España para tratar las diferencias regionales. Carabaña (2008) realiza un análisis de las diferencias regionales relacionando éstas con distintas variables explicativas como el estatus socioeconómico y cultural de las familias, el estatus inmigrante de los alumnos, los recursos materiales y gasto educativo, así como variables que indican características de los colegios. El autor destaca la importancia del estatus socioeconómico en el resultado académico medio de las regiones describiéndolo como el único factor que tiene una influencia importante y consistente sobre las diferencias de aprendizaje entre países y regiones. Ciccone y García-Fontes (2008) hacen un análisis de la importancia de la educación de los padres y el estatus inmigrante en las comparaciones entre Cataluña y el resto de regiones españolas, por un lado; y España y el resto de países europeos, por otro. Respecto a las comparaciones regionales se llega a la conclusión de que el peor resultado educativo mostrado por Cataluña mejora tan sólo ligeramente cuando se controla por el nivel educativo de los padres. También se encuentra que el mayor nivel de inmigración que tiene Cataluña respecto al resto de regiones españolas no es un factor importante para explicar el peor resultado educativo que esta consigue en PISA. Finalmente, Cordero, Crespo, y Santín (2009) hacen un análisis de las diferencias entre Comunidades Autónomas y aplican un Análisis Envolvente de Datos tratando de explicar qué variables están detrás de una mayor eficiencia de los inputs educativos (entorno familiar; recursos de los colegios; y características de los compañeros). Según sus palabras, los autores llegan a la conclusión de que: “..Galicia, La Rioja y Aragón ostentan en este orden los sistemas educativos más eficientes mientras que “Resto de España”, País Vasco y Cataluña son los más ineficientes”.

En el presente trabajo se quiere completar la literatura existente a través de diversas aportaciones. Se ha realizado un análisis multivariante teniendo en cuenta una serie de variables disponibles en la base de datos de PISA-2006 adicionales a las utilizadas en los anteriores trabajos, tales como: las horas de clases de ciencias en el colegio, la actitud de los estudiantes, el

idioma que el alumno habla en casa en aquéllas regiones en las que hay lengua regional, y el porcentaje de chicas en el colegio, todas ellas significativas. La principal aportación de este trabajo a la literatura es la propuesta de una forma de tratar los datos de PISA-2006 cuya población objetivo, aún siendo todos los estudiantes nacidos en el año 1990, está formada por alumnos matriculados en distintos cursos académicos. Dado que en España los alumnos deben empezar su vida escolar según el año de nacimiento, todos los alumnos que forman parte de la muestra de PISA-España deberían estar matriculados en 4º de la ESO, los estudiantes de la muestra que están en un curso inferior son *Repetidores*. A través de la estimación de un Modelo de Regresiones Alternantes (MRA) o *Endogenous Switching Regresion Model*, en términos anglosajones, se controla por la selección endógena existente entre los alumnos *No Repetidores* y *Repetidores* y permite obtener resultados diferentes para ambos grupos en cuanto a las diferencias regionales.

Con la estimación de este modelo se puede obtener cual es el efecto en el resultado educativo de pertenecer a una determinada región a igualdad en el resto de factores individuales, familiares y de colegio. Los resultados obtenidos indican claramente que las variables socioeconómicas y culturales, las horas de estudio de ciencias en el colegio, la actitud de los estudiantes y el porcentaje de repetidores explican gran parte del peor comportamiento de los alumnos andaluces. Así, cuando se comparan alumnos con características similares, las diferencias regionales se reducen respecto a las diferencias observadas. Tomando como referencia a Andalucía las diferencias se acortan totalmente con otras regiones de España, tales como Cataluña y el País Vasco. Sin embargo, Andalucía sigue estando peor que otras zonas, como Castilla y León y La Rioja, aunque la diferencia se reduce considerablemente. Los resultados indican que las diferencias regionales para los alumnos *repetidores* son inferiores que para los alumnos *no repetidores*.

La estructura del estudio es la siguiente. En primer lugar, se explicará brevemente en qué consiste la base de datos utilizada<sup>2</sup> y las variables analizadas. En la sección 3 se explica la metodología utilizada en el estudio. Por último, los resultados y las conclusiones se presentan en la sección 4.

## 2 Datos

---

<sup>2</sup> El lector interesado en conocer mejor la metodología de PISA puede acceder a las publicaciones originales (OCDE, 2008) y a otras publicaciones como por ejemplo el ofrecido por el Instituto de Evaluación del Ministerio de Educación y Ciencia (MEC, 2007), que tratan con detalle dicha encuesta.

## 2.1 *Muestra española de PISA-2006*

La muestra de PISA tiene un diseño estratificado en dos etapas. En primer lugar se seleccionan las escuelas con estudiantes de 15 años y en segundo lugar, se seleccionan a los estudiantes dentro de cada escuela.<sup>3</sup> En este trabajo utilizamos la muestra de PISA correspondiente a España y sus regiones con muestra ampliada. Las regiones con muestra representativa vienen recogidas en la Tabla 1. Como se puede ver, la muestra para el total de España alcanza los 19.604 estudiantes y las 686 escuelas. En cuanto a las muestras regionales observamos que la mayoría tienen aproximadamente 1.500 estudiantes aunque hay que destacar el caso del País Vasco, región para la cual la muestra asciende a casi 3.929 alumnos. Denominamos “Resto de España” al conjunto de regiones que no tienen muestra ampliada.

Una particularidad de los estudiantes de la muestra utilizada es que todos nacieron en el año 1990. Además, en España, de acuerdo a la normativa educativa, todos los alumnos nacidos en el mismo año deben comenzar su vida escolar también en el mismo curso académico. Por ello, los estudiantes de la muestra española de PISA deben estar matriculados en 4º de ESO, si no han repetido por cualquier causa. Entonces, los alumnos de la base de datos que están matriculados en ese curso son los *No Repetidores* y los estudiantes que se encuentran matriculados en un curso inferior, 3º o 2º de ESO son *Repetidores*. Esta distinción es crucial para entender el tratamiento que se hace en este trabajo al problema de selección endógena que existe entre los alumnos que forman parte de ambos grupos.

## 2.2 *Variables utilizadas*

La variable dependiente a analizar es la **puntuación obtenida en el examen de ciencias** por cada alumno. En PISA, debido a la posible existencia de errores de medida, la puntuación de cada estudiante se calcula en base a un procedimiento establecido en base a la *Item Response Theory*. Así, cada alumno tiene cinco *valores plausibles* de la puntuación en ciencias. No obstante, para este trabajo se utiliza una única puntuación descrita en la base de datos de PISA como *PVISCIE*. Los mismos resultados de este trabajo se han replicado para los otros cuatro *valores plausibles* sin que hayan variado los principales resultados.

En cuanto a las variables explicativas se han clasificado en los siguientes grupos:

### 1) Dotaciones iniciales.

Serán los factores de entorno familiar que al sistema educativo le vienen dados tales como las características socioeconómicas de la familia. Por tanto, en la medida en que dichas variables influyan en el resultado educativo de los alumnos, producirán diferencias regionales que no pueden ser atribuidas a un distinto funcionamiento de los sistemas educativos. Las

---

<sup>3</sup> Para más información sobre el diseño muestral de PISA ver OECD (2006).

variables que se incluyen en este grupo proceden del cuestionario que los alumnos de PISA-2006 contestaron de forma individual.

Dentro de las variables de dotaciones iniciales se incluye el **estatus educativo y ocupacional de los padres**. La educación de los padres y madres puede ser alta, si la titulación es de secundaria o mayor, o baja, para estudios primarios o inferiores. Para el análisis econométrico se han incluido las distintas interacciones posibles entre la educación del padre y la madre. Los padres de los alumnos están categorizados según su ocupación en alta o baja, según sea un trabajador de tipo administrativo-intelectual (o *white collar*) o tipo manual (o *blue collar*) respectivamente. Para el estatus laboral de las madres se incluyen las mismas categorías añadiendo un grupo de amas de casa<sup>4</sup>, para aquellas madres que se dedican principalmente a las tareas del hogar. La descripción de las variables de educación y ocupación de los padres se encuentra en las Tablas 2 y 3 respectivamente. La región con menos alumnos de madres y padres con titulación universitaria es Andalucía, tan sólo un 12% y el 13% respectivamente. Por el contrario, la región que más padres con el mismo nivel educativo tiene es País Vasco un 26% tanto de madres como de padres universitarios. El porcentaje de alumnos con madre de ocupación alta (20%) en Andalucía es menor que en cualquier otra comunidad autónoma, siendo Cataluña la región en donde dicho porcentaje es mayor (38%). En cuanto a la ocupación de los padres las regiones con menos padres de ocupación alta son Andalucía y Castilla y León con un 30% y 33% respectivamente, mientras que las regiones con mayor porcentaje de padres de alta ocupación son País Vasco y Cataluña con un 44% y 43% respectivamente.

Se han clasificado los alumnos según un índice categórico de **cultura educativa del hogar**. En la base de datos de PISA existe un índice de posesiones culturales de la familia, denominada CULTPOSS, sin embargo éste no incluye el número de ordenadores que tiene el estudiante en casa. Por ello, se ha realizado un análisis cluster para estas dos variables. Para llevar a cabo este análisis el índice cultural de PISA y la variable que indica el número de ordenadores se categorizaron en tres grupos dejando aproximadamente el mismo número de individuos en cada uno. El resultado es la creación de una variable de cultura educativa que distingue entre alumnos con baja cultura educativa, media, y alta. En la Tabla 3 se observa que Andalucía es la región que menos porcentaje de alumnos pertenecen a un hogar con una cultura educativa alta (un 30%) mientras que dicho porcentaje es del 50% en Aragón, la región con más alumnos con cultura educativa alta.

También se tuvieron en cuenta otras características de los alumnos que podían, en principio, influir en el resultado educativo. Fueron las siguientes: **nacionalidad y; lengua que normalmente habla en casa**. En cuanto a la nacionalidad del alumno, se ha apuntado en otros trabajos su relación con el resultado educativo. Para el caso español Zinovyeva, Felgueroso y Vazquez (2009) utilizan la base de datos de PISA y ponen de manifiesto el peor resultado

---

<sup>4</sup> Este grupo tiene interés únicamente para las madres puesto que el número de padres en esta categoría es insuficiente para realizar un análisis adecuado.

educativo para los inmigrantes incluso a igualdad de características socioeconómicas con los nativos. Estos mismos autores, no encuentran evidencia de que haya diferencias entre aquéllos alumnos que hablan una lengua nacional o no nacional, una vez que se controla por características como el tiempo que llevan los inmigrantes residiendo en España, sin embargo en el presente trabajo se ha querido diferenciar a los alumnos de aquellas regiones con dos idiomas oficiales según hablan la lengua regional o nacional. En la Tabla 3, se describe que las regiones que menos inmigrantes tienen son Andalucía, Asturias, Castilla y León y Galicia con menos de un 4% de alumnos inmigrantes. El caso contrario lo presentan Cataluña y Resto de España cuyo porcentaje de alumnos inmigrantes es de un 10%.

## 2) Variables individuales.

Son aquellas variables que son características propias del alumno. Aquí se incluyen características como el **sexo** del alumno o el **mes de nacimiento** así como las variables que se explican a continuación: frecuencia de uso del ordenador para usos no educativos; horas de estudio en ciencias en el colegio; la actitud ante la ciencia; y el carácter repetidor o no de los estudiantes.

Los alumnos indicaron la **frecuencia de uso de los ordenadores** para jugar y para bajarse música de Internet. De manera que han sido clasificados entre los que realizan estas actividades casi diariamente y los que lo hacen con menor frecuencia.

En este trabajo se tienen en cuenta las horas **semanales de ciencias** que el alumno da **en el colegio** habiéndose distinguido tres categorías: los alumnos que dan menos de 2 horas; aquéllos que tienen entre 2 y 4 horas; y los alumnos que dan más de 4 horas semanales. Como se puede ver en la Tabla 4 la mayor parte del alumnado de cada región tiene entre 2 y 4 horas de ciencias a la semana. Sin embargo, esto no es así para Navarra, región en la que la mayor parte de sus alumnos tienen más de 4 horas de ciencias, un 44%. Por lo que respecta a Cataluña también presenta un caso particular, es la región con mayor porcentaje de alumnos que tienen entre 2 y 4 horas a la semana con un 65%, mientras que tan sólo un 19% de los alumnos tienen más de 4 horas de ciencias a la semana.

Para medir la **actitud ante las ciencias** se ha realizado un índice con cada una de las preguntas del cuestionario que reflejan el interés de los estudiantes ante la misma. Dado el elevado grado de correlación entre los índices, se procedió a realizar un análisis *cluster* que permitió clasificar a los alumnos en aquéllos con actitud alta, media y baja. Diversos estudios dentro de la literatura sobre economía de la educación apuntan a la importancia de esta variable, en concreto De Fraja, Oliveira y Zanchi (2005) utiliza datos de Gran Bretaña del *National Child Development Study* y ponen énfasis en la forma en que la actitud o esfuerzo de los alumnos influye en el resultado educativo, encontrando que hay variables como el propio esfuerzo de los padres, de los colegios, y otras variables, que influyen en el logro educativo indirectamente a través de su influencia en la actitud de los alumnos. En la Tabla 4 se observa la distribución de

la variable actitud de los alumnos por regiones. A la vista de la misma, no parece haber grandes diferencias regionales en cuanto a esta variable. Por un lado, la región que más alumnos con actitud alta tiene es Castilla y León con un 30% de alumnos y la región con menor porcentaje en dicha categoría es Andalucía con un 24%.

Otra variable incluida en este grupo es aquella que indica si el alumno es **Repetidor** o **No Repetidor**. En la base de datos nos encontramos con alumnos que están en 1º de Bachillerato, 4º, 3º, 2º y 1º de ESO. Dado que todos los alumnos son de la misma edad (15 años) y nacieron en el mismo año (1990), todos los alumnos deberían estar matriculados en el mismo curso académico, 4º de ESO. Sin embargo, en la base de datos se encuentran también alumnos que están matriculados en 3º y 2º de ESO, que se han diferenciado como *Repetidores*. Es un hecho preocupante el que para la gran mayoría de las CCAA el porcentaje de alumnos que han repetido al menos un curso a los 15 años es de más del 30% (Tabla 4). Andalucía es la región con mayor porcentaje de repetidores con un 48% frente al 24% del País Vasco, la región que menos repetidores tiene.

Las variables de actitud ante las ciencias y el carácter repetidor de los estudiantes pueden ser problemáticas de clasificar como variables exógenas, ya que pueden influir en el resultado educativo y, a la vez, verse influenciadas por el resultado educativo previo del alumno. La variable que más claramente presenta un problema de endogeneidad es aquella que diferencia entre *Repetidores* y *No Repetidores*. Por un lado, el hecho de repetir influye en el resultado educativo del alumno, ya que los alumnos repetidores han llegado a estudiar un contenido curricular inferior a aquéllos alumnos que no repiten. Y por otro lado, el hecho de que un alumno haya repetido se debe al propio resultado educativo previo obtenido. Entonces, la diferencia en resultado educativo encontrada entre alumnos Repetidores y No Repetidores no se debe únicamente al mero hecho de repetir sino también al hecho de que cada grupo presenta características diferentes que influyen en el resultado educativo. En el apartado de metodología se explica el Modelo de Regresiones Alternantes que controla por la selección endógena de esta variable.

### 3) Variables de colegio.

Son aquellas que presenta un colegio y que serán las mismas para todos los alumnos del mismo. La procedencia de estas variables se encuentra en el cuestionario que rellenaron los directores de los centros educativos. Estas variables son: el porcentaje de chicas en el colegio; la titularidad; la presión de los padres; el nivel educativo medio de los padres; y el tamaño de la clase de lengua.

El **porcentaje de chicas** que hay en el colegio es una variable que diferencia entre los colegios en los que hay más del 50% de chicas y el resto de colegios. Esta variable es también utilizada por Calero y Escardíbul (2007) para los datos de España en PISA 2003 encontrándose un efecto positivo y significativo para los colegios con mayoría de chicas. También

encontramos evidencia de este efecto positivo en Hoxby (2000). En la Tabla 5 se observa la distribución de esta variable destacando Cataluña y País Vasco en las que menos del 40% de los alumnos asisten a colegios con mayoría de chicas. El caso opuesto lo presenta Asturias con un 58% de alumnos que asisten a colegios con mayoría de chicas.

En la base de datos original existe una variable que viene a diferenciar **tipos de colegio según la titularidad**: colegios públicos y colegios privados. A su vez, los colegios privados se clasifican en: privados puros, aquellos que tienen un porcentaje de financiación pública inferior al 50%; y colegios concertados, cuando dicho porcentaje es superior o igual al 50%. La evidencia empírica acerca de la influencia de la propiedad de los colegios en el resultado educativo es algo controvertida. Con los datos de PISA se encuentran los trabajos de Fertig (2003) y Aburra (2005), relativos a Alemania e Italia respectivamente, que no encuentran efecto significativo de la titularidad del centro. Por último, utilizando los datos de la edición de PISA 2003 en España, Calero y Escardíbul (2007) obtienen que las variables de propiedad de los colegios no son significativas una vez que se controlan por otras variables personales, familiares y características relativas al centro escolar. Las regiones presentan grandes diferencias en cuanto al porcentaje de alumnos que asisten a distintos tipo de colegio según la propiedad. En la Tabla 5 se observa que la región con mayor porcentaje de estudiantes en colegios públicos es Andalucía con un 75% mientras que el caso contrario lo presentan el País Vasco y Cataluña con un 42% y un 55% respectivamente. Sin embargo, estas dos últimas regiones se diferencian bastante, ya que mientras que País Vasco presenta un 58% de estudiantes de colegios concertados, dicho porcentaje para Cataluña es de tan sólo un 21%, siendo esta región la de mayor porcentaje de alumnos en colegios privados.

El **porcentaje de repetidores** que tuvo el colegio en el último año académico en ESO se ha tenido en cuenta como una variable de colegio. Los colegios se dividen entre aquéllos cuyo porcentaje de repetidores es mayor o menor del 15%, cifra que divide aproximadamente en dos partes de igual número a la muestra de alumnos.

Para el análisis se ha incluido la **presión que ejercen los padres** sobre el colegio. Se han diferenciado aquéllos colegios para los que existe una mayoría de padres exigentes que ejercen presión sobre la dirección del colegio y aquéllos otros para los que solo hay una minoría de padres que la ejercen o ésta es inexistente. Según la Tabla 5, la región en la que los padres son más exigentes es Cataluña. En esta región un 17% de los alumnos asisten a colegios donde la mayoría de los padres presionan para elevar el nivel educativo. Este dato contrasta con la región de Cantabria y Castilla y León donde dicho porcentaje es del 2% y 0% respectivamente.

Para cada colegio se ha computado el porcentaje de alumnos cuyos progenitores tienen al menos titulación secundaria. Así, los colegios se han categorizado en dos grupos delimitados por la mediana de dicha variable porcentual. De esta manera se crea la variable de **nivel educativo de los padres de los compañeros**. Esta variable viene a relacionarse con lo que ha



venido denominándose en la literatura con el término anglosajón “peer group” de tal manera que refleja las características de los compañeros de clase. Desde el informe Coleman (1966) han sido varios los autores que han incluido el grupo de compañeros de clase como factor explicativo del logro educativo. Así, se pueden citar Summers y Wolfe (1977) y Henderson, Mieszkowski y Sauvageau (1978) que encuentran un efecto positivo de las características de los compañeros de clase, al igual que Feinstein y Symons (1999).

Otra variable tenida en cuenta es el **tamaño de la clase de la lengua** en que se hizo la evaluación de los alumnos. Los alumnos se han clasificado en dos tipos de colegios cuyo tamaño de la clase es de más de 25 alumnos y aquéllos otros para los que el tamaño es de 25 o menos. A pesar de que hay amplia literatura que utilizando técnicas econométricas no encuentran evidencia de la influencia de esta variable en el resultado educativo, véase Hanushek (1998), esta variable se utiliza en este trabajo para identificar la probabilidad de repetir, siendo fundamental para la estimación del modelo de Regresiones Alternantes. Según lo observado en la Tabla 5 Cataluña y Andalucía son las regiones que presentan un mayor número de alumnos por clase, el 65% y el 63% de los alumnos de estas dos regiones, respectivamente, asisten a colegios con 26 estudiantes o más en las clases de lengua. El caso contrario lo presenta Asturias, tan sólo un 18% de sus estudiantes se encuentran en dicha categoría.

### 3 METODOLOGÍA

En el presente trabajo se pretende obtener las diferencias regionales resultantes una vez que se controla por el resto de las variables explicativas observables. Para ello se realiza la estimación de un Modelo de Regresiones Alternantes (MRA) que permite la estimación de coeficientes insesgados al controlar por la selección endógena de los alumnos repetidores y no repetidores. Por otro lado, se estima un modelo MCO que sirve para comparar con los resultados del MRA.

La estimación de un MRA se debe al problema de selección endógena de los alumnos repetidores y no repetidores. Los alumnos incluidos en PISA tienen la misma edad, 15 años, por lo que deberían estar en el mismo curso, concretamente 4º de ESO. Sin embargo, hay alumnos que han indicado estar en un curso inferior lo que implica que han repetido. La variable que indica si el alumno repite es claramente endógena debido a que el repetir influye en el resultado educativo, pero a su vez, la repetición se ve influenciada por el resultado educativo previo del alumno. Dado que los alumnos repetidores han completado un contenido curricular menos avanzado, se espera que obtengan menos nota en PISA. Pero las diferencias entre los alumnos repetidores y no repetidores no se van a deber únicamente al hecho de repetir. Los alumnos que han repetido presentan características diferentes que influyen en el resultado educativo. Más

concretamente la hipótesis de partida puede ser la creencia de que los alumnos que repiten son aquéllos que presentan peores características para el resultado educativo. En la medida en que estas características sean inobservables, las diferencias en resultado educativo entre Repetidores y No Repetidores bajo MCO tendrán un sesgo.

Para solventar este problema se propone la estimación por Máxima Verosimilitud de un MRA (*Switching Regression Model* en inglés) descrita por Lokshin y Sajaia (2004). Igualmente la idea del modelo es análoga a la establecida por Maddala (1983) para el caso general.

Se supone que un alumno repite dependiendo del resultado que obtenga en una determinada variable latente  $I_i^*$  que depende a su vez de una serie de características  $Z_i$  del alumno según la siguiente ecuación de selección (1).

$$I_i^* = Z_i \gamma_1 + e_i \quad (1)$$

La realización dicotómica de la variable latente es como sigue:

$$I_i = \begin{cases} 1 & \text{si } I_i^* > 0 \text{ (el alumno repite)} \\ 0 & \text{si } I_i^* \leq 0 \text{ (el alumno no repite)} \end{cases} \quad (2)$$

Así mismo, el modelo supone que la nota de un alumno  $y_i$  depende de unas variables incluidas en el vector  $X_i$ , donde por simplificación se recogen todas las variables tanto individuales, dotaciones iniciales, y de colegio. Sin embargo, hay una ecuación para los repetidores ( $y_{1i}$ ) y otra para los no repetidores ( $y_{2i}$ ):

$$y_{1i} = X_i \beta_1 + u_{1i} \quad (3)$$

$$y_{2i} = X_i \beta_2 + u_{2i} \quad (4)$$

Se asume que  $e_i$ ,  $u_{1i}$  y  $u_{2i}$  tienen una distribución normal trivariante con media cero y la siguiente matriz de covarianzas:

$$\Omega = \begin{bmatrix} \sigma_1^2 & & \\ \sigma_{12} & \sigma_2^2 & \\ \sigma_{1e} & \sigma_{2e} & \sigma_e^2 \end{bmatrix} \quad (5)$$

Donde  $\sigma_e^2$  es la varianza del término de error en la ecuación de selección (1), y  $\sigma_1^2$  y  $\sigma_2^2$  son las varianzas de los términos de error en las ecuaciones de rendimiento educativo (3) y (4). En cuanto a  $\sigma_{1e}$  es la covarianza de  $u_{1i}$  y  $e_i$  (muestra correlación entre la probabilidad de repetir y el rendimiento educativo de los repetidores), y  $\sigma_{2e}$  es la covarianza entre  $u_{2i}$  y  $e_i$  (muestra correlación entre la probabilidad de repetir y el rendimiento educativo de los no repetidores).

Dependiendo del signo de  $\sigma_{1e}$  las interpretaciones serán diferentes. Si  $\sigma_{1e} < 0$  existe una relación negativa entre las variables inobservables que hacen que un alumno repita (no repita) y las que hacen que un alumno repetidor obtenga mejores (peores) notas. En otras palabras esto significa que lo que hace que un alumno tenga más probabilidad de repetir hace que el mismo obtenga un bajo resultado académico. Por el contrario, si ocurre  $\sigma_{1e} > 0$  la interpretación será que lo que hace que un alumno repetidor tenga más probabilidad de repetir hace que un alumno repetidor obtenga un resultado educativo mayor. Por último, podría ocurrir que  $\sigma_{1e} = 0$  lo que implicaría que no hay correlación entre los errores de la ecuación de selección y los errores de la ecuación de rendimiento educativo de los repetidores.

Por otro lado, si  $\sigma_{2e} > 0$  esto indicaría que lo que hace disminuir la probabilidad de repetir también disminuye el resultado educativo de los no repetidores. Por el contrario, si  $\sigma_{2e} < 0$  las variables inobservables que hacen disminuir la probabilidad de repetir hacen aumentar el resultado educativo de los no repetidores. También podría pasar que  $\sigma_{2e} = 0$  en cuyo caso no hay correlación entre los errores de la ecuación de selección y el término de error de la ecuación de rendimiento educativo de los repetidores.

Este modelo, al ser estimado por máxima verosimilitud, mejora respecto a otros modelos anteriormente usados en la literatura PISA porque controla la selección endógena de los *repetidores* y *no repetidores* y además permite que el efecto de las variables explicativas sea distinto para ambos grupos de alumnos.

## 4 RESULTADOS Y CONCLUSIONES

En los modelos estimados aparecen todas las regiones con muestra ampliada excepto Andalucía que está en la constante, es decir, todas las comparaciones se harán con el estudiante andaluz. Además, para las regiones de Cataluña, País Vasco y Navarra, los estudiantes tienen patrones de resultado diferentes según la lengua que el alumno habla en casa. Por ello se ha procedido a distinguir entre aquéllos que hablan la lengua regional, los que hablan castellano, y los que hablan otra distinta a las anteriores. Según un análisis bivalente parece haber diferencias entre los alumnos gallegos según la lengua que hablan en casa, sin embargo éstas desaparecen una vez que se controla por el resto de variables explicativas<sup>5</sup> por lo que para esta comunidad autónoma no se ha diferenciado por la lengua que se habla en casa.

---

<sup>5</sup> La diferencia entre los alumnos gallegos que hablan la lengua regional y aquéllos que hablan castellano es de un punto y no es significativa.

Otras variables que han tenido un tratamiento especial son: la variable de educación de la madre ha sido interaccionada con la educación del padre; y la variable de titularidad del colegio se ha interaccionado con la variable que indica presión de los padres de los alumnos.

En la Tabla 6 se muestran los resultados del MRA, que incluye la estimación de la ecuación de selección y la estimación del resultado educativo para los mismos grupos. En la misma tabla se pueden observar los resultados por MCO.

Al final de la Tabla 6 se encuentran las estimaciones de los coeficientes de correlación del término de error de la ecuación de selección con el término de error de la ecuación de rendimiento de los *repetidores* ( $\rho_1$ ) y los *no repetidores* ( $\rho_2$ ). Concretamente se ha estimado un coeficiente de correlación  $\rho_2$  de -0,25 significativo al 1%. El signo negativo de dicho coeficiente indica que aquéllos factores inobservables que hacen disminuir la probabilidad de repetir también hacen incrementar el resultado educativo de los *no repetidores*. En cuanto al coeficiente de correlación estimado  $\rho_1$  no es significativo, por lo que no se ha encontrado relación entre el término de error de la ecuación de selección y el término de error de la ecuación de resultado de los alumnos *repetidores*.

El hecho de que  $\rho_2$  sea significativo implica que al estimar las diferencias entre los alumnos *repetidores* y *no repetidores* por MCO el resultado presenta un “sesgo de selección”. Esto lo podemos observar viendo las diferencias entre alumnos repetidores y no repetidores estimadas por el MRA y por MCO. Por ejemplo, los alumnos *no repetidores* que presentan las características de la constante del modelo<sup>6</sup> en la ecuación de resultados obtienen un resultado de 448,25 según la estimación por el MRA y un resultado de 463,81 según la estimación por MCO. Los mismos alumnos de referencia *repetidores* obtienen un resultado de 413,99 según la estimación por el MRA y un resultado de 407,27 según la estimación por MCO. De manera que si se controla por la selección endógena las diferencias entre *repetidores* y *no repetidores* se reducen con respecto a la estimación por MCO.

En la Tabla 6 se puede ver también si el sesgo se produce en la estimación de los efectos que tienen las distintas variables explicativas para cada uno de los grupos, *Repetidores* y *No Repetidores*.

En cuanto a los *no repetidores* en general no parece haber diferencias considerables entre los coeficientes de las variables de región estimados por las dos técnicas tanto en importancia como en significatividad. Para las variables de educación y ocupación de los padres, cultura educativa del hogar, horas de estudio de ciencias y actitud de los alumnos el

---

<sup>6</sup> En este modelo se está considerando como individuo de referencia a un alumno con residencia en Andalucía, nativo, de sexo masculino, con padre y madre de nivel educativo y ocupación baja, cuya madre trabaja y que vive en un hogar con una dotación cultural baja. Es un alumno que asiste a menos de dos horas de ciencias en el colegio a la semana, tiene una actitud ante la ciencia baja y que estudia en un colegio público con un nivel educativo de los padres bajo, con mayoría de chicos en la clase y que nació en el mes de enero.

coeficiente estimado por Regresiones Alternantes es algo mayor que por MCO a pesar de que la significatividad no varía apenas.

Por lo que respecta a los *repetidores* las estimaciones varían en alguna medida de una técnica a otra, sobre todo para las variables regionales. Bajo MCO los alumnos navarros que hablan eusquera y los vascos que hablan castellano no son significativamente distintos a Andalucía, sin embargo se estima un menor resultado para estas regiones bajo MRA. En general, los coeficientes de las variables regionales son menores cuando se estiman controlando por el sesgo de selección, lo que significa que las diferencias regionales se acortan aún más. Por último, la variable de mucha presión en colegios públicos pasa a ser significativa si se estima por MRA.

En la Tabla 6 se pueden comparar las estimaciones de los coeficientes de las variables explicativas para *repetidores* y *no repetidores* según el MRA. Las diferencias regionales estimadas para los alumnos *repetidores* son, para la mayoría de las regiones, menores que las estimadas para los alumnos *no repetidores*. A modo de ejemplo, para los *no repetidores* se estiman unas diferencias con Cantabria, Galicia y La Rioja de 15,91, 18,87 y 21,61 respectivamente, las mismas estimaciones para los alumnos *repetidores* son de 10,38, 10,89 y 15,29 respectivamente. A pesar de ello, tanto para un grupo como para otro se estima que las regiones de Aragón, Asturias, Cantabria, Castilla y León, Galicia y La Rioja obtienen mejor nota que la región andaluza. También para ambos grupos, los alumnos catalanes que hablan catalán y el resto de España no son diferentes a los alumnos andaluces. Las regiones que para el grupo de *no repetidores* obtienen peor resultado a Andalucía son Cataluña si hablan castellano y Navarra y País Vasco si hablan eusquera. Mientras que para los *repetidores* las regiones de peor rendimiento educativo son Navarra si hablan eusquera y País Vasco si hablan castellano. En cuanto a las variables de dotaciones iniciales se observa un mayor efecto de la educación de los padres para los alumnos *repetidores* y, sin embargo, un menor efecto positivo de las variables de ocupación de los padres. Llama la atención el hecho de que el tener madre ama de casa respecto a tener madre de ocupación baja supone 5,65 puntos menos para los *no repetidores* y 16,01 puntos menos para los *repetidores*, un efecto negativo mucho mayor. Otras diferencias que se encuentran son: un mayor efecto negativo de ser mujer para los *repetidores* y un mayor efecto negativo de ser inmigrante para los *no repetidores*. En cuanto a las variables individuales la actitud influye para ambos grupos de alumnos sin embargo su efecto es mucho mayor en los *no repetidores*. Por último, se encuentran dos variables de colegio que tienen un efecto distinto en uno y otro grupo de alumnos. Parece que la presión de los padres en los colegios públicos tiene influencia para los alumnos *repetidores* y no para los *no repetidores*. Igualmente, la presión ejercida por los padres en los colegios privados sólo afecta a los alumnos *no repetidores*.

La especificación del MRA permite la identificación del modelo al introducir variables en la ecuación de selección que son significativas y no aparecen en las ecuaciones de resultado

educativo. La existencia de estas variables permite distinguir el efecto que tiene en las notas el hecho de repetir del efecto que tienen las variables inobservables que se relacionan con la probabilidad de repetir. En concreto, las variables utilizadas para la identificación del modelo son: el tiempo que dedica el alumno al uso de ordenadores con fines distintos al estudio, como jugar, y bajarse música de Internet; y determinadas variables de colegio, como asistir a un colegio concertado; la presión que ejercen los padres en el colegio concertado; asistir a una clase de lengua de más de 25 alumnos; y asistir a un colegio en el que hay un alto porcentaje de repetidores. Estas variables son significativas a la hora de explicar la probabilidad de repetir según la ecuación de selección del MRA.

El análisis realizado ha permitido comparar el resultado educativo de estudiantes de diversas comunidades autónomas una vez ajustamos por diversas variables que pueden estar influyendo en dicho resultado. El hecho de que estudiantes de Aragón, Asturias, Cantabria, Castilla y León, Galicia y La Rioja con iguales características obtengan mayor rendimiento educativo que los estudiantes andaluces, indica que hay aún factores relacionados con la región que explican los resultados de PISA y que sin embargo no han podido ser captados con el modelo especificado. Entre estas variables pueden estar factores relacionados con el sistema educativo regional, o incluso variables de los estudiantes o sus colegios que no son observables en este trabajo. Las variables del sistema educativo que pudieran estar influyendo podemos considerar las siguientes: formas organizativas o metodológicas de los colegios; el nivel educativo o titulación del profesorado; la estabilidad de la plantilla de profesores; o el contenido curricular de las asignaturas impartidas. Igualmente si las diferencias estimadas entre la región de referencia y las demás regiones no son significativas (los alumnos catalanes que hablan catalán y el Resto de España) o incluso cambian de signo (Cataluña si hablan castellano y Navarra si hablan eusquera y País Vasco) se debe interpretar que las diferencias observadas son explicadas totalmente por las variables del modelo.

Por último, la estimación del MRA ha permitido controlar por el sesgo de selección existente entre *repetidores* y *no repetidores*. De hecho los resultados indican que la endogeneidad se debe a que las variables inobservables que afectan a la probabilidad de repetir están correlacionadas con el rendimiento de los alumnos no repetidores. Al controlar por este sesgo las diferencias entre *repetidores* y *no repetidores* se reducen considerablemente respecto a la estimación por MCO. Este modelo también ha permitido estimar el efecto de las covariables, por un lado, en el resultado educativo de los *repetidores* y, por otro, en el resultado de los *no repetidores*. Esta diferenciación parece tener sentido ya que los coeficientes varían para ambos grupos de alumnos.

## Referencias

- (1) Calero, J. y Escardíbul, J-O. (2007), “Evaluación de servicios educativos: el rendimiento en los centros públicos y privados medido en PISA-2003”, Hacienda Pública Española/Revista de Economía Pública, 183: pp. 33-66.
- (2) Carabaña, J. (2008), “Las diferencias entre países y regiones en las pruebas PISA”.Mimeo.www.colegiodeemeritos.es/docs/repositorio/es\_ES/documentos/pisa\_carabana\_(vf).pdf
- (3) Ciccone, A. y Garcia-Fontes, W. (2008) , “*The quality of the Catalan and Spanish Education Systems: A perspective from PISA*”. Mimeo. [www.antonioiciccone.eu/wp-content/uploads/2008/11/cc-quality-of-education-system-9-with-tables.pdf](http://www.antonioiciccone.eu/wp-content/uploads/2008/11/cc-quality-of-education-system-9-with-tables.pdf)
- (4) Coleman, J. S., J. S. Campbell, , E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld y R. L. York (1966), “*Equality of Educational Opportunity.*” Washington, D.C.: *US Department of Health, Education, and Welfare, US Government Printing Office.*
- (5) Cordero, J.M., Crespo, E., Pedraja, F., y Santín, D. (2009), “Análisis de las diferencias entre Comunidades Autónomas españolas en PISA 2006”. XVI ENCUENTRO DE ECONOMÍA PÚBLICA. Granada, 5-6 Febrero de 2009.
- (6) De Fraja, Gianni & Oliveira, Tania & Zanchi, Luisa, 2005. "*Must Try Harder. Evaluating the Role of Effort in Educational Attainment,*" *CEPR Discussion Papers 5048, C.E.P.R. Discussion Papers.*
- (7) Feinstein, L. y J. Symons (1999), “*Attainment in secondary education.*” *Oxford Economic Papers*, 51: 300-321.
- (8) Fertig, M. (2003), “*Who’s to Blame? The Determinants of German Students’ Achievement in the PISA 2000 Study.*” *IZA Discussion Paper Series*, 739.
- (9) Hanushek, E. A. (1998), “*The evidence on Class Size*”, W. Allen Wallis *Institute of Political Economy, Occasional Paper 98-1.*
- (10) Henderson, V., P. Mieszkowski y Y. Sauvageau (1978), “*Peer group effects in educational production functions.*” *Journal of Public Economics*, 10 (1): 97-106.
- (11) Hoxby, C. (2000). *Peer Effects in the Classroom: Learning from Gender and Race Variation.*, NBER working paper 7867.
- (12) Lokshin, M. & Sajaia, Z. (2004). "Maximum likelihood estimation of endogenous switching regression models," *Stata Journal*, StataCorp LP, vol. 4(3), pages 282-289, September.

- (13) Maddala, G., (1983) *Limited-Dependent and Qualitative Variables in Econometric*, Econometric Society Monographs No. 3, Cambridge University Press, New York.
- (14) OCDE. (2005) "*PISA 2003 Technical Report.*"
- (15) OCDE. (2006) "*Assessing Scientific, Reading and Mathematical Literacy: A framework for PISA 2006*"
- (16) OCDE. (2008) "Informe PISA 2006. Competencias científicas para el mundo del mañana"
- (17) Summers, A. A. y B. L. Wolfe (1977), "*Do schools make a difference?*" *American Economic Review*, 67 (4): 639-652.
- (18) Zinovyeva, N., Felgueroso, F. y Vázquez, P. (2008). "*Immigration and Students' Achievement in Spain,*" [\*Working Papers\*](#) 2008-37, FEDEA.



## APÉNDICE

**Tabla 1 Muestras regionales y resultado en ciencias**

Región	Muestra	Resultado en ciencias
<b>Andalucía</b>	1.463	520
<b>Aragón</b>	1.526	520
<b>Asturias</b>	1.579	513
<b>Cantabria</b>	1.496	511
<b>Castilla y León</b>	1.512	509
<b>Cataluña</b>	1.527	508
<b>Galicia</b>	1.573	505
<b>La Rioja</b>	1.333	495
<b>Navarra</b>	1.590	491
<b>País Vasco</b>	3.929	484
<b>Resto de España</b>	2.076	474
<b>Total España</b>	<b>19.604</b>	<b>488</b>

**Tabla 2 Nivel educativo de los padres por regiones**

Región	Sin educación		Primaria		Secundaria o FP		Enseñanza superior	
	Padres	Madres	Padres	Madres	Padres	Madres	Padres	Madres
<b>Andalucía</b>	16%	15%	23%	24%	48%	49%	13%	12%
<b>Aragon</b>	5%	3%	11%	11%	63%	64%	21%	23%
<b>Asturias</b>	5%	3%	10%	8%	64%	66%	22%	23%
<b>Cantabria</b>	3%	3%	9%	9%	69%	70%	19%	18%
<b>Castilla y León</b>	5%	4%	13%	11%	63%	63%	19%	22%
<b>Cataluña</b>	7%	7%	15%	13%	56%	58%	22%	22%
<b>Galicia</b>	7%	5%	15%	15%	61%	62%	18%	18%
<b>La Rioja</b>	4%	3%	13%	10%	62%	66%	22%	21%
<b>Navarra</b>	6%	4%	12%	10%	59%	62%	23%	23%
<b>País Vasco</b>	5%	3%	9%	10%	60%	61%	26%	26%
<b>Resto de España</b>	10%	7%	15%	16%	57%	58%	18%	19%

**Tabla 3 Estatus Ocupacional de los padres, cultura educativa y porcentaje de inmigrantes por regiones**

Región	Ocupación de los padres					Cultura educativa			% de inmigrantes
	Ama de casa (Madres)	Baja ocupación		Alta ocupación		Baja	Media	Alta	
		Padres	Madres	Padres	Madres				
<b>Andalucía</b>	37%	70%	43%	30%	20%	34%	36%	30%	2%
<b>Aragon</b>	21%	60%	44%	40%	35%	20%	30%	50%	5%
<b>Asturias</b>	30%	63%	39%	37%	31%	24%	34%	43%	3%
<b>Cantabria</b>	27%	63%	44%	37%	29%	22%	36%	42%	4%
<b>Castilla y León</b>	29%	67%	42%	33%	30%	20%	34%	46%	3%
<b>Cataluña</b>	13%	58%	49%	43%	39%	29%	32%	39%	10%
<b>Galicia</b>	25%	66%	45%	34%	30%	31%	34%	35%	3%
<b>La Rioja</b>	23%	62%	45%	38%	31%	19%	32%	48%	6%
<b>Navarra</b>	19%	65%	46%	36%	35%	21%	33%	46%	7%
<b>País Vasco</b>	20%	56%	43%	44%	37%	21%	35%	44%	4%
<b>Resto de España</b>	27%	62%	41%	38%	32%	27%	34%	39%	10%

**Tabla 4 Horas semanales de clases de ciencias en el colegio por regiones**

Región	Horas semanales de ciencias			Actitud ante la ciencia			% de repetidores
	Menos de 2 horas	Entre 2 y 4 horas	Más de 4 horas	Actitud Alta	Actitud Media	Actitud Baja	
Andalucía	30%	44%	26%	24%	48%	28%	48%
Aragón	26%	44%	30%	27%	45%	29%	38%
Asturias	30%	42%	29%	29%	46%	26%	34%
Cantabria	25%	44%	30%	27%	47%	26%	39%
Castilla y León	24%	44%	31%	30%	48%	22%	38%
Cataluña	15%	65%	19%	25%	47%	28%	30%
Galicia	30%	45%	25%	26%	48%	25%	39%
La Rioja	23%	48%	29%	29%	47%	24%	38%
Navarra	31%	26%	44%	28%	40%	33%	30%
País Vasco	26%	48%	25%	27%	42%	31%	24%
Resto de España	31%	40%	29%	26%	47%	27%	43%

**Tabla 5 Distribución regional de alumnos según el porcentaje de chicas en el colegio, la titularidad del centro, tamaño de la clase de lengua, y la presión de los padres**

Región	Titularidad del centro educativo				Clases de 26 o más	Muchos padres ejercen presión
	Igual o más del 50% de chicas	Privado	Concertado	Público		
Andalucía	48%	1%	24%	75%	63%	5%
Aragón	55%	2%	34%	64%	38%	6%
Asturias	58%	13%	23%	64%	18%	6%
Cantabria	57%	2%	32%	66%	24%	2%
Castilla y León	54%	6%	31%	64%	37%	0%
Cataluña	38%	23%	21%	55%	65%	17%
Galicia	45%	11%	18%	71%	34%	3%
La Rioja	51%	5%	33%	61%	44%	4%
Navarra	53%	3%	38%	59%	50%	12%
País Vasco	40%	0%	58%	42%	33%	12%
Resto de España	55%	12%	22%	66%	57%	10%

**Tabla 6 Estimación de la ecuación de selección y ecuación del resultado educativo en ciencias. Modelo de Regresiones Alternantes versus MCO**

VARIABLE	MRA (Ecuación de selección) $I_i^* = Z_i\gamma_i + e_i$		Rendimiento No Repetidores			Rendimiento Repetidores				
	COEF.	z	COEF.	z	COEF.	t	COEF.	z	COEF.	t
<b>REGIONES</b>										
<i>Aragón</i>	-0,05	-0,75	11,73***	3,52	12,61***	3,90	9,93**	2,02	12,88***	2,70
<i>Asturias</i>	-0,13**	-2,06	10,58***	3,26	10,12***	3,22	9,14*	1,81	12,16**	2,51
<i>Cantabria</i>	-0,03	-0,53	15,91***	4,74	16,88***	5,16	10,38**	2,15	12,07***	2,60
<i>Castilla y León</i>	-0,02	-0,40	17,65***	5,34	18,58***	5,87	20,97***	4,37	25,10***	5,49
<i>Cataluña hablan Catalán</i>	-0,40***	-4,23	5,15	1,22	0,86	0,21	-8,02	-0,83	-3,81	-0,42
<i>Cataluña hablan Castellano</i>	-0,25***	-3,51	-7,04*	-1,85	-10,09***	-2,74	-9,67	-1,55	-6,85	-1,18
<i>Cataluña hablan otra lengua</i>	0,08	0,40	-24,65*	-1,66	-18,59	-1,33	-27,73**	-2,10	-19,81	-1,51
<i>Galicia</i>	-0,03	-0,56	18,87***	5,82	19,19***	6,11	10,89**	2,37	14,80***	3,36
<i>La Rioja</i>	-0,08	-1,30	21,61***	6,45	21,42***	6,56	15,29***	2,95	19,96***	4,06
<i>Navarra que hablan Eusquera</i>	-0,49**	-2,49	-20,29**	-2,33	-23,08***	-2,72	-34,00*	-1,69	-29,24	-1,46
<i>Navarra que hablan castellano</i>	-0,22***	-3,38	15,39***	4,62	13,49***	4,26	-2,00	-0,35	4,49	0,86
<i>Navarra que hablan otra lengua</i>	0,00	0,00	48,23***	3,02	42,59***	2,74	-8,42	-0,49	-11,72	-0,74
<i>País Vasco que hablan Eusquera</i>	-0,64***	-6,77	-13,23***	-3,33	-13,37***	-3,69	-16,91	-1,57	-22,13**	-2,47
<i>País Vasco que hablan castellano</i>	-0,34***	-5,97	-1,54	-0,52	-3,57	-1,28	-10,12*	-1,90	-6,27	-1,41
<i>País Vasco que hablan otra</i>	0,31	1,36	5,74	0,36	-1,17	-0,08	-31,03**	-2,14	-20,72	-1,51
<i>Resto</i>	-0,07	-1,17	-3,91	-1,25	-6,66***	-2,22	-4,84	-1,06	-2,85	-0,66
<b>DOTACIONES INICIALES</b>										
<i>Madre educación alta y padre</i>	-0,29***	-5,32	9,79***	2,87	7,65**	2,37	14,34***	3,34	14,64***	3,78
<i>Madre educación baja y padre</i>	-0,23***	-4,04	7,29**	2,01	5,36	1,56	5,43	1,23	5,97	1,47
<i>Madre y padre educación alta</i>	-0,37***	-9,02	11,17***	3,94	8,07***	3,11	14,48***	4,01	15,65***	5,52
<i>Padre de ocupación alta</i>	-0,26***	-8,93	12,23***	8,33	10,83***	8,02	5,93*	1,89	5,72**	2,23
<i>Madre de ocupación alta</i>	-0,34***	-10,58	11,52***	7,06	9,08***	6,12	3,02	0,81	4,78*	1,67
<i>Madre ama de casa</i>	-0,15***	-4,90	-5,65***	-3,37	-6,30***	-3,97	-16,01***	-6,20	-14,96***	-6,37
<i>Cultura educativa media</i>	-0,25***	-8,88	11,19***	6,87	10,39***	6,94	13,41***	4,93	13,61***	6,12
<i>Cultura educativa alta</i>	-0,24***	-6,72	15,26***	8,24	14,29***	8,25	17,49***	5,03	17,86***	5,83
<i>País de nacimiento (inmigrante)</i>	0,70***	11,49	-28,85***	-6,43	-24,18***	-6,00	-22,88***	-4,14	-24,07***	-6,20
<b>VARIABLES INDIVIDUALES</b>										
<i>Sexo del alumno (mujer)</i>	-0,30***	-11,77	-14,50***	-10,64	-16,73***	-13,92	-20,42***	-7,11	-18,73***	-9,24
<i>Mes de nacimiento del alumno</i>	0,03***	7,52	-1,04***	-5,64	-0,91***	-5,28	-1,31***	-3,89	-1,21***	-4,16
<i>De 2 a 4 de clases de ciencias</i>	-0,13***	-4,45	21,05***	11,74	19,86***	11,83	18,88***	7,70	20,29***	9,16
<i>Más de 4 hrs. de clases de ciencias</i>	-0,42***	-11,76	44,08***	21,66	41,89***	23,08	31,48***	7,86	31,12***	10,57
<i>Actitud media</i>	-0,11***	-3,85	21,21***	12,50	19,41***	12,02	8,68***	3,61	10,42***	4,72
<i>Actitud alta</i>	-0,60***	-16,13	53,51***	26,15	49,31***	27,58	31,73***	6,54	33,16***	10,34
<i>Casi a diario juega con el PC</i>	0,09**	2,36								
<i>Casi a diario baja música</i>	-0,02	-0,57								
<i>Juega y baja música casi a diario</i>	0,17***	3,35								
<b>VARIABLES DE COLEGIO</b>										
<i>Mayoría de chicas en el colegio</i>	0,04	1,50	4,92***	3,91	4,13***	3,44	6,93***	3,26	6,23***	3,10
<i>Colegio privado</i>	0,05	0,73	-8,78***	-3,05	-7,99***	-3,04	7,04	1,22	7,01	1,29
<i>Mucha presión en públicos</i>	0,00	-0,04	-4,44	-0,81	-0,56	-0,11	21,07**	2,18	15,60	1,64
<i>Mucha presión en privados</i>	-0,25**	-2,01	23,09***	4,84	22,10***	4,83	4,55	0,34	5,56	0,43
<i>Nivel educativo de los padres alto</i>	0,03	0,89	6,24***	4,26	6,53***	4,66	7,87***	3,18	8,33***	3,56
<i>Colegio concertado</i>	-0,12***	-3,65								
<i>Mucha presión en concertados</i>	-0,21***	-2,76								
<i>Clase de muchos alumnos</i>	-0,07**	-2,57								
<i>Alto porcentaje de repetidores</i>	0,28***	9,32								
<i>Constante</i>	0,48***	6,52	448,25***	79,26	463,817**	124,92	413,99***	71,25	407,27***	87,06

*Estimación de las varianzas y correlaciones entre los términos de error del modelo de selección endógena:*

$\rho_1 = \frac{\sigma_{1e}^2}{\sigma_1 \sigma_e}$ (Error Estándar)	<b>0,02 (0,13)</b>	$\sigma_1$ (Error Estándar)	<b>62,83*** (0,58)</b>
$\rho_2 = \frac{\sigma_{2e}^2}{\sigma_2 \sigma_e}$ (Error Estándar)	<b>-0,25*** (0,06)</b>	$\sigma_2$ (Error Estándar)	<b>64,28*** (0,73)</b>

**Notas:** \*, \*\* y \*\*\* significa que el coeficiente es significativo al 10%, al 5% o al 1%, respectivamente. El número de observaciones es 14.428.