



Working papers series

WP ECON 10.12

*DIFERENCIAS REGIONALES EN RENDIMIENTO
EDUCATIVO EN ESPAÑA:
¿LA FAMILIA LO EXPLICA TODO?*

J. Ignacio García Pérez (U. Pablo de Olavide,
FEDEA & FCEA)

Marisa Hidalgo Hidalgo (U. Pablo de Olavide)

J. Antonio Robles Zurita (U. Pablo de Olavide)

JEL Classification numbers: H52, I21, I28, D61, R50

Keywords: PISA, Rendimiento educativo, Análisis Regional,
Descomposición de las diferencias, Sistema educativo, Políticas
Públicas



Department of Economics

**DIFERENCIAS REGIONALES EN RENDIMIENTO EDUCATIVO EN ESPAÑA:
¿LA FAMILIA LO EXPLICA TODO?***

J. Ignacio García Pérez

Universidad Pablo de Olavide, FCEA y FEDEA

Marisa Hidalgo Hidalgo

Universidad Pablo de Olavide

J. Antonio Robles Zurita

Universidad Pablo de Olavide

Noviembre, 2010

* Los autores desean agradecer la financiación recibida del Ministerio de Educación (SEJ2006-04803 y SEJ 2007-67734), Junta de Andalucía (SEJ-2905 y SEJ-426) y Centro de Estudios Andaluces. Asimismo, agradecemos los comentarios y sugerencias recibidos de José Luis Pinto y el resto de compañeros del Area de Análisis Económico de la Universidad Pablo de Olavide.

Resumen

Los resultados del Programa PISA 2006 evidenciaron enormes diferencias en las puntuaciones obtenidas por los alumnos de distintas regiones españolas. El objetivo de este artículo es analizar los factores explicativos de estas diferencias. Nuestros resultados indican que las características socio-familiares de los estudiantes procedentes de distintas regiones son un factor crucial. Sin embargo, también encontramos que las diferencias en los sistemas educativos son importantes. Las variables relacionadas con particularidades del sistema educativo regional son importantes para explicar los distintos resultados académicos en distintas regiones. Las implicaciones en términos de políticas educativas de este último resultado son muchas, especialmente de cara a desarrollar medidas para la disminución de las desigualdades económicas regionales.

Abstract

PISA 2006 results showed huge differences in educational attainment across Spanish regions. The aim of this paper is to analyze the factors explaining these differences. Our results indicate that differences in family characteristics of students from different regions are crucial in explaining them. However, we also find that differences in regional educational systems are important factors explaining differences in education outcome. The variables related to regional particularities of the educational system are important in explaining the different results in different academic areas. The implications in terms of educational policies of the latter result are many, especially for the implementation of policies to reduce regional economic disparities.

Palabras clave: PISA, Rendimiento educativo, Análisis Regional, Descomposición de las diferencias, Sistema educativo, Políticas Públicas.

Clasificación JEL: H52, I21, I28, D61, R50

1 Introducción

Los resultados del programa PISA 2006 (Programme for International Student Assessment) han evidenciado las diferencias existentes entre las puntuaciones de alumnos de distintos países de la OCDE.¹ En efecto, la puntuación media de los alumnos españoles en PISA 2006 está por debajo de la media de los alumnos de la OECD en las tres áreas objeto de evaluación.² Dada la creciente evidencia empírica sobre la relación entre el nivel educativo y el crecimiento económico, los resultados para España están siendo objeto de interés por parte de políticos y académicos.³ Del mismo modo, resultan alarmantes las diferencias existentes entre las puntuaciones obtenidas por alumnos procedentes de distintas regiones españolas, en tanto que éstas pueden originar futuras desigualdades en renta y crecimiento económico interregional.

El análisis de estas diferencias regionales en resultado educativo tiene interés por la magnitud de las mismas, pero también porque la toma de decisiones en relación a la gestión y organización del sistema educativo en España está descentralizada en los distintos gobiernos autonómicos representativos de las regiones, de tal manera que a través de un análisis comparativo se pueden obtener implicaciones de política económica identificando cuáles son los sistemas educativos que mejor pueden estar funcionando. La diversidad regional en nivel educativo no sólo se encuentra entre los alumnos que son población objetivo de PISA sino

¹ El estudio de PISA 2006 se engloba dentro de un programa de evaluación de la educación más amplio que empezó en el año 2000. Cada tres años este programa mide el resultado de estudiantes de 15 años que se encuentran escolarizados, en países miembros de la OCDE y en otros países (llamados asociados). La evaluación se centra en el resultado de las áreas de ciencias, matemáticas, y lectura.

² La media de los alumnos españoles en PISA es de 488, 480 y 461 puntos en ciencias, matemáticas y lectura respectivamente. Mientras que la media de la OCDE es de 500 en todas las disciplinas.

³ Existe una literatura muy extensa sobre la relación entre educación y crecimiento económico. Ver, entre otros, Acemoglu (2009).

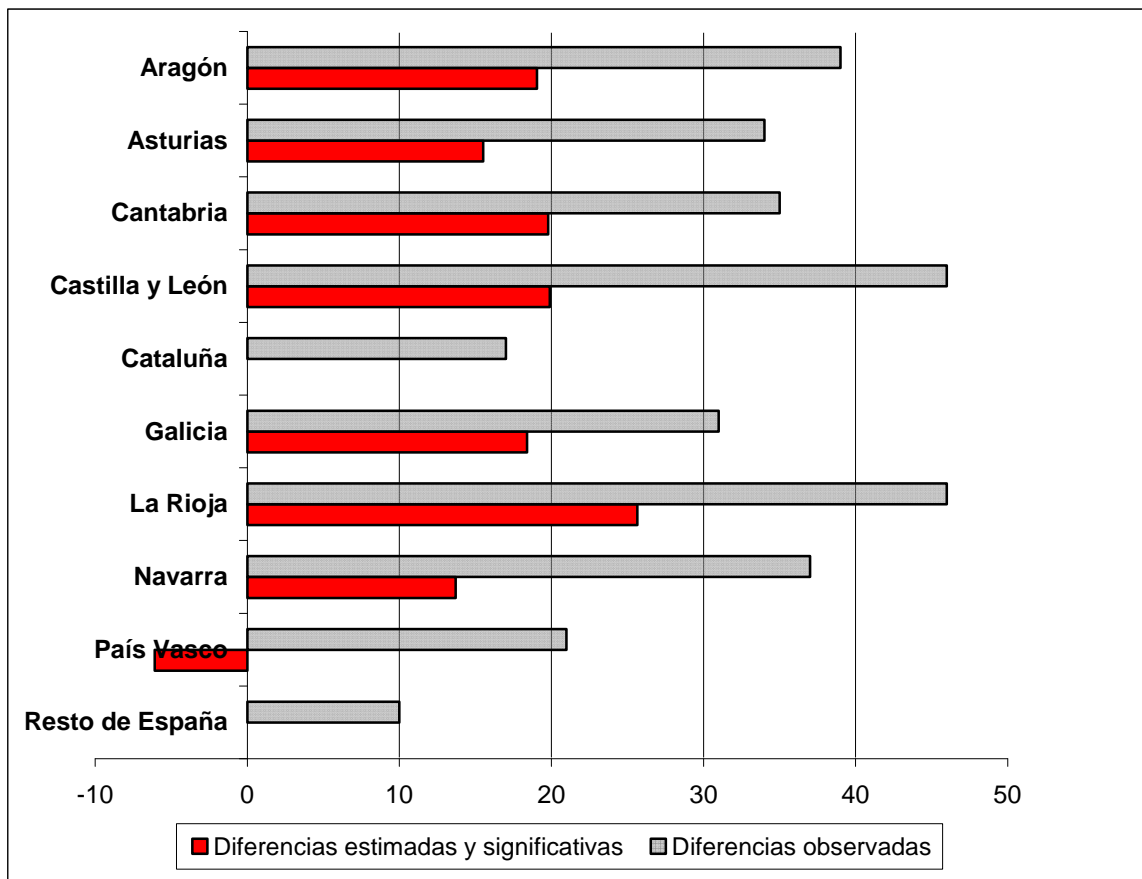
que tiene una presencia más amplia y explicitada en el conjunto de la población. Según la Encuesta de Población Activa en el segundo trimestre de 2006, año de la edición de PISA que analizamos en nuestro estudio, el 42% de la población de más de 16 años de Castilla la Mancha, el 39% en Castilla y León y el 35% en Andalucía tenían tan sólo educación primaria o inferior mientras que el mismo porcentaje era mucho menor para Madrid, 27%, Navarra, 28%, o el País Vasco, 28%. Igualmente, las administraciones autonómicas tienen capacidad para regular muchos aspectos del sistema educativo. En concreto, establecen dentro de unos límites fijados a nivel estatal el currículo correspondiente a la Educación Secundaria Obligatoria (ESO). Así, por ejemplo, los alumnos de 1º de ESO en Galicia y Asturias tienen 4 horas semanales de Ciencias de la Naturaleza mientras que en otras Comunidades como Andalucía, País Vasco o Aragón (entre otras), tienen sólo 3 horas a la semana. Los alumnos de 3º de ESO de Andalucía y Aragón tienen una hora más de ciencias a la semana (biología y geología, y física y química) que los alumnos que optan por el modelo lingüístico castellano en País Vasco y Navarra.

Como primera impresión de las diferencias en rendimiento educativo a nivel regional, y tomando a Andalucía como región de referencia, el cuadro 1 muestra las diferencias en las puntuaciones obtenidas en el examen de ciencias respecto al resto de comunidades autónomas españolas. Por una parte observamos las diferencias en las puntuaciones observadas de PISA 2006 entre las distintas regiones (barras azules) y por otra, las diferencias en las puntuaciones estimadas (barras rojas) mediante un modelo simple de regresión lineal donde incluimos variables *dummy* para las distintas regiones además de un conjunto de variables explicativas de las puntuaciones de PISA.⁴ Así podemos observar que, si bien las diferencias en las puntuaciones entre Andalucía y cada una de las regiones disminuyen considerablemente una

⁴ Entre las variables explicativas consideramos el sexo del alumno, la edad, el estatus inmigrante, el interés del estudiante, si el alumno repite o no, el tiempo de estudio propio, la educación y ocupación de los padres, la disponibilidad de ordenador en el hogar, si tiene clases de recuperación dentro o fuera del colegio, y las horas de clases de ciencias en el colegio.

vez controlamos por este conjunto de variables, en la mayoría de los casos, éstas siguen siendo significativas e importantes. Sin embargo, para la comparación con las regiones de Cataluña, Resto de España y País Vasco, las diferencias con Andalucía dejan de ser significativas en los dos primeros casos, y cambian de signo en el último. Esto nos llama poderosamente la atención sobre la necesidad de comparar los resultados entre regiones teniendo en cuenta adecuadamente las diferencias existentes en las distintas variables que determinan el rendimiento educativo.

Cuadro 1. Comparación de las diferencias regionales observadas en la puntuación del examen de ciencias con las diferencias estimadas con MCO



Existe una extensa literatura sobre los posibles factores determinantes del rendimiento académico. Esta literatura agrupa dichos factores en tres grandes categorías: características

individuales, factores determinados fuera del sistema educativo o factores socio-familiares y educación recibida dentro del sistema educativo o características del colegio.⁵

Las diferencias en resultado académico medio entre regiones admiten dos posibles explicaciones. En primer lugar podemos atribuir las a la diferente composición de la población de alumnos en relación a estos tres grupos de variables explicativas. En este trabajo denominaremos “diferencias en características” a este primer componente. En segundo lugar, las diferencias pueden deberse a que las variables explicativas no tienen el mismo efecto sobre el resultado educativo en las distintas regiones (por ejemplo, el efecto de tener una madre universitaria sobre la nota del alumno es distinto entre regiones). A este segundo componente nos referiremos como “diferencias en impacto”.⁶ En este trabajo tratamos de cuantificar estos dos componentes aplicando una descomposición Oaxaca-Blinder. Éste análisis requiere estimar previamente el impacto de las distintas variables explicativas del rendimiento educativo y conocer la composición de las distintas regiones en cuanto a dichas variables. De este modo, es posible calcular cuánto se reducirían las diferencias regionales si la composición en relación a estas variables explicativas fuera la misma y cuánto disminuirían las diferencias si los impactos de las variables fueran los mismos entre regiones.

El objetivo de este artículo es, por tanto, explicar las diferencias existentes en los resultados académicos entre las regiones españolas participantes en el programa PISA-2006.⁷ Nuestra contribución al debate anteriormente mencionado consiste en dar respuesta a las siguientes preguntas: ¿Cuáles son los determinantes de las diferencias existentes entre

⁵ Las referencias son innumerables, desde el original Informe Coleman (1966) hasta trabajos más recientes como Heckman (2006).

⁶ Si bien la clásica literatura minceriana distingue entre impacto y retorno de una variable (el primer término se utiliza normalmente para variables relativas al entorno del alumno mientras que el segundo se utiliza con aquellas variables cuyo valor depende de decisiones que toma el estudiante o su familia) en este trabajo utilizaremos ambos términos de forma indistinta.

⁷ El objetivo de la OCDE es el de establecer comparaciones entre países y no regiones por lo que, a menos que las regiones decidan ampliar su muestra, la muestra será representativa solo a nivel de país. En España ampliaron muestra Andalucía, Aragón, Cantabria, Castilla y León, Cataluña, Galicia, La Rioja, Navarra y el País Vasco. Denominamos “Resto de España” al resto de regiones sin muestra ampliada, que analizaremos de forma conjunta.

regiones? ¿Cuáles serían las diferencias entre regiones si compartieran el mismo valor de las variables explicativas del rendimiento educativo, en concreto, el mismo nivel socioeconómico en las familias?

Existen varios trabajos con objetivos similares al nuestro. Los más cercanos son Ciccone y García-Fontes (2008), Bratti, Checchi y Filippin (2007) y Cordero, Crespo, y Santín (2009). El primero de estos trabajos cuantifica el efecto del nivel socio-familiar del alumno, en concreto el nivel educativo de sus padres y el estatus inmigrante, en las comparaciones entre los resultados de PISA 2006 para Cataluña y el resto de regiones españolas, por un lado; y para España y el resto de países europeos, por otro. Encuentran que el peor resultado educativo de Cataluña respecto a otras regiones mejora tan sólo ligeramente cuando controlan por el nivel educativo de los padres. Bratti *et al.* (2007) estudian las diferencias existentes en los resultados de PISA 2003 entre tres macro áreas italianas (Norte, Centro y Sur). Se centra en las diferencias debidas a diferente impacto de variables relacionadas con el sistema educativo de estas regiones (ej. infraestructuras educativas) y encuentran que este factor es determinante. Por último, Cordero *et al.* (2009) evalúan la eficiencia alcanzada por los alumnos pertenecientes a las diferentes regiones españolas, entendida como el grado de aprovechamiento de los recursos que éstos tienen a su disposición (entorno familiar, recursos de los colegios y características de los compañeros). Sus resultados muestran que las regiones más eficientes en el proceso educativo son Galicia, La Rioja y Aragón mientras que “Resto de España”, País Vasco y Cataluña son las más ineficientes.

El presente artículo complementa los trabajos anteriores al aplicar una descomposición Oaxaca-Blinder al análisis de las diferencias regionales en las puntuaciones obtenidas por los alumnos en PISA 2006. Este tipo de análisis permite tener en cuenta la heterogeneidad regional en cuanto a los impactos de las distintas variables explicativas en el resultado educativo y posteriormente descomponer las diferencias observadas en diferencias en impacto

y diferencias en características. Bratti *et al.* (2007) utilizan un enfoque similar. Sin embargo, a diferencia de ellos, nosotros no realizamos ningún supuesto sobre las diferencias en el impacto de las variables explicativas del resultado educativo en las distintas regiones.⁸

Nuestros resultados indican claramente que las variables familiares y socioeconómicas explican una parte importante de las diferencias entre regiones. Así, por ejemplo, cuando se comparan alumnos con entornos familiares similares, Andalucía, que es la región con peor resultado educativo, no está peor que regiones como Cataluña y País Vasco. Esto se debe a que las diferencias en características entre estas regiones son muy importantes. Sin embargo, Andalucía sigue obteniendo peor resultado educativo medio que otras regiones como Castilla y León y La Rioja, lo que indica que las diferencias en el impacto de estas características socioeconómicas y familiares son también muy relevantes. Además, encontramos que existen también importantes diferencias tanto en los valores medios como en el impacto de variables de esfuerzo individual del alumno así como de aquellas relacionadas con el colegio del mismo.

La estructura del texto que continúa es la siguiente. La sección 2 explica la muestra española de PISA-2006 así como las variables explicativas que utilizamos en el análisis. La sección 3 explica la metodología empleada, los modelos estimados y la estrategia econométrica seguida. La sección 4 muestra los resultados de las estimaciones realizadas. La sección 5 analiza la descomposición Oaxaca-Blinder de las diferencias regionales. Y por último, la sección 6 muestra nuestras conclusiones.

⁸ Bratti *et al.* (2007) suponen que el efecto de las variables individuales, así como de otras variables de recursos locales que incluyen en el análisis, es el mismo en todas las macro-regiones italianas que estudian.

2 La muestra española de PISA

2.1 Datos

La muestra del estudio PISA tiene un diseño estratificado en dos etapas. En primer lugar se selecciona una muestra aleatoria entre las escuelas con estudiantes de 15 años y en segundo lugar, se seleccionan a los estudiantes dentro de cada escuela. A su vez, los países participantes en PISA pueden hacerlo proporcionando muestras representativas a nivel regional.⁹ En este trabajo utilizamos la muestra de PISA correspondiente a España y sus regiones con muestra ampliada. De este modo podemos realizar un análisis estadístico y econométrico para la totalidad del territorio nacional y, especialmente, para aquellas regiones con muestra ampliada. La Tabla 1 muestra la relación de regiones con muestra representativa, la mayoría de las cuales tienen aproximadamente 1.500 estudiantes aunque hay que destacar el caso del País Vasco, región para la cual la muestra asciende a casi 4.000 alumnos. Denominamos “Resto de España” al conjunto de regiones que no tienen muestra ampliada.

2.2 Variables utilizadas

PISA evalúa el rendimiento educativo de los estudiantes en tres disciplinas: ciencias, matemáticas y lectura. Sin embargo, dado que la edición de 2006 está especialmente dedicada a las ciencias, esta materia tiene un mayor número de preguntas en el examen de evaluación. Además los alumnos contestaron a un gran número de preguntas sobre actitud y relación con las ciencias. Por ello, la variable que analizamos como variable dependiente en este trabajo es el resultado educativo (puntuación) obtenido en las pruebas de ciencias por cada alumno. Sin embargo, y debido a la posible existencia de errores de medida, la puntuación de cada estudiante no es la puntuación directa de cada estudiante en el test. Nosotros seguiremos la

⁹ Este es el caso si la muestra la forman en torno a 1.500 alumnos. Para más información sobre el diseño muestral de PISA ver OCDE (2006).

metodología recomendada por la OCDE y usaremos los cinco valores plausibles para calcular el resultado educativo de cada estudiante así como los errores estándar de los coeficientes estimados.¹⁰

Podemos clasificar las variables explicativas de la puntuación en tres grupos: individuales, socio-familiares (*informal schooling*) y de escuela (*formal schooling*). Las Tablas 2 a 6 muestran las variables explicativas utilizadas en nuestro estudio que comentamos brevemente a continuación.

(i) Variables individuales

Dentro de esta categoría consideramos el sexo, el mes de nacimiento y la nacionalidad del alumno. La distribución de los estudiantes por sexo es bastante homogénea (51% de chicos y 49% de chicas). En la segunda columna de la Tabla 2 podemos observar, para cada una de las regiones analizadas, el porcentaje de estudiantes inmigrantes de primera generación (aquellos nacidos fuera España y cuyos padres también nacieron en otro país). Del mismo modo, en la tercera columna de la Tabla 2 podemos observar, para cada una de las regiones analizadas, el porcentaje de estudiantes inmigrantes de segunda generación (aquellos que ya han nacido en España aunque sus progenitores sean de nacionalidad extranjera). Como refleja esta tabla, el porcentaje total de estudiantes inmigrantes en España es un 6,9%, lo cual significa un ratio dos puntos porcentuales inferior a la media de la OCDE. La mayoría de las regiones con muestra ampliada tienen un porcentaje de inmigrantes inferior a la media de España. Tan sólo Cataluña supera la media con un 9,5% de estudiantes no nativos. El porcentaje más bajo se observa en Andalucía (2,5% sobre el total de alumnos en la muestra).

¹⁰ Siguiendo lo establecido en la Item Response Theory (ver una descripción de esta teoría en, por ejemplo, Hambleton, Swaminathan y Rogers, 1991), para cada una de las tres áreas evaluadas cada alumno tiene cinco valores plausibles. Estos valores plausibles son considerados como posibles puntuaciones que razonablemente se pueden aceptar de un alumno de acuerdo con la realización de su examen. A su vez, estas puntuaciones individuales son estandarizadas de tal modo que la media de la muestra completa de PISA 2006 en toda la OCDE es igual a 500 y la desviación típica igual a 100 (ver apéndice metodológico AM1).

La siguiente variable que consideramos es el interés declarado por el estudiante en la ciencia. Utilizando el análisis cluster hemos construido un índice resumen de las preguntas del cuestionario cuyas respuestas reflejan el interés de los estudiantes ante la ciencia clasificando a los estudiantes entre aquéllos con interés alto, medio y bajo según sus respuestas (ver apéndice metodológico AM2). En las tres primeras columnas de la Tabla 3 observamos el porcentaje de alumnos en las categorías de interés alto, medio y bajo para cada una de las regiones analizadas. En general existe cierta homogeneidad entre las distintas regiones. Destacan Castilla y León y Andalucía, la primera por ser la región con más alumnos dentro de la categoría de interés alto, un 30%, y la segunda por presentar el caso contrario, un 24%.

Otra variable incluida en este grupo es el carácter repetidor o no del estudiante. Puesto que todos los alumnos de la muestra tienen la misma edad (15 años), deberían estar cursando 4º de ESO. Así, serán repetidores todos aquellos alumnos que se encuentren en un curso inferior a éste es decir, 1º, 2º ó 3º de ESO. Las columnas 5 y 6 de la Tabla 3 muestran el porcentaje de estudiantes que repiten uno o más cursos para las diferentes regiones. Existen diferencias regionales muy importantes respecto a esta variable. Por ejemplo, mientras en el País Vasco únicamente el 24% de los estudiantes son repetidores, esta cifra se dobla para el caso de Andalucía, 47,6%, siendo esta región, junto con Galicia, la que muestra un mayor porcentaje de alumnos que han repetido más de un curso.

Por último, consideramos el esfuerzo del alumno a través de dos variables individuales: la asistencia por parte de éste a clases particulares fuera del colegio y el tiempo de estudio propio.¹¹ Las columnas 6 y 7 de la Tabla 3 muestran los descriptivos de estas variables. Para el total de España, un 42% de estudiantes recibe clases particulares mientras que dicho porcentaje para la media de la OCDE es bastante inferior, 26%. Entre las regiones españolas también hay diferencias a este respecto. Así, las regiones con mayor y menor

¹¹ Para más detalle sobre la construcción de estas variables, ver Apéndice Metodológico AM3.

porcentaje de alumnos que reciben clases particulares son Asturias y Andalucía, con un 61% y 39%, respectivamente. Por lo que respecta al tiempo de estudio propio clasificamos a los alumnos entre aquéllos que dedican al menos dos horas a la semana al estudio en ciencias y los que estudian menos de dos horas. La columna 8 de la Tabla 3 muestra el descriptivo de esta variable que nos dice que el 34% de los alumnos españoles estudian ciencias al menos dos horas a la semana, un porcentaje superior al de la OCDE, con 25%. Castilla y León es la región cuyos alumnos dedican más tiempo al estudio propio de las ciencias, 41% con al menos dos horas de estudio, por el contrario, el País Vasco es la región donde los alumnos dedican menos tiempo de estudio propio a esta disciplina (solo un 28% dedica al menos dos horas al estudio de las ciencias).

(ii) Variables socio-familiares

Dentro de esta categoría consideramos el nivel educativo y el estatus ocupacional de los padres, así como la posesión de ordenador por parte de la familia. En primer lugar, la Tabla 4 muestra el nivel educativo de los padres y madres para cada una de las regiones estudiadas así como los valores medios para España y para la OCDE. Cabe destacar aquí que en España, un 26% de los padres tiene un nivel educativo no superior a educación primaria. Este porcentaje es muy superior al correspondiente para el conjunto de la OCDE: 10%. Andalucía es la región con el porcentaje más elevado de padres con niveles educativos inferiores o iguales a primaria: 39%. Las cifras para madres son muy similares.¹²

En segundo lugar, la Tabla 5 muestra el análisis descriptivo de la variable correspondiente al estatus ocupacional de los padres. Así, distinguimos entre padres con

¹² Según la EPA del segundo trimestre de 2006, el porcentaje de varones mayores de 16 años con nivel educativo inferior o igual a Primaria en el conjunto de España era del 30% en 2006 mientras que en PISA este porcentaje es del 26%. Existen regiones como Cataluña donde las diferencias van en este mismo sentido (30% según la EPA y 22% según PISA) mientras que en otras, como Andalucía, el sesgo va en sentido contrario (32% según la EPA y 39% según PISA). Dado que la información sobre la educación de los padres es proporcionada por el alumno, esto puede estar indicando que esta variable puede estar medida con cierto margen de error.

ocupación alta o baja¹³. Para las madres consideramos además la posibilidad de ejercer como ama de casa (el porcentaje de estudiantes que indicaron tener padres en esta categoría es prácticamente nulo). Destacamos aquí un denominador común entre todas las regiones y es el elevado porcentaje de padres y madres con ocupación baja (más del 63% para los padres y aproximadamente el 43% para las madres para el total de España), por encima en ambos casos de la media de la OCDE. Las regiones en las que hay mayor proporción de padres con ocupación alta son País Vasco y Cataluña con un 44% y 42.5%, respectivamente mientras que aquéllas en las que hay menos padres con alta ocupación son Castilla y León y Andalucía con un 33% y un 30% respectivamente.

En la última columna de la Tabla 5 observamos la distribución regional de la posesión de ordenador en casa. Como vemos, en todas las regiones la gran mayoría de los alumnos han indicado tener ordenador.

(iii) Variables de colegio o “formal schooling”

Las variables de colegio que consideramos son las que tienen variabilidad a nivel de estudiante. En concreto, las horas de clase de ciencias en el colegio y el hecho de si el alumno tiene clases de apoyo dentro del colegio o no. El resto de variables que miden características del colegio (titularidad, régimen de enseñanza, etc.) no las consideramos en este artículo. El motivo es la técnica econométrica que empleamos para estimar las diferencias entre regiones que, como se explica en la siguiente sección, controla por las características inobservables a nivel de colegio lo que impide estimar el efecto de las características de colegio que no tienen variabilidad a nivel individual.

Las horas de clase de ciencias en el colegio las hemos extraído directamente del cuestionario individual de los estudiantes. Así, hemos dividido a los alumnos en tres grupos:

¹³ La variable “ocupación alta” agrupa a las categorías 1 al 4 según la clasificación del ISCO (directivos de empresas privadas o altos cargos de la administración, profesionales que trabajan por cuenta ajena o profesionales liberales). La variable “ocupación baja” se corresponden con los trabajos en las categorías 5 al 9 de esta misma clasificación (trabajadores del sector servicios y trabajadores u operarios del sector industrial, trabajadores no cualificados).

aquéllos que reciben menos de dos horas de ciencias a la semana; los que reciben entre dos y cuatro horas de ciencias a la semana; y los que reciben más de cuatro horas a la semana. En la Tabla 6 podemos observar que los países de la OCDE tienen de media un 29% de alumnos que reciben más de cuatro horas de ciencias a la semana, cifra algo superior a la de España, un 27%. Por lo que respecta a las diferencias regionales podemos observar una gran disparidad entre la región que menos alumnos tiene en esta categoría, Cataluña con casi un 20% de estudiantes, y la región que mayor porcentaje presenta, Navarra con casi un 44%. Las diferencias entre regiones pueden deberse a dos motivos: distintos currícula (lo que implica diferencias regionales en horarios escolares) y distinta composición del conjunto de estudiantes en cuanto a la proporción de alumnos no repetidores (puesto que éstos tienen acceso a distintos itinerarios formativos que difieren en el número de horas de estudio de ciencias) y repetidores (no tienen acceso a distintos itinerarios).¹⁴

Finalmente, en la cuarta columna de la Tabla 6 podemos observar que no hay grandes diferencias regionales en cuanto al porcentaje de alumnos que va a clases de apoyo dentro del colegio, variando este porcentaje entre el 14% en la Rioja y el 20% en las regiones agrupadas bajo el nombre de Resto de España. Es importante distinguir esta variable de la descrita en la séptima columna de la Tabla 3 (alumno recibe clases particulares fuera del colegio). Los alumnos que asisten a clases de apoyo dentro del propio colegio son aquellos que pertenecen a un “itinerario especial” de enseñanza. Los motivos de pertenecer a este tipo de itinerarios pueden ser varios. Por una parte, puede tratarse de una medida del colegio dirigida a los alumnos que tienen un rendimiento especialmente bajo. Por otra parte, en este grupo pueden encontrarse alumnos que han pasado de curso con algunas asignaturas pendientes de años

¹⁴ En la actual legislación, los alumnos de 3º de ESO pueden elegir una optativa de ciencias de 2 horas semanales en las regiones de Andalucía, País Vasco, Asturias, Galicia, Castilla y León, Cataluña, Aragón y Navarra. En Navarra la optativa a elegir es de 3 horas semanales si se opta por el modelo lingüístico castellano. Por otra parte, por ejemplo, en Andalucía hay 4 horas semanales de ciencias de la naturaleza en 3º de ESO mientras que en las regiones de País Vasco y Navarra (si se elige el modelo de lengua castellana) se da una hora menos. Por otro lado, Galicia y Asturias tienen una hora más de Ciencias de la Naturaleza en 1º de ESO que la región andaluza.

anteriores.¹⁵ De este modo podemos suponer que estos alumnos son diferentes a aquéllos que no tienen clases de apoyo o incluso a aquéllos que tienen clases particulares. Por último hay que destacar que, si bien la asistencia a clases particulares fuera de la escuela puede ser el resultado de una decisión por parte del alumno (o de los padres), la asistencia a clases de apoyo puede ser el resultado de una decisión por parte del colegio, en cumplimiento de la normativa escolar.

3 Estimación de los Modelos Regionales con Efectos Fijos de colegio

El modelo econométrico que estimamos supone que para cada región k la puntuación en PISA de un determinado alumno j que asiste a un colegio concreto i , y_{ij} depende de: una constante μ^k ; $h=1, \dots, n$ variables individuales y socio-familiares recogidas del cuestionario individual ($x_{h,ij}$) cuyos valores pueden variar entre alumnos de un mismo colegio; de $l=1, \dots, m$ variables relativas al colegio del estudiante ($z_{l,i}$) que son recogidas en el cuestionario de los directores cuyos valores serán los mismos para los alumnos que asisten a un mismo colegio; y de una serie de variables inobservables relativas al colegio del alumno que se recogen en el parámetro α_i^k y que serán las mismas para todos los estudiantes que asisten a un mismo centro educativo. El modelo se puede representar por la siguiente expresión:

$$y_{ij} = \mu^k + \sum_{h=1}^n \beta_h^k x_{h,ij} + \sum_{l=1}^m \gamma_l^k z_{l,i} + \alpha_i^k + u_{ij} \quad (1)$$

¹⁵ De acuerdo con la actual legislación los alumnos de ESO pueden pasar de curso aunque no hayan superado todas las asignaturas, quedando pendientes de recuperación para el siguiente curso escolar.

La estrategia econométrica seguida, para conseguir una estimación consistente de los coeficientes de las variables, ha sido la utilización de la metodología generalmente aplicada a datos de panel con heterogeneidad inobservable. En concreto, hemos empleado el estimador intragrupos o de Efectos Fijos. En la medida en que las variables inobservables α_i^k influyan en el resultado educativo de un alumno y además estén correlacionadas con los regresores utilizados en el modelo, los coeficientes β_h^k y γ_i^k estarían sesgados si no se controla adecuadamente por la presencia de dichos efectos inobservables. Esta técnica permite una estimación consistente de β_h^k a partir de la siguiente transformación del modelo:

$$(y_{ij} - \bar{y}_i) = \sum_{h=1}^n \beta_h^k (x_{h,ij} - \bar{x}_{h,i}) + (u_{ij} - \bar{u}_i) \quad (2)$$

Siendo \bar{y}_i la nota media en el colegio i , $\bar{x}_{h,i}$ el valor medio de la variable h en el colegio i , y \bar{u}_i el error medio en el colegio i . Esta transformación de las variables permite que la heterogeneidad inobservable de los colegios no perjudique la estimación. Nótese que con esta especificación del modelo desaparecen las variables de colegio cuyos coeficientes, γ_i^k , no podrán ser estimados. En cualquier caso sus efectos quedan recogidos por la técnica econométrica utilizada ya que el estimador de Efectos Fijos controla por la heterogeneidad observable e inobservable a nivel de colegio. Una manera de no perder la estimación de estas características de colegio es estimar el modelo mediante el supuesto de efectos aleatorios. Sin embargo, hemos realizado el test de especificación de Hausman, para contrastar qué supuesto es más adecuado, dando como resultado que es preferible utilizar el modelo de Efectos Fijos para el total de la muestra española y para la mayoría de las regiones.¹⁶

¹⁶ El test de Hausman tiene como objetivo el rechazo o aceptación de la hipótesis nula de que no hay diferencias sistemáticas entre los coeficientes estimados con Efectos Fijos y los estimados con Efectos Aleatorios. El

4 Resultados de la estimación.

Los resultados de la estimación se muestran en la Tabla 7, donde cada columna muestra los coeficientes estimados del modelo (2) para cada una de las regiones. Como individuo de referencia, cuyo resultado aparece estimado en el término constante de cada regresión, consideramos a un alumno de sexo masculino, que nació en enero, es nativo, con un interés bajo en la ciencia, que no es repetidor, que no recibe clases particulares ni de apoyo, estudia ciencias menos de dos horas semanales, tiene padre y madre de nivel educativo y ocupación baja, y que vive en una casa en la que no hay ordenador. La estimación del resultado para los alumnos de referencia vascos es de 478, la menor de entre todas las regiones. Seguidamente se encuentra Andalucía y Resto de España con un coeficiente para la constante de 484 y 487 respectivamente. Las regiones cuyos alumnos de referencia obtienen mejores puntuaciones son Galicia (527) y Cantabria (525).

4.1. Variables individuales

Coincidiendo con cierta evidencia empírica previa (ver Hoxby, 2000, y Zinovyeva, Felgueroso y Vázquez, 2008) encontramos que el sexo del alumno tiene un efecto significativo en todas las regiones, en concreto existe un diferencial positivo a favor de los chicos.¹⁷ Galicia es la región donde es mayor el diferencial negativo estimado para las chicas, -26,37, seguida por Andalucía (-25,91), Cantabria (-24,12), y Cataluña (-22,48). En el otro extremo está Aragón, la región en la que las diferencias entre mujeres y hombres son menores, -7,27 puntos.

estadístico χ^2 que obtenemos para la muestra total española es de 84,44 por lo que podemos rechazar la hipótesis nula con un error de menos del 1%. En este caso es preferible utilizar Efectos Fijos ya que es el estimador insesgado. Hemos realizado este mismo test para las distintas muestras regionales, obteniendo el mismo resultado para la mayoría de ellas. Tan sólo para tres regiones (Andalucía, Aragón y Castilla y León) se acepta la hipótesis nula.

¹⁷ Encontramos un resultado similar para la disciplina de matemáticas. Sin embargo, para la disciplina de lectura las chicas obtienen mejor nota que los chicos. Estos resultados no los mostramos por razones de espacio pero están disponibles para el lector interesado.

La variable edad sólo es significativa para las regiones de Cantabria, Aragón, Galicia, Navarra y País Vasco. Para todas ellas el efecto es negativo, lo que podemos interpretar como que los estudiantes que nacieron en los últimos meses del año tienen peor resultado educativo que los que nacieron a principio de año.¹⁸

La variable de estatus inmigrante es significativa y negativa en la mayoría de las regiones, destacando Castilla y León (-56,63). Es decir, en general los inmigrantes obtienen peor nota en PISA que los alumnos similares a ellos pero de padres nacionales. Nótese, no obstante, que el coeficiente estimado para algunas regiones como Asturias es prácticamente cero. Zinovyeva et al. (2008), con datos de PISA 2003 y 2006, obtienen un resultado similar.

En las distintas regresiones el interés del alumno en la ciencia tiene un efecto estimado sobre la puntuación muy importante y estadísticamente significativo. En cuanto a la variable de “interés medio”, el coeficiente va desde los 13,84 de La Rioja a los 27,83 de Aragón. El coeficiente para el grupo de “interés alto” en Andalucía es el menor de todos con 38,19 puntos. Por el contrario el coeficiente de la región de Cataluña es el mayor con 60,12 puntos. Por tanto, existen importantes diferencias regionales en el impacto que tiene sobre el rendimiento educativo que el alumno muestre un interés alto sobre las ciencias, aún cuando comparamos alumnos de iguales valores en el resto de variables.

La variable que indica la repetición de curso por parte del alumno tiene un coeficiente negativo y muy importante para todas las regiones si bien hay diferencias entre ellas. Las regiones donde el impacto de esta variable es más negativo son Navarra y Galicia. Aquí, los alumnos que han repetido un curso (los matriculados en 3º de ESO) obtienen 69,18 y 68,84 puntos menos, respectivamente, y los que han repetido dos o más cursos (los matriculados en un curso inferior a 3º de ESO) 116,06 y 116,68 puntos menos, respectivamente. El caso opuesto lo encontramos en Cataluña y Asturias. Aquí, los alumnos que han repetido un curso

¹⁸ Nuestros resultados coinciden con los encontrados en la literatura en relación al efecto positivo de la edad sobre el rendimiento académico (ver, entre otros Acemoglu y Angrist, 1999 y Angrist y Krueger, 1990).

obtienen 35,67 y 38,78 puntos menos respectivamente, y los que han repetido dos o más cursos 83,62 y 86,41 puntos menos respectivamente.¹⁹

Respecto a las variables que miden el esfuerzo del alumno encontramos en primer lugar que las horas de estudio propio son no significativas para la mayoría de regiones excepto Cataluña cuyo coeficiente estimado es -10,24. En segundo lugar vemos que el coeficiente estimado para los alumnos que reciben clases particulares fuera del colegio es negativo y significativo para todas las regiones. Sin embargo hay que tomar la interpretación de este efecto con cierta precaución porque esta variable puede estar sufriendo problemas de endogeneidad.²⁰ En general cabría esperar una relación positiva entre esfuerzo del estudiante y su resultado educativo.²¹ Sin embargo, también es posible encontrar una relación negativa entre ambas variables. La intuición es la siguiente. Cuando observamos que un alumno asiste a clases particulares o que dedica mucho tiempo al estudio puede ser porque su resultado educativo previo ha sido malo y/o porque su habilidad es peor que la del resto de alumnos, por lo que necesita dedicar más tiempo al estudio. Esta hipótesis solo se podría contrastar adecuadamente comparando el rendimiento del alumno en momentos distintos del tiempo pero lamentablemente no existe esta posibilidad al ser PISA un estudio de corte transversal.

4.2. Variables socioeconómicas

El número de categorías de las variables de educación del padre y de la madre las hemos reducido a dos en nuestra especificación econométrica: baja, si el padre (madre) tiene

¹⁹ En García-Pérez, Hidalgo-Hidalgo y Robles (2010) realizamos un estudio pormenorizado de la variable “ser repetidor” y su impacto sobre las puntuaciones de los alumnos, haciendo especial hincapié en el posible carácter endógeno de esta variable.

²⁰ La presencia de endogeneidad en alguno de los regresores del modelo puede hacer que el coeficiente asociado con dicha variable no se estime de manera consistente, lo cual puede causar problemas importantes si queremos interpretar causalmente el efecto estimado de dicha variable. Sin embargo, dado que el objetivo de este trabajo es predecir y comparar el resultado educativo por regiones, este problema puede ser menos relevante ya que la predicción con el modelo completo no tiene por qué verse afectada de manera importante por el posible sesgo de alguno de sus coeficientes.

²¹ De Fraja et al. (2010) encuentra que el impacto de esfuerzo del estudiante sobre su resultado académico es positivo, aunque muy bajo.

estudios secundarios medios (EGB) o una titulación menor; y alta, si el padre (madre) tiene una titulación mayor a ésta. En la estimación incluimos la interacción de las variables educación y ocupación, tanto para el padre como la madre. Así, en el término independiente o constante de nuestro modelo tenemos a los estudiantes de padre con educación y ocupación baja, incluyendo una serie de variables binarias que presentan las distintas combinaciones entre padre con educación baja o alta, y ocupación baja o alta. Incluimos la educación y ocupación de la madre de forma análoga añadiendo una variable binaria que indica si la madre es ama de casa o no. Nuestros resultados coinciden con los ya establecidos en la literatura: el nivel de educativo de los padres influye de forma positiva e importante en el resultado académico del estudiante.²² Por ejemplo, los estudiantes de Asturias que tienen padre (madre) con educación y ocupación alta obtienen 18,08 (13,99) puntos más que los estudiantes con padre (madre) de baja educación y ocupación. Lo mismo ocurre en La Rioja y Aragón. Por el contrario, en algunas regiones es sólo uno de los padres el que parece influir más sobre el rendimiento del alumno. Así, por ejemplo, en Andalucía o Cantabria tener una madre con estudios y ocupación alta hace que el rendimiento suba 13 y 18 puntos, respectivamente. Por otra parte, en algunas regiones como Andalucía, La Rioja y País Vasco, los estudiantes que tienen una madre que ejerce de ama de casa, obtienen un resultado académico inferior a aquellos cuya madre trabaja fuera del hogar. Finalmente, es interesante que en algunas regiones el estatus socioeconómico de las familias no muestre un impacto significativo. Se trata de las regiones de Castilla y León, Cataluña y Galicia, donde ninguna variable de este grupo parece causar ningún efecto sobre el rendimiento de los alumnos.

²² Entre otros, ver Heckman (2006), que encuentra que las diferencias en rendimiento académico entre estudiantes con distinto perfil socioeconómico aparecen a edades muy tempranas, y se mantienen en niveles educativos posteriores.

Por último encontramos que tener un ordenador en casa no implica una mayor puntuación en el examen de ciencias salvo para tres regiones donde el efecto es positivo (entre 15 y 18 puntos): Andalucía, Aragón y Resto de España.²³

4.3. Variables de colegio

En primer lugar es importante recordar que en este grupo de variables sólo es posible estimar los coeficientes asociados a regresores con variación a nivel individual, esto es, a nivel de alumno. Concretamente estimamos el efecto de que el alumno tenga clases de apoyo dentro del colegio y el de las horas semanales de clases de ciencias en el colegio. Hemos interactuado estas variables entre sí, así como con la variable que indica si el alumno es repetidor o no. El objetivo de esta especificación es captar el efecto que tienen las horas de ciencias en el colegio en función de la asistencia a clases de apoyo en el mismo colegio y en función de si el alumno es repetidor o no. La razón de las interacciones es buscar posibles diferencias en el impacto de estas variables para alumnos con/sin clases de apoyo o que son repetidores o no.

En primer lugar, encontramos que el efecto que tiene asistir a clases de apoyo dentro del colegio es negativo para todas las regiones, aunque el coeficiente es no significativo para dos de ellas: Aragón y Galicia. A modo de ejemplo, en Cataluña los alumnos con este tipo de clases de apoyo tienen 35,5 puntos menos.

En segundo lugar, el efecto que tienen las horas de clase de ciencias en el colegio es positivo y muy significativo en todas las regiones analizadas. A modo de ilustración la región en la que esta variable es más importante es La Rioja. Para esta región se estima que los alumnos que tienen entre 2 y 4 horas de ciencias obtienen 19,60 puntos más que los que tienen menos de 2 horas, y aquéllos alumnos que tienen más de 4 horas obtienen 45,36 puntos

²³ Recientemente, Spiezia (2010), también con datos de PISA 2006, encuentra una relación positiva entre el uso de ordenador en casa y el rendimiento académico del estudiante.

más. En general los resultados muestran que para los alumnos repetidores, independientemente de si tienen clases de apoyo en el mismo colegio o no, el efecto de las horas de ciencias en el colegio es menor que para los no repetidores. Sin embargo, la significatividad de estas variables es muy reducida, tan sólo se encuentra este efecto significativo para Asturias, Cantabria y País Vasco.

5 Descomposición *Oaxaca-Blinder*.

El análisis anterior permite conocer, para cada región, los efectos que tienen las variables influyentes en el resultado educativo. Nuestro objetivo ahora es descomponer las diferencias en rendimiento entre las distintas regiones en “diferencias en características” entre alumnos de dichas regiones y “diferencias en el impacto” de dichas características sobre el propio resultado del alumno. Para ello realizamos la siguiente descomposición *Oaxaca-Blinder*:

$$\bar{Y}^b - \bar{Y}^a = \hat{\beta}^b (\bar{X}^b - \bar{X}^a) + \bar{X}^a (\hat{\beta}^b - \hat{\beta}^a) \quad (3)$$

donde \bar{Y}^b e \bar{Y}^a representan la nota media en PISA de las regiones b y a respectivamente; \bar{X}^b y \bar{X}^a son los vectores que representan las características medias de ambas regiones, y $\hat{\beta}^a$ y $\hat{\beta}^b$ son los vectores de coeficientes estimados para cada región, representando el impacto que tienen las variables del modelo estimado en dicha región. Por tanto, se puede interpretar que $\hat{\beta}^b (\bar{X}^b - \bar{X}^a)$ es la parte de las diferencias regionales que se debe únicamente a diferencias en las características de las regiones, esto es, diferencias en las variables individuales de los alumnos, en las variables familiares y en las variables de colegio o “*formal schooling*”. Y por otro lado, $\bar{X}^a (\hat{\beta}^b - \hat{\beta}^a)$ se interpreta como la parte de las diferencias regionales que se debe

únicamente a diferencias en los impactos de las anteriores características sobre el rendimiento de los alumnos.

En la medida en que los coeficientes estimados para las distintas regiones sean similares, las diferencias regionales se deberían sobre todo a las “diferencias en características” entre dichas regiones. Por el contrario, si los coeficientes son muy distintos entre regiones, las “diferencias en impacto” serán más importantes.

La Tabla 8 muestra las diferencias en resultado educativo observadas entre las distintas regiones y Andalucía y la descomposición de estas diferencias en *impacto* y en *características*.²⁴ Las mayores diferencias se encuentran entre la región andaluza y La Rioja (44,90). Por el contrario, las diferencias son menores con País Vasco (20,20), Cataluña (14,13) y Resto de España (11,05).

En la Tabla 8 observamos que hay seis regiones para las que las diferencias en características explican menos del 50% de las diferencias totales con Andalucía. Éstas son Galicia, Cantabria, La Rioja, Castilla y León, Aragón y Asturias. Es decir, las diferencias entre las puntuaciones medias de estas regiones y Andalucía se explican sobre todo debido a diferencias en los impactos de dichas características sobre el resultado del alumno.

Sin embargo, las diferencias entre las puntuaciones medias de regiones como País Vasco, Cataluña y Navarra respecto a Andalucía, se explican principalmente a partir de las diferencias en características entre las mismas. Es decir, si los estudiantes del País Vasco o Cataluña tuvieran características similares a los de Andalucía, los primeros obtendrían peor nota media que los últimos.

La Tabla 8 proporciona información adicional sobre qué parte de las diferencias totales se debe a las diferencias en las constantes de los modelos, diferencias en características

²⁴ Las diferencias observadas entre las distintas regiones mostradas en la Tabla 8 son ligeramente distintas a las diferencias a nivel descriptivo que mostramos en la Tabla 1. Esto es debido a que la descomposición de *Oaxaca-Blinder* tiene en cuenta sólo las muestras regionales que quedan después de eliminar a los estudiantes que presentan *missing values* en las variables explicativas del modelo estimado.

o diferencias en impacto de cada grupo concreto de variables. Las diferencias en el término constante entre las regresiones de dos regiones pueden interpretarse como diferencias en aspectos no medidos por el modelo estimado. Dada la interpretación de dichos coeficientes (el valor de la variable dependiente para el caso que todos los regresores tienen un valor igual a cero), parece razonable concluir que las diferencias se deben a factores que no se pueden asociar a ninguna de las variables explicativas usadas en el modelo estimado. Nótese que estas diferencias explican una parte muy importante del total de diferencias observadas para la mayoría de regiones. El caso extremo es Cataluña: la diferencia en constantes entre esta región y Andalucía explica en un 146,3% la diferencia total entre las mismas, es decir que si Andalucía tuviera la constante de Cataluña superaría con creces la nota media de ésta última. Sin embargo, son las diferencias en el impacto de las características sí tenidas en cuenta en nuestro modelo, en especial las referidas a las características familiares de los alumnos, las que están reduciendo esta diferencia hasta el porcentaje mostrado en la tabla. Lo contrario ocurre para el País Vasco donde la diferencia en constantes es negativa, indicando que las diferencias no observables entre Andalucía y esta región están reduciendo las diferencias observadas, al igual que ocurre con el impacto de las características familiares.

Las variables familiares o socioeconómicas tienen un gran peso al explicar las diferencias regionales, sobre todo debido a diferencias en características. Por ejemplo, sólo las diferencias en características referidas a variables familiares explican un 65,35% de las diferencias regionales entre Cataluña y Andalucía. Sin embargo, las diferencias en el impacto de las variables familiares no contribuyen a explicar las diferencias regionales observadas sino todo lo contrario: si Andalucía tuviese el impacto observado en Cataluña de dichas variables socioeconómicas, las diferencias con esta región aumentarían en un 236,44%. Esto se debe a que el impacto que tienen algunas variables familiares en Andalucía (como por ejemplo la ocupación de la madre o la existencia de un PC en casa) es bastante mayor que el

que tienen en Cataluña. Por ello, a igualdad de dotaciones, Andalucía muestra diferencias mayores entre alumnos de madres trabajadoras y aquéllos cuyas madres son, por ejemplo, amas de casa. Como el porcentaje de madres trabajadoras es menor en Andalucía, estas diferencias en el impacto contribuyen a reducir las diferencias observadas entre ambas regiones.

Las características de los alumnos también tienen un peso importante en la explicación de las diferencias regionales debido a diferencias en características. Por ejemplo, representan más de un 25% de las diferencias totales observadas entre Andalucía y País Vasco, Cataluña, Navarra, y Resto de España. Las diferencias en impacto de las variables individuales son muy importantes en la descomposición de las diferencias observadas con Galicia (-43,91%), País Vasco (52,58%) y Cantabria (-28,40%).

Por último encontramos que el conjunto de variables que hemos considerado como variables de colegio o “*formal schooling*” no son muy importantes a la hora de explicar las diferencias regionales aunque, en el caso de Andalucía, País Vasco y Cataluña llegan a explicar hasta un 15% del total de diferencias observadas.

6 Conclusiones.

Este trabajo tiene como objetivo explicar las diferencias en resultado educativo medido con la puntuación obtenida en el programa PISA 2006 entre los estudiantes de distintas regiones españolas. Para ello realizamos una descomposición Oaxaca-Blinder de las diferencias regionales, distinguiendo así entre diferencias debidas a características y diferencias debidas al distinto impacto de las variables explicativas del resultado educativo.

Nuestros resultados indican que las diferencias en las características socio-familiares de los alumnos explican una parte importante de las diferencias observadas en las

puntuaciones de los estudiantes. Esto es particularmente llamativo en el caso de las comparaciones entre Andalucía y regiones como País Vasco, Cataluña y Navarra. Sin embargo, las diferencias persisten cuando, manteniendo constantes las características socio-familiares, comparamos Andalucía con regiones como La Rioja, Castilla y León, Aragón, Cantabria, Galicia y Asturias. En estos casos existen diferencias tanto en dotaciones como en el impacto de distintas variables relacionadas con el esfuerzo del alumno y con el colegio del mismo que explican la persistencia de dichas diferencias. Por ejemplo, encontramos que las diferencias en características de los sistemas educativos explican más del 15% de las diferencias en las puntuaciones medias entre Andalucía, País Vasco y Cataluña.

Por tanto, son diferencias en el sistema educativo, ya sean las captadas por nuestras variables de colegio así como otras no incluidas en nuestro análisis, y reflejadas en las diferencias entre los términos independientes de cada regresión regional, las que están detrás de un porcentaje muy importante de las diferencias observadas. Las diferencias en la constante de nuestros modelos indican que hay características inobservables, y que deben estar relacionadas con el propio sistema educativo de cada región, que son muy importantes de cara a explicar las diferencias regionales.

Estos resultados tienen gran interés desde una perspectiva de política educativa, pues indican que existe cierto margen de reducción de las diferencias regionales en los resultados educativos mediante el correcto diseño de los sistemas educativos regionales. Si no se toman medidas a tiempo, las diferencias regionales en nivel educativo aumentarán en el futuro, y tal y como indican nuestros resultados, esto originará aún más diferencias en los resultados educativos de futuras generaciones de estudiantes, dado el enorme impacto de las variables socioeconómicas y de contexto educativo sobre los mismos.

Referencias bibliográficas

- (1) Acemoglu, D. (2009) Introduction to modern economic growth, Princeton, NJ: Princeton University Press.
- (2) Acemoglu, D. y Angrist, J. (1999) "How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws", NBER Working Papers 7444,
- (3) Angrist, J. D. y Krueger, A. B. (1990) "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples", NBER Working Papers 3571, National.
- (4) Bratti, M.; Checchi, D.; y Filippin, A. (2007) "Geographical differences in Italian students' mathematical competencies: evidence from PISA 2003", *Giornale degli Economisti e Annali di Economia* (66) 3: 299-333.
- (5) Ciccone, A. y Garcia-Fontes, W. (2009) "The quality of the Catalan and Spanish education systems: A perspective from PISA", IESE Research Papers D/810, IESE Business School.
- (6) Coleman, J.; Campbell, E.Q.; Hobson, C.F.; McPartland, J.; y Mood, A.M. (1966) "Equality of Educational Opportunity", Washington, US Office of Education.
- (7) Cordero, J.M.; Crespo, E.; y Santín, D. (2009) "Factores explicativos de los resultados de las Comunidades Autónomas españolas en PISA 2006". Instituto de Estudios Fiscales Papeles de Trabajo, 14/09.
- (8) De Fraja, G.; Oliveira, T.; y Zanchi, L. (2010) "Must try harder. Evaluating the role of effort on examination result", *Review of Economics and Statistics*, 92: 577-597
- (9) García-Pérez, J. I.; Hidalgo-Hidalgo, M.; y Robles, J. A. (2010) "Does grade retention play any role on accounting for the differences in PISA test scores across Spanish regions?", Mimeo, Universidad Pablo de Olavide.

- (10) Hambleton, R.K.; Swaminathan, H.; y Rogers, H.J. (1991) "Fundamentals of item response theory", Newbury Park, California: Sage Publications.
- (11) Heckman, J.J. (2006) "Skill formation and the economics of investing in disadvantaged children" *Science*, **312**, 5782: 1900-1902.
- (12) Hoxby, C. (2000) "Peer Effects in the Classroom: Learning from Gender and Race Variation" NBER Working Paper 7867.
- (13) OCDE (2006) "Assessing Scientific, Reading and Mathematical Literacy: A framework for PISA 2006".
- (14) OCDE (2008) "Informe PISA 2006. Competencias científicas para el mundo del mañana".
- (15) OCDE (2009) "PISA 2006 Technical Report".
- (16) Spieza, V. (2010) "Assesing the impact of ICT use on PISA scores" Mimeo OECD.
- (17) Zinovyeva, N., Felgueroso, F. y Vazquez, P. (2008). "Immigration and Students' Achievement in Spain," Working Papers 2008-37, FEDEA.

Tabla 1: Resultado educativo por Comunidades Autónomas

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

Tabla 2: Estatus inmigrante de los alumnos en cada región.

Región	Estatus inmigrante(%)		
	Nativos	Primera Generación	Segunda generación
Andalucía	97,5	2,1	0,4
Aragón	94,6	5,0	0,4
Asturias	97,4	2,1	0,6
Cantabria	96,1	3,6	0,3
Castilla y León	97,2	2,7	0,1
Cataluña	90,6	7,9	1,6
Galicia	97,6	1,5	0,8
La Rioja	93,8	5,8	0,4
Navarra	92,7	6,9	0,4
País Vasco	96,2	3,5	0,3
Resto de España	89,8	9,2	1,0
Total España	93,1	6,1	0,8
OCDE	90,8	5,1	4,1

Nota: Los datos de la OCDE que se toman como referencia en este trabajo se obtienen a partir de la media aritmética de los 30 países miembros.

Tabla 3: Interés, carácter repetidor, clases particulares, y estudio propio en ciencias por regiones.

Región	Índice de interés(%)			Carácter repetidor del alumno(%)			Clases particulares (%)	Estudio propio (%)
	Alto	Medio	Bajo	4º ESO (No Repite)	3º ESO (Repite un curso)	< de 3º ESO (Repite más de un curso)		
Andalucía	24,0	48,4	27,6	52,3	37,2	10,4	39,5	35,6
Aragón	26,7	44,6	28,7	62,4	30,6	7,0	48,0	37,6
Asturias	28,8	45,5	25,7	65,7	28,1	6,2	61,4	37,3
Cantabria	27,3	47,1	25,6	60,6	32,9	6,5	57,9	36,3
Castilla y León	30,4	47,9	21,7	62,3	31,4	6,4	47,8	40,3
Cataluña	25,2	46,9	27,9	69,8	27,1	3,1	33,6	30,9
Galicia	26,5	48,2	25,3	60,6	28,7	10,7	49,7	31,5
La Rioja	29,1	46,6	24,4	62,2	32,6	5,3	57,9	36,5
Navarra	27,6	39,8	32,6	70,2	25,1	4,6	47,9	35,5
País Vasco	27,2	42,1	30,7	76,4	19,7	3,9	56,3	27,8
Resto de España	26,0	47,3	26,7	57,4	35,6	7,0	41,9	35,9
Total España	25,9	47,2	27,0	59,8	33	7,1	42,4	34,8
OCDE	(3)	(3)	(3)	(3)	(3)	(3)	26,2	25,3

Notas: Los datos de la OCDE que se toman como referencia en este trabajo se obtienen a partir de la media aritmética de los 30 países miembros.

Las clases de apoyo con un profesor particular son individuales o en grupo y se realizan fuera del colegio.

El índice *cluster* de interés y la variable que indica repetición de curso no se muestra para la OCDE por falta de homogeneidad de los datos entre países. En el caso del índice *cluster* éste fue creado con la muestra española únicamente. Igualmente puede que lo que se considere un alumno repetidor en España no sea igual en el resto de países.

Tabla 4: Nivel educativo de los padres por regiones

Región	Primaria o menos (%)		Secundaria (%)		Enseñanza superior (%)	
	Padres	Madres	Padres	Madres	Padres	Madres
Andalucía	39	39,6	39,5	44,6	21,5	15,9
Aragón	16,1	13,9	53,7	57,3	30,3	28,8
Asturias	14,3	10,8	54,2	60,9	31,5	28,4
Cantabria	12,5	11,7	58,7	62,6	28,8	25,8
Castilla y León	17,6	15,1	54,3	56,8	28,2	28,1
Cataluña	21,6	20	45	49,3	33,4	30,8
Galicia	21,1	19,9	54,3	56,7	24,7	23,5
La Rioja	16,2	13,6	53,5	58,7	30,4	27,7
Navarra	18,1	14,7	48	55,1	33,8	30,2
País Vasco	13,8	13	46,9	52,5	39,2	34,4
Resto de España	24,8	23,4	49,9	53,1	25,3	23,5
Total España	25,7	24,6	47,5	51,5	26,8	24
OCDE	10	11,2	54,9	55,2	36	34,8

Nota: Los datos de la OCDE que se toman como referencia en este trabajo se obtienen a partir de la media aritmética de los 30 países miembros.

Tabla 5: Estatus Ocupacional de los padres y porcentaje de alumnos con ordenador en casa por regiones

Región	Ama de casa(%)	Baja ocupación(%)		Alta ocupación(%)		Tiene PC en casa (%)
	(Madres)	Padres	Madres	Padres	Madres	
Andalucía	37,1	69,9	42,8	30,1	20,1	85,3
Aragón	21	60,3	43,8	39,7	35,2	92,9
Asturias	29,9	63,1	39,4	36,9	30,8	90,2
Cantabria	26,7	62,6	44,4	37,5	29	90,5
Castilla y León	28,6	67,3	41,7	32,7	29,7	92,5
Cataluña	12,7	57,5	48,8	42,5	38,6	94,4
Galicia	25,4	65,7	45	34,3	29,6	83,9
La Rioja	23,3	61,8	45,4	38,2	31,3	93,6
Navarra	19,3	64,5	45,7	35,6	35,1	92,4
País Vasco	19,5	56,0	43,2	44,0	37,4	94,7
Resto de España	27,3	62,0	41,2	38,0	31,6	90,5
Total España	26,7	63,3	43,1	36,7	30,2	90,0
OCDE	12,9	46,2	13,1	53,8	70,0	88,9

Nota: Los datos de la OCDE que se toman como referencia en este trabajo se obtienen a partir de la media aritmética de los 30 países miembros.

Tabla 6: Distribución del tiempo dedicado al estudio de las ciencias en el colegio y clases de apoyo en el colegio por regiones

Región	Clases de ciencias en el colegio(%)			Clases de apoyo en el colegio(%)
	Menos de 2 horas	Entre 2 y 4 horas	Más de 4 horas	
Andalucía	29,7	43,6	26,7	17,6
Aragón	26,1	44,0	30,0	17,0
Asturias	29,3	42,1	28,6	16,7
Cantabria	25,8	43,6	30,6	17,0
Castilla y León	24,5	44,3	31,2	16,0
Cataluña	15,8	64,5	19,7	19,4
Galicia	30,1	45,1	24,8	16,0
La Rioja	23,5	47,6	28,9	13,9
Navarra	30,8	25,6	43,6	16,1
País Vasco	26,4	48,5	25,2	16,6
Resto de España	31,1	40,2	28,7	20,4
Total España	27,7	45,4	26,9	18,7
OCDE	33,1	38,7	29,2	-

Nota: Los datos de la OCDE que se toman como referencia en este trabajo se obtienen a partir de la media aritmética de los 30 países miembros.

	Andalucía	Aragón	Asturias	Cantabria	Cast. León	Cataluña	Galicia	La Rioja	Navarra	P. Vasco	Resto Esp.
V. INDIVIDUALES											
Sexo (mujer=1)	-25,91*** (4,23)	-7,27* (4,32)	-20,62*** (3,72)	-24,12*** (3,73)	-14,42*** (3,44)	-22,48*** (3,58)	-26,37*** (3,95)	-14,38*** (5,56)	-16,14*** (4,17)	-13,17*** (2,05)	-16,08*** (2,99)
Mes de nacimiento (Enero=0)	-0,70 (0,64)	-1,45** (0,69)	-0,43 (0,53)	-1,39** (0,54)	-0,06 (0,47)	-0,66 (0,58)	-1,04* (0,58)	-0,84 (0,59)	-0,89* (0,51)	-0,71** (0,34)	-0,79 (0,49)
Inmigrante	-18,63 (13,11)	-30,70*** (9,77)	-6,46 (16,66)	-32,53** (13,00)	-56,63*** (12,84)	-42,80*** (8,53)	-14,76 (16,77)	-30,22*** (10,25)	-12,68 (11,05)	-19,25** (9,79)	-21,73*** (6,27)
Interés medio	17,62*** (3,86)	27,83*** (5,32)	19,90*** (4,52)	14,86*** (4,22)	18,57*** (5,66)	16,77*** (5,10)	16,85*** (5,51)	13,84** (6,35)	21,11*** (5,62)	20,57*** (2,95)	16,79*** (3,63)
Interés alto	38,19*** (6,10)	50,07*** (6,29)	40,78*** (6,38)	48,19*** (4,11)	47,43*** (6,29)	60,12*** (5,17)	48,76*** (6,19)	39,12*** (7,62)	53,44*** (7,71)	49,81*** (3,83)	41,30*** (4,30)
Repite un curso	-53,36*** (6,95)	-44,72*** (7,27)	-38,78*** (9,20)	-62,12*** (6,10)	-54,16*** (7,10)	-35,67*** (13,83)	-68,84*** (8,49)	-60,55*** (11,41)	-69,18*** (9,70)	-57,50*** (6,08)	-53,36*** (5,76)
Repite 2 o más cursos	-95,41*** (9,53)	-108,36*** (8,01)	-86,41*** (12,25)	-100,42*** (9,50)	-92,98*** (13,28)	-83,62*** (16,09)	-116,68*** (12,02)	-106,93*** (13,75)	-116,06*** (16,71)	-97,67*** (9,34)	-99,95*** (8,41)
Clase particulares	-12,83*** (3,50)	-21,26*** (3,97)	-22,58*** (4,14)	-12,20*** (3,79)	-19,12*** (4,09)	-14,52*** (5,38)	-21,51*** (4,01)	-20,30*** (4,63)	-11,89*** (3,93)	-9,00*** (2,03)	-20,87*** (3,96)
Estudio propio (>=2 horas/sem.)	6,86 (4,84)	-3,45 (5,71)	-5,74 (3,96)	-6,23 (4,29)	0,71 (4,23)	-10,28* (5,69)	-4,10 (4,94)	4,97 (5,29)	-5,21 (5,21)	0,30 (3,63)	0,56 (3,36)
V. SOCIOECO.											
P_OcuAlta_EduAlta	9,21 (7,24)	11,91** (4,73)	18,08*** (6,09)	5,80 (5,21)	10,02 (6,55)	8,73 (6,48)	5,63 (5,87)	16,08** (6,64)	18,61*** (6,39)	5,43 (4,03)	7,97 (5,80)
P_OcuAlta_EduBaja	-3,42 (8,42)	19,44*** (7,38)	17,39 (10,64)	10,55 (9,28)	9,38 (9,33)	-7,86 (7,63)	12,11 (7,67)	8,64 (8,51)	7,87 (10,92)	4,15 (5,94)	9,26 (7,15)
P_OcuBaja_EduAlta	-3,26 (6,76)	-4,05 (4,47)	8,07 (5,92)	-0,27 (5,21)	0,52 (4,67)	1,65 (6,14)	-2,42 (5,55)	13,10** (6,25)	-2,86 (4,80)	2,21 (3,45)	-0,31 (5,27)
M_OcuAlta_EduAlta	13,09** (6,57)	16,63*** (4,64)	13,99** (6,22)	17,92*** (5,91)	7,60 (6,99)	11,12 (7,01)	8,15 (7,04)	12,49* (7,17)	6,13 (5,75)	6,50* (3,36)	12,32* (6,57)
M_OcuAlta_EduBaja	24,66*** (9,21)	5,60 (8,97)	16,17 (11,49)	13,93 (10,00)	-9,30 (8,91)	-4,96 (8,45)	7,22 (10,00)	-0,48 (13,19)	2,37 (12,21)	6,43 (6,53)	12,49 (8,94)
M_OcuBaja_EduAlta	15,90** (8,06)	1,19 (6,54)	10,53* (5,38)	0,57 (4,56)	2,47 (5,49)	0,35 (6,86)	4,09 (7,40)	-1,88 (6,83)	4,22 (6,02)	1,28 (3,56)	8,99 (5,86)
M_ama de casa	-12,29** (4,82)	-5,98 (5,46)	-4,53 (4,83)	-5,34 (4,48)	-4,62 (4,84)	-10,45 (6,80)	-6,27 (6,34)	-11,34* (6,71)	-10,32 (6,82)	-9,55** (4,13)	-2,32 (3,77)
PC>=1	17,28** (6,85)	17,38* (9,87)	3,43 (8,50)	6,52 (7,26)	-0,87 (7,43)	-14,09* (7,97)	7,33 (5,64)	11,98 (8,89)	2,70 (10,53)	10,86 (7,11)	14,78* (7,87)
V. DE COLEGIO											
Clase apoyo colegio	-32,84*** (6,90)	-10,50 (9,87)	-30,64*** (10,97)	-19,89*** (6,00)	-27,40** (11,02)	-35,55*** (12,28)	-16,24 (10,02)	-26,20** (11,95)	-34,27*** (8,30)	-34,81*** (6,92)	-33,08*** (7,30)
Entre 2 y 4 h ciencias	21,85*** (6,93)	17,01** (6,77)	16,28*** (5,62)	15,67*** (5,43)	16,62** (6,84)	28,53*** (9,78)	18,47** (9,31)	19,60*** (6,83)	5,15 (9,44)	19,34*** (3,49)	17,23*** (5,28)
NR_Clase apoyo col.	-3,51 (13,77)	-13,92 (15,44)	12,97 (16,14)	-6,43 (13,74)	9,77 (16,26)	-1,12 (13,74)	-20,44 (17,78)	-7,56 (18,18)	5,78 (14,44)	-1,00 (8,49)	7,22 (13,27)
R_Clase apoyo col.	-11,29 (9,02)	-10,16 (11,94)	-19,20* (10,95)	5,67 (7,28)	-2,78 (8,74)	-15,25 (14,28)	0,82 (12,63)	6,80 (13,04)	-4,65 (16,14)	4,51 (7,16)	-4,12 (8,57)
R_No clase apoyo col.	2,54 (13,21)	-20,39 (16,85)	-14,46 (14,15)	-22,55 (13,99)	3,64 (13,84)	-24,18 (18,89)	6,37 (19,04)	1,82 (16,67)	8,62 (16,52)	10,63 (10,27)	-0,97 (12,44)
Más de 4 h. ciencias	38,05*** (8,18)	39,77*** (8,13)	44,46*** (6,92)	36,98*** (5,07)	34,99*** (7,91)	35,80*** (11,41)	31,80*** (9,06)	45,36*** (8,31)	31,86*** (6,44)	44,05*** (4,79)	37,36*** (6,42)
NR_Clase apoyo col.	11,61 (20,47)	-31,69* (16,54)	-13,06 (18,81)	-0,96 (15,13)	18,37 (17,26)	11,63 (22,27)	11,21 (15,04)	-9,40 (19,50)	24,92* (13,58)	1,46 (12,26)	13,78 (11,98)
R_Clases apoyo col.	-19,23 (14,72)	-17,85 (14,10)	-52,04*** (13,63)	-3,69 (10,96)	6,19 (11,99)	-10,51 (19,33)	-7,35 (18,38)	-14,10 (14,51)	-6,35 (14,60)	-22,89* (12,74)	-1,25 (11,53)
R_No clases apoyo col.	-28,62 (23,12)	-21,12 (22,97)	-31,58 (19,54)	-39,46** (19,76)	13,55 (16,95)	-77,32 (25,39)	-11,22 (23,91)	-16,65 (36,25)	-7,01 (20,45)	-14,61 (16,45)	-20,46 (15,64)
CONSTANTE	484,05*** (4,51)	498,18*** (5,65)	509,44*** (5,05)	525,95*** (3,86)	519,65*** (5,04)	505,58*** (5,51)	527,85*** (5,64)	515,81*** (6,04)	518,11*** (4,44)	478,57*** (3,11)	487,89*** (5,08)
Nº observaciones	1.232	1.261	1.272	1.249	1.253	1.302	1.250	1.127	1.273	3.078	1.748
R2	0,48	0,45	0,43	0,50	0,45	0,38	0,45	0,46	0,47	0,40	0,41

Notas: *, ** y *** significa que el coeficiente es significativo al 10%, al 5% o al 1%, respectivamente. Los errores estándar se muestran entre paréntesis.

Tabla 8: Descomposición Oaxaca-Blinder de las diferencias regionales en resultado educativo entre las distintas regiones con Andalucía. Características e impactos en valor absoluto y en porcentaje

Grupo de Variables	Aragón				Asturias				Cantabria			
	Caract.		Impac.		Caract.		Impac.		Caract.		Impac.	
	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%
VARIABLES INDIVIDUALES	5,61	15,30	6,34	17,31	8,00	24,08	-0,52	-1,57	2,96	9,00	-9,51	-28,90
VARIABLES FAMILIARES	6,72	18,34	1,42	3,88	5,29	15,92	-4,03	-12,12	5,30	16,11	-9,94	-30,23
VARIABLES DE COLEGIO	2,09	5,70	0,33	0,91	2,08	6,25	-2,99	-9,00	2,05	6,22	0,14	0,42
CONSTANTE			14,14	38,56			25,40	76,44			41,90	127,37
Subtotal	14,42	39,33	22,24	60,67	15,37	46,25	17,86	53,75	10,31	31,34	22,59	68,66
Total Diferencias Observadas con Andalucía			36,66				33,23				32,90	
Grupo de Variables	Castilla y León				Cataluña				Galicia			
	Caract.		Impac.		Caract.		Impac.		Caract.		Impac.	
	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%
VARIABLES INDIVIDUALES	8,81	21,97	5,69	14,19	10,47	71,09	3,64	24,71	4,88	15,86	-13,52	-43,91
VARIABLES FAMILIARES	5,02	12,51	-18,17	-45,28	9,62	65,35	-34,82	-236,44	3,50	11,37	-10,72	-34,81
VARIABLES DE COLEGIO	2,09	5,21	1,07	2,66	2,61	17,69	1,67	11,34	1,82	5,90	1,04	3,39
CONSTANTE			35,61	88,75			21,54	146,26			43,80	142,19
Subtotal	15,92	39,68	24,20	60,32	22,70	154,14	-7,97	-54,14	10,20	33,13	20,60	66,87
Total Diferencias Observadas con Andalucía			40,12				14,73				30,80	
Grupo de Variables	La Rioja				Navarra				País Vasco			
	Caract.		Impac.		Caract.		Impac.		Caract.		Impac.	
	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%
VARIABLES INDIVIDUALES	5,65	12,58	-4,94	-11,00	9,88	26,11	2,04	5,39	12,80	63,35	10,62	52,58
VARIABLES FAMILIARES	6,18	13,77	-2,12	-4,72	7,69	20,31	-15,14	-40,00	8,71	43,11	-11,17	-55,27
VARIABLES DE COLEGIO	3,96	8,83	4,40	9,80	3,77	9,96	-4,45	-11,77	3,22	15,96	1,49	7,36
CONSTANTE			31,77	70,75			34,06	90,01			-5,47	-27,08
Subtotal	15,79	35,18	29,11	64,82	21,33	56,37	16,51	43,63	24,73	122,42	-4,53	-22,42
Total Diferencias Observadas con Andalucía			44,90				37,84				20,20	
Grupo de Variables	Resto de España											
	Caract.		Impac.									
	Abs.	%	Abs.	%								
VARIABLES INDIVIDUALES	3,74	33,89	-1,47	-13,29								
VARIABLES FAMILIARES	4,99	45,15	-0,29	-2,59								
VARIABLES DE COLEGIO	0,21	1,94	0,02	0,15								
CONSTANTE			3,84	34,75								
Subtotal	8,95	80,98	2,10	19,02								
Total Diferencias Observadas con Andalucía			11,05									

Apéndices metodológicos (AM)

AM1. Los Valores Plausibles y la estimación de los errores estándar

La obtención de los errores estándar de los parámetros estimados requiere un comentario detenido. La información en PISA proviene de un diseño que trata de corregir por errores de muestreo de manera que se obtienen distintos valores posibles para el rendimiento de cada alumno, los denominados *valores plausibles* en la nomenclatura de la encuesta, y además se realiza un muestreo estratificado de manera que podamos elevar a poblacional los resultados muestrales mediante los correspondientes pesos asignados a cada alumno.

La metodología que, por recomendación de la OCDE, se ha seguido en este estudio para estimar los coeficientes de las distintas variables explicativas introducidas en el modelo, consiste en estimar el modelo con cada uno de esos cinco valores plausibles. El coeficiente mostrado en las tablas de resultados será la media de los obtenidos en cada una de estas cinco estimaciones.²⁵

Con respecto a los errores estándar, se corrige la varianza de cada coeficiente por el método de muestreo de nuestra base de datos. Concretamente, la muestra obtenida en PISA se ha llevado a cabo en dos etapas. Primero se escoge a los colegios que participan de la muestra y posteriormente se selecciona un número de alumnos dentro de cada uno de los colegios. Debido a este tipo de muestreo en dos etapas, la varianza muestral de un determinado parámetro va a ser mayor que en el caso del muestreo aleatorio simple. Para estimar las varianzas en este contexto se utilizará, siguiendo de nuevo la metodología propuesta por la OCDE la técnica *Balanced Repeated Replication (BRR)*.²⁶ Esta técnica consiste en utilizar submuestras para calcular el estimador de interés de tal manera que la variabilidad que presenta este estimador de interés será la utilizada para calcular los errores típicos de cada coeficiente estimado y por tanto los intervalos de confianza. Para el uso de esta técnica hay 80 pesos replicados en la base de datos de PISA que se obtienen a partir replicaciones de la muestra. Por lo que la regresión con cada valor plausible hay que hacerla para cada una de esas replicaciones.

²⁵ La explicación del análisis con los 5 Valores Plausibles se puede ver en *PISA 2006 Technical Report. Chapter 9: Scaling PISA cognitive data. OECD (2009)*.

²⁶ Este método de estimación de la varianza puede consultarse en *PISA 2006 Technical Report. Chapter 8: Survey Weighting and the Calculation of Sampling Variance. OECD*.

AM2. Creación del índice de interés.

El análisis cluster es una herramienta que permite dividir a los individuos de una muestra en C grupos atendiendo a una serie de variables o características seleccionadas. En este trabajo se han agrupado a los alumnos en tres grupos (interés alto, medio, y bajo) atendiendo a 11 preguntas de interés del cuestionario de estudiantes²⁷.

Cada individuo es asignado a un grupo concreto porque es con cual tiene más proximidad de acuerdo con un criterio de distancia. El criterio elegido en este caso ha sido la distancia *euclídea* que se calcula de la siguiente manera:

$$d(p, q) = \sqrt{\sum_{i=1}^{11} (p_i - q_i)^2}$$

Siendo p y q dos vectores de dos estudiantes que representan los valores de 11 variables, $d(p, q)$ es la distancia *euclídea* entre los dos vectores p y q .

Hay distintos procedimientos para la formación de los grupos que llevan a distintos resultados. En este trabajo se ha utilizado el procedimiento de *conglomerado de k medias* del paquete estadístico *SPSS*, el cuál conlleva distintas fases:

Etapas 1. Se seleccionan los centros de grupo. Éstos son un vector inicial para cada uno de los 3 grupos. Se puede optar por algún criterio para elegir estos *centros de grupo* o se puede optar por seleccionar de forma aleatoria 3 observaciones de la muestra eligiendo sus vectores de valores de las 11 variables. En este caso se ha optado por el criterio por defecto del programa informático que tiene por finalidad escoger *centros de grupo* lo suficientemente distantes entre sí.

Etapas 2. Se calculan los centros de grupo de clasificación. Ahora se empieza a asignar cada individuo uno a uno al grupo más cercano. Entendiendo como grupo más cercano aquél con el cual la distancia *euclídea* entre el vector del individuo y el *centro de grupo* es la menor. Así mismo, cada vez que un individuo se asigna a un grupo el *centro de grupo* es actualizado y remplazado por la media del grupo de tal manera que cuando todos los individuos estén asignados la media actualizada de cada grupo es denominada *centro de clasificación* de los grupos y se va a tener en cuenta para la siguiente etapa.

Etapas 3. Asignación de los individuos al grupo más cercano. Ahora la distancia con respecto a un grupo es la distancia *euclídea* entre el individuo y los *centros de clasificación* de los grupos. Un individuo es asignado al grupo con el que la distancia es menor. De este último paso se obtiene la *media final* de cada grupo como la media de todos los individuos que están en un determinado grupo. Cuando el número de iteraciones es mayor que uno la *media final* de cada grupo obtenida en esta etapa 3 es fijada como *centro de clasificación* y se repite la etapa 3 de nuevo tantas veces como se indique o hasta que las *medias finales* no cambien más que un determinado valor entre una iteración y la siguiente. En este trabajo se ha realizado tan sólo una iteración.

²⁷ Estas preguntas vienen establecidas en la parte 3 del cuestionario (P16-P19, P21), en la que los alumnos daban su opinión sobre la ciencia, la parte 4 (P22), en la que se preguntaba acerca del medio ambiente, la parte 5 (P27, P29), que estaba destinada a la profesión que le gustaría desempeñar en el futuro, y la parte 7 (P35-P37), que trataba de las clases de ciencias en el colegio.

AM3. Construcción de las variables de tiempo de estudio y clases de ciencias.

- **El tiempo de estudio propio de las ciencias.** En la pregunta 31 del cuestionario individual se pregunta a los estudiantes el número de horas semanales que dedican al estudio de las ciencias o a hacer deberes de ciencias él solo. Las posibles respuestas son: nada; menos de 2 horas; más de 2 pero menos de 4 horas; más de 4 pero menos de 6 horas; y 6 o más horas. Con el objetivo de simplificar las categorías se han diferenciado dos grupos: los alumnos que estudian 2 o menos horas a la semana, y; los alumnos que estudian más de dos horas semanales.
- **La asistencia a clases de recuperación dentro y fuera del colegio.** En la pregunta 32 del cuestionario individual se pregunta a los alumnos el tipo de clases de recuperación, apoyo o particulares asisten. Los alumnos podían responder afirmativa o negativamente a las siguiente tipos de clases: a) Clases individuales con un profesor que también enseña en el colegio o instituto; b) Clases individuales con un profesor que no enseña en el colegio o instituto; c) Clases en grupos reducidos (menos de 8 alumnos) con un profesor que también enseña en el colegio o instituto; d) Clases en grupos reducidos (menos de 8 alumnos) con un profesor que no enseña en el colegio o instituto; e) Clases en grupos más numerosos (8 o más alumnos) con un profesor que también enseña en el colegio o instituto; f) Clases en grupos más numerosos (8 o más alumnos) con un profesor que no enseña en el colegio o instituto. Los alumnos que indicaron que respondieron afirmativamente a alguno de los tipos de clase a), c) o e) son alumnos que se han definido como aquéllos que tienen clases de recuperación dentro del colegio, por el contrario los alumnos que no respondieron afirmativamente ninguna de las letras indicadas son aquéllos que no tienen clases de recuperación dentro del colegio. Por otro lado, los alumnos que respondieron afirmativamente a alguno de los tipos de clase b), d) o f) se han definido como los alumnos que tienen clases particulares fuera del colegio, y el resto cómo los que no tienen clases particulares fuera del colegio. Nótese que un mismo alumno puede tener ambos tipos de clases de recuperación, dentro y fuera del colegio.
- **Número de horas semanales que el estudiante tiene de clases regulares de ciencias en el colegio.** En la pregunta 31 del cuestionario individual también se pregunta a los estudiantes el número de horas semanales que tienen de clases de ciencias regulares en el colegio. Las posibles respuestas son: nada; menos de 2 horas; más de 2 pero menos de 4 horas; más de 4 pero menos de 6 horas; y 6 o más horas. En este caso los estudiantes se han agrupado entre aquéllos que tienen 2 o menos horas de ciencias a la semana, más de 2 pero menos de 4 horas, y más de 4 horas semanales.