

Diferencias regionales y rendimiento educativo

Análisis de los resultados
del estudio PISA-2006

J. Ignacio García Pérez
J. Luis Pinto Prades
J. Antonio Robles Zurita
Centro de Estudios Andaluces



Diferencias regionales y rendimiento educativo

Análisis de los resultados
del estudio PISA-2006

J. Ignacio García Pérez
J. Luis Pinto Prades
J. Antonio Robles Zurita
Centro de Estudios Andaluces
IF011/09



Centro de Estudios Andaluces
CONSEJERÍA DE LA PRESIDENCIA



Edita:
Fundación Centro de Estudios Andaluces,
Consejería de la Presidencia, JUNTA DE ANDALUCÍA.

© Del texto, sus autores.
© Fundación Centro de Estudios Andaluces
Bailén, 50, 41001 – Sevilla
Tel.: 955 055 210
Fax: 955 055 211
www.centrodeestudiosandaluces.es

Octubre de 2009
ISBN: 978-84-692-6702-8

Índice

1. Introducción y motivación	7
2. El estudio PISA	13
2.1. Información recogida	13
2.2. Características de las muestras de los países	15
2.3. La muestra española	15
3. Metodología	19
3.1. Variables utilizadas	19
3.1.1. Dotaciones iniciales	20
3.1.2. Variables de esfuerzo individual.....	21
3.1.3. Características de los colegios	22
3.2. Metodología de análisis.....	25
3.2.1. Modelos a estimar	26
Modelo para el total de España	26
Modelo para cada región	26
3.2.2. Estimación econométrica. Coeficientes y errores estándar.....	27
3.2.3. Análisis <i>Shift-Share</i>	28
4. Análisis bivalente	33
4.1. Variables relacionadas con el resultado educativo	33
4.2. Características de las Comunidades Autónomas	37
5. Modelo para el total de España.....	41
5.1. Especificación del modelo	42
5.2. Resultados del modelo estimado	42
6. Modelos regionales	51
6.1. Resultados de la estimación	51
6.2. Diferencias en resultados para distintos grupos de estudiantes.....	54
7. Análisis <i>Shift-Share</i>.....	59
8. Conclusiones.....	63
Bibliografía	67
Apéndices	71

1. Introducción y motivación



1. Introducción y motivación

En Diciembre de 2007 se publicó por la OCDE (Organización para la Cooperación y el Desarrollo Económico) la base de datos de PISA 2006 (*Programme for International Student Assessment*). A través de ésta se ha estudiado en profundidad el resultado académico de los estudiantes de 15 años de diversos países (cincuenta y siete) y regiones participantes. Los ficheros con la información individual de los participantes en este proceso de evaluación suponen una fuente de recursos de enorme riqueza para la investigación sobre educación.

El objetivo de este trabajo es tratar de explicar las diferencias de nivel educativo detectadas entre la Comunidad Autónoma de Andalucía y el resto de regiones de España participantes en el programa. La OCDE estimó el tamaño muestral para comparar países y no regiones, lo cual haría difícil comparar regiones españolas entre sí, si no fuera porque en nuestro país algunas Comunidades¹ ampliaron la muestra de forma voluntaria para tener datos representativos de su región. Eso permitirá comparar Andalucía con las Comunidades Autónomas (CCAA) que tienen una muestra ampliada. El

resto de CCAA sin muestra ampliada formarán un grupo denominado en lo que sigue como “Resto de España”. A la hora de comparar el resultado educativo en Andalucía con otras zonas, se ha optado por restringir dicha comparación a otras regiones de España. La razón de excluir a otros países radica en que, si ya es complejo comparar regiones entre sí, es mucho más complejo extraer conclusiones de dichas comparaciones cuando las regiones a comparar se encuentran en diferentes contextos culturales o sociales y los sistemas educativos de éstas tienen también estructuras muy distintas. Al reducir la comparación a las regiones españolas se reduce el riesgo de que existan variables inobservables que estén influyendo en los resultados y que no sepamos identificar. Asimismo, si se compara Andalucía con otra región o país muy distinto, es más difícil sacar conclusiones de política educativa de dichas comparaciones mientras que este ejercicio es más directo cuando se compara con una región más similar.

La información recogida en PISA y difundida por distintos medios de comunicación refleja que la Comunidad Autónoma de Andalucía queda claramente por debajo del resto de regiones españolas comparables. Se han dado diversas explicaciones de estos resultados por parte de la administración pública y por el sector académico. La Consejería de Educación de la Junta de Andalucía

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

publicó en diciembre de 2007 su propio informe en el que se hace referencia al nivel cultural y socioeconómico de la Comunidad Autónoma como razón principal de su bajo resultado. Sin embargo, este estudio tiene varias limitaciones que intentaremos superar en este documento. Por un lado, algunas comparaciones se realizan utilizando las medias entre regiones, perdiendo gran parte de la riqueza de una base de datos con miles de observaciones individuales. Por otro lado, se realizan comparaciones bivariadas, es decir, comparaciones del valor obtenido en PISA controlando solo por una variable explicativa cuando, debido a que hay interrelaciones entre las variables que pueden estar influyendo en dicho valor, es más correcto realizar un análisis multivariante, que refleje los efectos aislados de cada variable, sin que el resto de posibles influencias contamine dicha comparación. Así, se podrán predecir mejor los resultados esperados.

Desde el sector académico ha habido ya otras publicaciones que han utilizado los datos de PISA de España para tratar las diferencias regionales. Carabaña (2008) viene a destacar también la importancia del estatus socioeconómico en el resultado académico medio de las regiones y países. Ciccone y García-Fontes (2008) hacen un análisis de la importancia de la educación de los padres y el estatus inmigrante en las comparaciones internacionales y regionales. Finalmente, Cordero *et al.* (2009) hacen un análisis de las diferencias entre Comunidades Autónomas aplicando un Análisis Envolvente de Datos.

El hecho de que Andalucía obtenga una menor nota media se puede deber a dos posibles efectos, normalmente presentes de manera conjunta. En primer lugar, se puede deber a que en esta Comunidad Autónoma exista

un mayor porcentaje de alumnos con características que normalmente están relacionadas con un bajo resultado educativo (padres de bajo nivel educativo, bajo estatus ocupacional o bajo nivel cultural de la familia), esto es lo que se llama en este trabajo como *diferencias en características*. Por otro lado, las diferencias regionales pueden deberse también a que incluso a igualdad de características los alumnos andaluces obtienen peor resultado educativo, lo que se denomina como *diferencias en rendimientos*. En el presente trabajo se compara el comportamiento relativo de Andalucía intentando estimar los dos componentes anteriormente citados respecto a otras zonas de España aplicando dos enfoques diferentes que se llevan a cabo mediante la estimación de dos modelos econométricos complementarios: uno estimado para el total de España, *Modelo España*; y otro estimado para cada una de las CCAA, *Modelo Regional*. Estos dos modelos econométricos tratan al resultado obtenido en PISA por los alumnos como variable dependiente y a las variables familiares, personales y de colegio como explicativas o variables independientes.

Con la estimación del Modelo España se puede obtener cual es el efecto en el resultado educativo de pertenecer a una determinada región a igualdad en el resto de factores personales, familiares y de colegio. Los resultados obtenidos indican claramente que las variables familiares socioeconómicas y culturales explican gran parte del peor comportamiento de los alumnos andaluces. Así, cuando se comparan alumnos con entornos familiares similares, Andalucía no está peor que otras regiones de España, tales como Cataluña y el País Vasco, lo que quiere decir que las diferencias en características son importantes. Sin embargo, sigue estando peor que otras zonas, como Castilla y León y La Rioja, aunque la diferencia se reduce considerablemente cuan-

do se consideran variables que reflejan tanto el esfuerzo individual del alumno como el nivel de implicación de la familia en conseguir un buen ambiente de estudio. Es decir, aún a igualdad de características un alumno de Andalucía obtiene peores notas que los alumnos de algunas otras regiones.

La parte de las diferencias debidas a rendimientos se ha calculado bajo el supuesto de que los coeficientes de las distintas variables explicativas son los mismos en todas las regiones. Para relajar este supuesto, estimamos el *Modelo Regional* con el que se pretende estimar el valor de las diferencias en rendimientos sin imponer restricciones sobre los coeficientes del modelo para cada región. Se estima un modelo distinto para cada comunidad autónoma de tal manera que tendremos unos coeficientes distintos para cada una de ellas. Así podremos comprobar las diferencias regionales en los efectos que tienen las distintas variables explicativas sobre el resultado educativo. El objetivo último de este apartado del trabajo es estimar qué importancia tiene la composición sociodemográfica de Andalucía en las diferencias observadas. Por tanto, partiendo del modelo regional se descompondrán las diferencias regionales totales en, por un lado, *diferencias en características* y, por otro, *diferencias en rendimientos*, gracias a la aplicación de un análisis *Shift-Share*.

La estructura del estudio es la siguiente. En primer lugar, se explicará brevemente en qué consiste la base de datos utilizada. Únicamente se mencionarán aquellos detalles que más afectan a este estudio. 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. En la sección 3 se explica la metodología utilizada en el estudio, modelos a estimar, estrategia econométrica y la descomposición de las diferencias regionales a través del análisis *shift-share*. En el apartado 4 se lleva a cabo un análisis previo de carácter descriptivo en base a comparaciones bivariantes, que van a permitir observar las variables que más relación tienen con la nota de PISA y las características de cada Comunidad Autónoma en relación a éstas variables. Posteriormente, los resultados de la estimación de los modelos, *Modelo España* y *Modelo Regional*, se muestran en las secciones 5 y 6 respectivamente. La contabilidad de las diferencias regionales a través del análisis *Shift-Share* es presentada en el apartado 7, en el que se lleva a cabo la descomposición de las diferencias regionales en las características y los rendimientos. Por último, se llegará a una conclusión en la que se comentarán las implicaciones de los resultados de todo el análisis realizado.

2. El estudio PISA



2. El estudio PISA

El estudio de PISA-2006, cuyos datos se utilizan para este trabajo, se engloba dentro de un programa de evaluación de la educación más amplio que empezó en 2000 y que tiene previsto acabar en 2015. A lo largo del citado periodo, cada tres años se mide el resultado de los estudiantes en los países miembros de la OCDE y en países llamados asociados que se quieren unir al programa. A continuación se presentan varios aspectos que caracterizan PISA.

2.1. Información recogida

La información contenida en las bases de datos tiene diversos orígenes:

- Los exámenes de evaluación.
- Un cuestionario de los alumnos.
- Un cuestionario de los colegios.
- Un cuestionario de los padres.

Los exámenes de evaluación, de dos horas de duración, han servido para medir los conocimientos de los alumnos. La evaluación se centra en el resultado en las

áreas de matemáticas, ciencias y lectura. Cada edición del estudio pone mayor énfasis en alguna de estas tres áreas del conocimiento. Así, PISA 2000 estuvo destinada a hacer una evaluación más profunda en lectura, PISA 2003 se centró en matemáticas, y en la edición de 2006 le ha tocado el turno a las ciencias.

El contenido de las preguntas que tenían que resolver los alumnos era un tanto distinto al de muchos exámenes de conocimientos habituales en un curso académico. Los responsables de PISA hicieron un esfuerzo especial por elaborar un test que no evaluase únicamente el conocimiento curricular adquirido en el centro educativo, sino la capacidad del alumno para utilizar los conocimientos adquiridos de las distintas disciplinas en la resolución de situaciones que se pueden presentar en la vida de una persona. Por ejemplo, en el caso de la evaluación de lectura se pretendía evaluar si comprendían la aplicación de los textos a problemas concretos. Los alumnos se enfrentaron tanto a textos escritos continuos como discontinuos tales como diagramas, tablas o gráficos. En el caso de las matemáticas, los alumnos debían emitir juicios basados en el razonamiento matemático. Por lo que respecta a la evaluación de ciencias, las preguntas se clasificaron según los siguientes criterios:

- *Competencias del estudiante:* identificación de cuestiones científicas, explicación de fenómenos científicos, y utilización de pruebas científicas.
- *Subárea del conocimiento:* Sistemas físicos, vivos, terrestres y espaciales y tecnológicos.

Dentro de la parte de evaluación de ciencias se incluyeron preguntas que intentaban medir el nivel de comprensión de las ciencias como método para adquirir conocimiento. También se incluyeron cuestiones relativas al interés de los alumnos por las ciencias y al apoyo a la investigación científica.

Los cuestionarios restantes de PISA tratan de recoger información, que se considera relacionada con el resultado educativo, y que es facilitada por diversos agentes como los propios estudiantes, los directores del centro al que éstos asisten, y los padres de los mismos.

El cuestionario que contestaron los estudiantes participantes de la muestra contenía preguntas sobre:

- Entorno familiar;
- Su opinión sobre distintos aspectos relacionados con las ciencias;
- Información y la actitud hacia el medio ambiente;
- Expectativas profesionales y su vinculación con las ciencias.
- El tiempo dedicado al estudio;
- Las clases de ciencias en el colegio;
- Utilización de ordenadores;

Los directores de los colegios respondieron a preguntas sobre:

- Las características del centro;
- El alumnado;
- Los recursos del centro;
- El profesorado;
- La organización del centro;
- La inclusión del medio ambiente en el currículo del centro.
- La orientación profesional y la preparación de los alumnos para estudios posteriores.

El cuestionario de padres tan sólo se realizó en 16 países, entre los que no se incluyó España. Este cuestionario recogía datos sobre:

- Las actividades pasadas sobre ciencias de los estudiantes.
- La visión de los padres sobre el colegio del alumno.
- La opinión del padre acerca de la profesión deseada por su hijo y su relación con las ciencias. La necesidad de tener capacidades científicas para el mercado laboral.
- La opinión de los padres sobre las ciencias y el medio ambiente.
- El gasto familiar en educación.
- La educación y la ocupación de los padres.

2.2. Características de las muestras de los países

En la edición de 2006 participaron 57 países, 30 de ellos miembros de la OCDE. El número de individuos que forman las muestras de los distintos países varía desde los 339 de Liechtenstein hasta los 30.971 de México. El tamaño de la muestra depende no sólo del tamaño de la población objeto de estudio de cada país sino del hecho de que dentro de cada país se quieran obtener muestras representativas para algunas regiones (en cuyo caso la muestra va a ser mayor).

La población objeto de estudio en cada país, es la comprendida por los alumnos de 15 años que se encuentran escolarizados y que han terminado la educación obligatoria básica. Debido a esta definición de población los alumnos pueden estar matriculados en distintos cursos y no necesariamente al que les corresponde de acuerdo a su edad.² Se excluyó a estudiantes con discapacidad intelectual o funcional, así como aquéllos que presentaban una incapacidad en el dominio del idioma de evaluación. Se excluyeron de la población objeto de estudio algunos colegios por tener unas características especiales establecidas por PISA (colegios geográficamente inaccesibles; o colegios en los que la totalidad de los alumnos se encuentran en alguna de las causas de exclusión).

2.3. La muestra española

En este estudio que se presenta se ha utilizado la parte de la base de datos de PISA, que corresponde a España y sus regiones. Los países pueden haber participado teniendo muestras regionales representativas como es el caso de España. En este caso se puede realizar el análisis estadístico para la totalidad del territorio nacional y para aquellas regiones que tienen muestra ampliada. Una región tiene muestra representativa si la forman alrededor de 1.500 alumnos. Las regiones con muestra representativa vienen recogidas en la Tabla A1. Como se puede ver la muestra para el total de España (donde entran todas las regiones) alcanza los 19.604 estudiantes. En cuanto a las muestras regionales observamos que la mayoría tienen aproximadamente los 1.500 estudiantes necesarios aunque hay que destacar el caso del País Vasco, región para la cual la muestra asciende a casi 3.929 alumnos. Se establecen las distintas regiones que no tienen muestra ampliada, con la denominación de “Resto de España”.

2 Para España la muestra la componen alumnos de están cursando desde 1º de la ESO hasta 1º de Bachillerato.

3. Metodología



3. Metodología

3.1. Variables utilizadas

La principal variable a utilizar será la nota del alumno/a en cada una de las tres materias evaluadas. Como variables explicativas de la nota, se utilizan un total de 25 variables que corresponden a 29 preguntas del cuestionario de alumnos y a 16 preguntas del cuestionario de colegios.

La variable a estudiar, o variable dependiente, es concretamente el **resultado educativo (puntuación) obtenido en los exámenes de ciencias, matemáticas y lectura** por cada alumno. Se tiene por tanto tres variables dependientes, una por cada disciplina.

Un elemento un tanto característico del estudio PISA (y que genera no poca confusión) es el hecho de que no se utiliza una sola medida del resultado de un alumno porque se considera por la propia OCDE que puede haber errores de medición. Esto da lugar a que cada alumno, para cada una de las tres áreas evaluadas, disponga de lo que se denomina 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 test. En definitiva PISA no estima una puntuación de cada alumno sino una distribución de probabilidad de la

puntuación de cada alumno, tratando a ésta como una variable aleatoria. Las diferencias en la puntuación se consideran estadísticamente significativas si son de 5 puntos o más.³

En este trabajo, se clasifican las variables explicativas en los siguientes grupos:

i. **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 en el resultado educativo entre regiones que no pueden ser atribuidas a un distinto funcionamiento de los sistemas educativos. Además, estas variables tienen un rasgo significativo como variables explicativas, ya que se pueden considerar como totalmente exógenas, esto es, influyen en el resultado educativo del alumno pero no ocurre lo contrario, el logro educativo de un alumno no hace cambiar las características de su familia. A modo de

3 En la Sección 3.2.2. se explica detenidamente como se ha procedido para estimar nuestros modelos usando esta distribución de probabilidad basada en los cinco valores plausibles para cada individuo.

ejemplo, el hecho de que un alumno tenga padres con nivel educativo y ocupacional alto puede influir en el resultado educativo, pero no se espera que la causalidad se de en el sentido inverso. Por ello, las dotaciones iniciales son las variables que con más seguridad podemos tratar como exógenas.

ii. Variables de esfuerzo individual. Son aquéllas variables que presenta el alumno y que se basan en decisiones individuales del estudiante (por ejemplo, la actitud ante la ciencia). Estas variables, como se explica seguidamente, pueden ser problemáticas de clasificar como variables exógenas (como en el caso de las dotaciones iniciales) o endógenas al sistema.

iii. Variables de colegio. Son aquéllas que presenta un colegio y que serán las mismas para todos los alumnos del mismo, como la titularidad del colegio o el número de profesores.

A continuación, se ven por separado cada conjunto de variables explicativas:

3.1.1. Dotaciones iniciales

El concepto de “dotaciones iniciales” se corresponde con las siguientes características socioeconómicas de la familia del alumno:

A. Nivel educativo de los padres. En este trabajo se han hecho cuatro grupos: **sin estudios; primaria; secundaria y FP** (EGB, FP 1^{er} grado, BUP, COU y FP media, y FP superior); y **Titulación universitaria**. Se ha considerado por separado el nivel educativo del padre y de la madre, así como la presencia en el hogar de padres del mismo o de distinto nivel educativo.

B. Ocupación de los padres. A partir de lo indicado por los estudiantes, se tipificaron las ocupaciones de los padres de acuerdo a la Clasificación Internacional Uniforme de Ocupaciones (ISCO, *International Standard Classification of Occupations*,⁴ en sus siglas inglesas). El informe PISA agrupa estas categorías entre trabajadores manuales (o de cuello azul) y los no manuales (o de cuello blanco), clasificación que seguiremos en nuestro estudio. Para las mujeres se incluye el grupo de **amas de casa** diferenciado de los dos grupos antes descritos con el fin de identificar las no participantes en el mercado de trabajo. Esta categoría en el caso de los padres ha sido eliminada por falta de observaciones.

4 Se trata de una clasificación de ocupaciones realizada por la Organización Internacional del Trabajo. Ésta pertenece a la Familia Internacional de Clasificaciones Económicas y Sociales que comprende clasificaciones en diversos ámbitos que son registradas en el Inventario de Clasificaciones de las Naciones Unidas. Las categorías en las que se pueden agrupar las ocupaciones del ISCO son:

1. Dirección de empresas y de las Administraciones Públicas.
2. Técnicos y profesionales científicos e intelectuales.
3. Técnicos y profesionales de apoyo.
4. Empleados de tipo administrativo.
5. Trabajadores de los servicios de restauración, personales, protección y vendedores de los comercios.
6. Trabajadores cualificados en la agricultura y en la pesca.
7. Artesanos y trabajadores cualificados de las industrias manufactureras, la construcción, y la minería, excepto los operadores de instalaciones y maquinaria.
8. Operadores de instalaciones y maquinaria, y montadores.
9. Trabajadores no cualificados.
10. Ama de casa.
11. Estudiante.
12. Beneficiario de la SS.

C. Nivel económico o riqueza. En PISA no existe ninguna variable que mida directamente el **nivel económico** o de **riqueza** de las familias. Sin embargo, hay algunas variables que pueden estar relacionadas con ésta. En el cuestionario de los estudiantes había una pregunta en la que los estudiantes indicaban si la familia disponía o no de ciertos bienes.⁵ Dado que algunos de estos bienes son poseídos por prácticamente la totalidad de la muestra estudiada, se han utilizado cuatro tipos de posesiones (lavavajillas; videocámara; televisión satélite; y equipo de cine en casa) para crear un **índice de posesiones familiares de riqueza** que va de 0 (el alumno que no posea ninguno) a 4 (el alumno que posea 4).

D. Posesiones educativas. Dado que es lógico suponer que el entorno educativo y cultural de la familia puede influir en el resultado académico del alumno, se elaboró un índice cultural-educativo a partir de tres cuestiones. En primer lugar, la pregunta de si se disponía en casa de⁶: software educativo; conexión a Internet; libros de literatura; libros de poesía; y obras de arte. En segundo lugar, se utilizó el **número de libros** en casa y, en tercer lugar, el **número de or-**

denadores que poseía la familia. El índice elaborado clasifica a las familias en tres categorías, esto es, cultura educativa baja, media y alta.

E. Otras variables. También se tuvieron en cuenta otras características de los alumnos que podían, en principio, influir en el resultado educativo manifestado con la nota en PISA. Fueron las siguientes: a) nacionalidad; b) sexo; c) lengua en que normalmente habla en casa y d) mes de nacimiento. Otros estudios han tenido en cuenta algunas de estas variables para el caso español: Calero y Escardíbul (2007) utilizan la edad, el género y la nacionalidad como variables explicativas. También Zinovyeva *et al* (2009) utilizan la base de datos de PISA y ponen de manifiesto el peor resultado educativo para los inmigrantes aún a igualdad de características socioeconómicas de las familias.

3.1.2. Variables de esfuerzo individual

En PISA 2006 existe una novedad con respecto a las anteriores ediciones. Se ha querido medir una serie de **variables de actitud** de los alumnos hacia las ciencias, tanto en la propia evaluación como en el propio cuestionario que completó cada estudiante. Estas preguntas vienen establecidas en la parte 3 del cuestionario, en la que los alumnos daban su opinión sobre la ciencia, la parte 4, en la que se preguntaba acerca del medio ambiente, la parte 5, que estaba relacionada con la profesión que le gustaría desempeñar en el futuro, y la parte 7, que trataba de las clases de ciencias en el colegio. Para cada pregunta (un total de once) se ha creado un índice de actitud. 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 (o muy buena), media y baja.

5 Se preguntó a los alumnos por la posesión de: a) Una mesa para el estudio; b) Una habitación para el alumno; c) Un sitio tranquilo para estudiar; d) Un ordenador que puedas utilizar para estudiar; e) Programas educativos de ordenador; f) Conexión a Internet; g) Su propia calculadora; h) Libros de literatura clásica (p. ej., Cervantes, Galdós, García Lorca, etc.); i) Libros de poesía; j) Obras de arte (p. ej., cuadros, esculturas, etc.); k) Libros para ayudarte con tus estudios; l) Un diccionario; m) Un lavavajillas o lavaplatos; n) DVD o vídeo; o) Cámara de vídeo; p) TV satélite; q) Home cinema.

6 Igualmente en este caso hemos descartado los artículos que aunque indiquen entorno cultural no muestran variación muestral suficiente, en el sentido de que su posesión se extiende a la práctica totalidad de la población de la muestra.

Diversos estudios dentro de la literatura sobre economía de la educación apuntan a la importancia de esta variable, en concreto De Fraja et al (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.

Otra variable incluida en este grupo es si el alumno está o no **repitiendo curso**. Esto se puede saber fácilmente en la base de datos porque se conoce el curso en que se encuentran matriculados los estudiantes. Así en la base de datos nos encontramos con alumnos que están en 1º de Bachillerato, 4º, 3º, 2º y 1º de ESO. Por tanto, los alumnos que han repetido un curso son los de 3º de ESO y los alumnos que han repetido dos o tres cursos son los de 2º y 1º de ESO respectivamente.

Estas dos variables se han separado de las variables individuales que hemos denominado dotaciones iniciales por su potencial carácter endógeno al propio resultado del alumno. Dada la importancia de este concepto, lo explicaremos brevemente. Este estudio intenta explicar las diferencias en el resultado educativo en distintas regiones de España. Como se ha comentado anteriormente, dado que se sabe que (por ejemplo) el nivel educativo de los padres influye en las notas de los hijos, es de esperar que en aquella región donde el nivel educativo medio de los padres sea mayor, la nota media también sea mayor. Por tanto, esa diferencia de notas no indica (obligatoriamente) que el sistema educativo en la zona donde los estudiantes sacan peores notas sea peor. Son variables exógenas (se determinan previamente) al

propio resultado del alumno. Ahora bien, si en una zona hay más o menos alumnos/as con actitud alta o baja y eso influye en la nota media no está claro si la actitud es una variable exógena al sistema educativo o si es el funcionamiento del propio sistema lo que genera la distinta actitud de los alumnos. En el caso del número de repetidores, parece que estamos, claramente, ante un caso de endogeneidad claro. Esto es, si los repetidores sacan, por término medio, peores notas que los no repetidores, el sistema educativo no puede ampararse en el hecho de que tiene más repetidores para justificar que en su zona la nota media es menor. En este estudio, siendo conscientes de este caso, utilizaremos la variable “repetidor” como una variable que explica la nota media. Esto quiere decir que este estudio únicamente tiene como objetivo explicar diferencias regionales en la nota media del informe PISA. Por tanto, el número de repetidores se considera como una variable explicativa y exógena. Esto implica que este estudio no realiza una evaluación global del resultado de los sistemas educativos regionales. Dicha evaluación global requeriría usar la variable “número de repetidores” como una variable de resultado y no como una variable explicativa. No obstante, también haremos una estimación alternativa separando los alumnos que repiten de los que no lo hacen para estudiar como cambian las diferencias regionales medias predichas por nuestro modelo cuando se permite que el efecto de repetir curso pueda afectar tanto a la nota media como al efecto del resto de variables explicativas del modelo.

3.1.3. Características de los colegios

Los directores de los colegios contestaron a un cuestionario relativo a las características de sus centros que pueden estar relacionadas con el resultado educativo

de sus alumnos. Las variables de este cuestionario que se han analizado en este estudio son:

- a. El **porcentaje de chicas o chicos** que hay en el colegio. Así para el análisis bivalente se ha utilizado una variable que diferencia entre los colegios en los que hay más del 50% de chicas y el resto de colegios.
- b. **Propiedad de los colegios.** En la base de datos original existe una variable que viene a diferenciar tres tipos de colegios. Para ello la OCDE tiene en cuenta dos criterios: la propiedad indicada por el director del centro y la financiación pública del colegio. En un primer momento se clasifica a los colegios en públicos o privados según lo establecido por el director en el cuestionario. A continuación, se vuelve a clasificar los colegios privados dependiendo de la financiación pública que también indicó el director. Así los privados se clasifican en: independientes, o *colegios privados puros*, que tienen un porcentaje de financiación pública inferior al 50%; y dependientes del Estado, *colegios privados concertados*, cuando dicho porcentaje es superior o igual al 50%. Por tanto en la variable **propiedad del colegio** se ha diferenciado entre colegios públicos, colegios privados y colegios concertados. No obstante, se observa que dentro de los colegios privados puede haber grandes diferencias en el **porcentaje de financiación pública**, al igual que ocurre para los colegios concertados. Por ello, se han tenido en cuenta dos variables que recogen el porcentaje de financiación pública de los colegios privados y concertados respectivamente. Se han clasificado los dos tipos de colegios en función del porcentaje de financiación pública. Concretamente, se dividieron los colegios entre los que tenían más o menos financiación

pública que el colegio mediano. La evidencia empírica acerca de la influencia de la propiedad de los colegios en el resultado educativo es algo controvertida. Hay estudios que se han hecho utilizando datos de otros países en los que no se encuentran relación entre la propiedad de los colegios y el logro educativo, como por ejemplo Noell (1982) y, más recientemente, Smith y Taylor (2005), para Estados Unidos y Reino Unido respectivamente. Sin embargo, otros trabajos apuntan hacia una influencia positiva de la asistencia a centros educativos privados: Coleman y Hoffer (1987), Hanushek (1986) y recientemente Opdenakker y Van Damme (2006) para el caso de EEUU. Con los datos de PISA se encuentran los trabajos de Fertig (2003) y Aburra (2005), relativos a Alemania e Italia respectivamente, sin encontrar efecto en 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 diferentes a la titularidad del centro.

- c. El **porcentaje de repetidores** que tuvo el colegio en el último año académico en ESO. Los colegios se dividen entre aquéllos cuyo porcentaje de repetidores es mayor o menor del 15%.
- d. **Nivel educativo medio de los padres de los compañeros de clase.** 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. 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 et al (1978) que encuentran un efecto positivo de las características de los compañeros de clase, al igual que Feinstein y Symons (1999).

e. Tamaño del municipio en el que se encuentra el colegio. En la base de datos nos encontramos que se diferencia entre: Pueblo pequeño (municipio de menos de 3.000 habitantes); Pueblo mediano (municipio de 3.000 a 15.000 habitantes); Ciudad pequeña (de 15.000 a 100.000 habitantes); Ciudad grande (de 100.000 a 1.000.000 habitantes) y Gran Ciudad (con más de 1.000.000 habitantes). En nuestro estudio hemos utilizado una clasificación binaria, que diferencia a las ciudades según tengan menos o más de 100.000 habitantes.

f. Existencia de una política de agrupación de los alumnos por capacidades dentro de la misma clase. Los directores informaron sobre como se agrupaban a los alumnos para las distintas asignaturas.

g. Número de profesores de los que dispone el centro. Así mismo, se pueden distinguir los profesores de tiempo completo de los de tiempo parcial. Se ha considerado que un profesor a tiempo parcial equivale a medio profesor a tiempo completo. Teniendo en cuenta dicha regla se ha dividido el número total de alumnos por el número de profesores, creando así la variable **número de alumnos por profesor** que hay en el centro. Ésta variable se ha estructurado como categórica y ordinal distinguiendo tres grupos, cada

uno de los cuáles engloban aproximadamente a 1/3 de la muestra de alumnos.

h. Información que ofrece el centro acerca de los resultados académicos así como el **uso de dicha información**. En las preguntas 15 y 17 el director del centro respondió acerca de la información que se presta a los padres de los alumnos y del uso de dicha información en la evaluación de la gestión y del personal. El responsable indicó si se informa a los padres de los resultados de los alumnos: a) en relación a alumnos de su mismo curso, b) en relación a indicadores nacionales, o c) con alumnos de otros centros en el mismo curso. Igualmente indicó si los datos de resultado se usan para evaluar al director, el profesorado, la asignación de recursos educativos, y si se efectúa un seguimiento de los datos de resultado a lo largo del tiempo. En base a esta información se ha construido un índice, a partir de un análisis cluster, que distingue entre colegios que tienden a informar del resultado académico y a usar dicha información en su gestión y toma de decisiones y aquéllos que no. Esta variable es también utilizada por Sprietma (2006) en un estudio en el que utiliza los datos de PISA 2003 para realizar comparaciones internacionales, encontrando que un mayor porcentaje de colegios que ofrecen información y comparan sus resultados a nivel regional o nacional viene acompañado de mejores resultados en los exámenes.

i. Presión que ejercen los padres. Se ha 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.

j. Nivel de competencia que existe entre el centro escolar con las escuelas de su entorno. La variable utilizada distingue entre los colegios en los que existen dos o más colegios compitiendo en la zona y aquéllos en los que la competencia se resume a un colegio más o ninguno.

k. Porcentaje de inmigrantes en el colegio.

l. Criterios para la selección de alumnos. Se ha agrupado a los alumnos en dos categorías, distinguiendo los centros que tienen un alto nivel de selección de alumnos y otros que eligen en función de variables tales como la proximidad al colegio, hermanos o similares.

3.2. Metodología de análisis

Antes de comenzar con el análisis multivariante se ha llevado a cabo un análisis descriptivo bivalente que permite obtener la primera toma de contacto con los datos. Esto facilita una visión sobre los factores más importantes que se relacionan con el resultado educativo de los alumnos. Además, se hace una presentación de las características que presentan las distintas CCAA con respecto a estos factores, lo cual facilita la búsqueda de posibles explicaciones a las diferencias existentes en los resultados de los sistemas educativos autonómicos.

Para analizar los microdatos disponibles se ha utilizado una metodología que se instrumenta para estimar de forma econométrica el efecto que tienen las distintas variables explicativas en el resultado educativo. Una vez

estimados estos efectos se puede saber si las características de Andalucía en relación a cada una de dichas variables explican las diferencias con las otras comunidades autónomas o no. Se han estimado dos modelos econométricos distintos que responden a dos enfoques diferentes a la hora de encontrar explicaciones a las diferencias regionales. En primer lugar, se estima un modelo para el total de España, *Modelo España*. Éste incluye unas variables binarias, llamadas *variables de región*, que indican la Comunidad Autónoma de pertenencia de los estudiantes, además de incluir las variables explicativas anteriormente descritas. La región omitida en este grupo de variables de región es Andalucía, que por tanto queda recogida en el término independiente de este modelo. Así, los coeficientes de las variables de región indican las diferencias en resultado educativo de Andalucía con el resto de regiones. Concretamente, dado que controlamos por las características anteriormente descritas, el coeficiente de estas variables de región miden las diferencias regionales que no se logran explicar con el resto de variables que se incluyen en el modelo: *dotaciones iniciales*, *variables de esfuerzo individual*, y *variables de colegio*. Adicionalmente, se ha estimado un segundo modelo para cada una de las regiones por separado, *Modelo Regional*. Esto ha permitido estimar efectos distintos de los factores educativos en cada una de las regiones. Además, a partir de este segundo modelo se ha procedido a la descomposición de las diferencias entre Andalucía y el resto de regiones, distinguiendo los efectos de las *características* y los *rendimientos* a partir del análisis *Shift-Share*.

3.2.1. Modelos a estimar

Modelo para el total de España

El modelo que estimaremos supone que la nota en PISA y_{ij} de un determinado alumno j que asiste a un colegio concreto i , depende de: una constante μ ; n variables individuales, $x_{h,ij}$, cuyos valores pueden variar entre alumnos de un mismo colegio; de m variables relativas al colegio asistido por el estudiante, $z_{l,i}$, cuyos valores serán los mismos para los alumnos que asisten a un mismo colegio; de r variables de región de pertenencia, $R_{k,i}$; y de una serie de variables inobservables relativas al colegio asistido y que se designan conjuntamente por el parámetro α_i . Estas variables inobservables, al ser características de los colegios, serán las mismas para todos los estudiantes que asisten al centro educativo. El modelo se puede representar por la siguiente expresión:

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

En línea con el objetivo de este trabajo, se pretende contrastar si las variables de región son significativas lo que va a suponer que, aún controlando por el resto de variables, el hecho de pertenecer a una Comunidad Autónoma u otra influye en la nota obtenida por el alumno en PISA. Si el coeficiente ϖ_k es positivo (negativo), esto indicaría que, a igualdad de características, los alumnos de la región k obtienen mejor (peor) resultado educativo que los alumnos de Andalucía. En cambio, si dicho coeficiente fuese cercano a cero se podría decir que las diferencias que a priori se encuentran entre los distintos *outputs* de los sistemas educativos autonómicos se explican en su totalidad por las variables incluidas en el modelo.

No obstante, el análisis que se pueda llevar a cabo con este modelo está suponiendo que los coeficientes de las variables explicativas son los mismos para cada Comunidad Autónoma. Puede que el efecto de una variable concreta como la ocupación de los padres influya de manera diferente en distintas regiones. Para captar estas diferencias en coeficientes entre regiones sería necesario estimar el modelo regional que se expone a continuación.

Modelo para cada región

En este caso, el modelo tendría la misma especificación que el anterior salvo porque los coeficientes de las variables individuales y de colegio serán específicos para cada región. Concretamente el modelo tendrá la siguiente especificación:

$$y_{ij} = \mu^k + \sum_{h=1}^n \beta_h^k x_{h,ij} + \sum_{l=1}^m \gamma_l^k z_{l,i} + \alpha_i^k + u_{ij} \quad (2)$$

En esta expresión se incluye un superíndice k , que identifica a la región, para cada uno de los parámetros a estimar: μ^k , β_h^k , γ_l^k y α_i^k . Además, tal y como se observa, la expresión (2) es igual a la (1) pero sin variables de región, que dejan de tener sentido en un modelo para cada sistema educativo autonómico.

Este modelo regional es el modelo que se va a tener en cuenta a la hora de llevar a cabo el Análisis *Shift-Share* que se explica posteriormente.

3.2.2. Estimación econométrica. Coeficientes y errores estándar⁷

La metodología aplicada en el análisis de regresión multivariante realizado tiene como objetivo evitar los posibles problemas de sesgo en la estimación de los coeficientes del modelo y de sus errores estándar.

La estrategia econométrica seguida, para conseguir la estimación consistente de los coeficientes de las variables incluidas en los modelos especificados para el total de España y para cada una de las regiones, ha sido la utilización de la metodología generalmente aplicada a datos de panel. En concreto, se ha empleado el estimador intragrupos o de efectos fijos. Dado que tenemos varias observaciones individuales para cada colegio, podemos plantear una estimación de datos de panel considerando que el panel tiene i colegios, cada uno con j estudiantes.

En la medida en que las variables inobservables α_i influyan en el resultado educativo de un alumno y además se correlacionen con los regresores utilizados en el modelo, los coeficientes de β_h y γ_i estarían sesgados si no se controla adecuadamente por la presencia de dichos efectos inobservables. Sin embargo, el estimador intragrupos permite una estimación consistente de β_h a partir de la siguiente transformación del modelo:

$$(y_{ij} - \bar{y}_i) = \sum_{h=1}^n \beta_h (x_{h,ij} - \bar{x}_{h,i}) + (u_{ij} - \bar{u}_i) \quad (3)$$

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. Sin embargo, al utilizar el estimador intragrupos se ha de prescindir de la estimación de los coeficientes relativos a las variables de colegio sin que ello presente ningún problema econométrico.⁸ Para la estimación consistente de los coeficientes de las variables de colegio habría que disponer de variables instrumentales que no estén correlacionadas con la heterogeneidad inobservable, pero sí con las variables de colegio cuyos coeficientes se quieren estimar. No obstante, esta opción no se considera viable dada la disponibilidad de información en la base de datos de PISA, de tal manera que para la estimación del Modelo España se ha procedido a la aplicación de Mínimos Cuadrados Ordinarios, con el fin de incluir las variables de colegio. Los resultados de esta estimación se han comparado con los resultados del estimador intragrupos para comprobar las posibles diferencias entre ambas técnicas. Por lo que respecta al caso del modelo regional se ha optado por utilizar el estimador intragrupos, sin incluir las variables de colegio en el análisis con el objetivo de obtener estimaciones insesgadas.

El modelo transformado (3) no tiene constante debido a que está expresado en diferencias respecto a la media del colegio, sin embargo, puede obtenerse el modelo estimado expresado en niveles y calcular una constante, una vez que se han estimado los coeficientes β_h , de la siguiente manera:

⁷ La lectura de los epígrafes 3.2.2. y 3.2.3 no es imprescindible para la comprensión del resto del texto.

⁸ Dado que estas variables son comunes para cada colegio, al tomar diferencias respecto a la media de cada colegio estas variables desaparecen de la estimación intragrupos.

$$\bar{y} = \underbrace{\mu + \bar{\alpha}}_{\text{Constante}} + \sum_{l=1}^m \gamma_l \bar{z}_l + \sum_{h=1}^n \beta_h \bar{x}_h \rightarrow \bar{y} - \sum_{h=1}^n \beta_h \bar{x}_h = c \quad (4)$$

Donde \bar{y} es la nota media de la población de estudiantes objeto de estudio (puede ser el total de estudiantes españoles, Modelo España, o el total de estudiantes regionales, Modelo Regional), $\bar{\alpha}$ es la media de los inobservables de colegio, \bar{z} es la media de las variables de colegio y \bar{x}_h es la media de las variables de estudiante utilizadas en el modelo. Por tanto, la constante que se pretende estimar está compuesta por la constante del modelo (1), más la media de los inobservables de colegio.

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 regresiones.⁹

⁹ La explicación del análisis con los 5 Valores Plausibles se puede ver en *PISA 2006 Technical Report. Chapter 9: Scaling PISA*

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)*.¹⁰ 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.

3.2.3. Análisis *Shift-Share*

Este análisis es utilizado normalmente en los trabajos sobre crecimiento económico para descomponer los cambios de una variable Y entre dos momentos del tiempo o las diferencias entre dos regiones en relación a dicha variable. También se encuentra el análisis *Shift-Share* en otros campos como la economía laboral. En la literatura reciente se pueden citar

cognitive data. OECD.

¹⁰ 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.*

algunas aplicaciones de esta metodología como: Castaldi (2009) que utiliza esta metodología para descomponer el crecimiento de la productividad agregada del trabajo en Europa, Estados Unidos y Japón; Bielik y Rajcainova (2008) que analizan una descomposición del crecimiento del empleo a lo largo de diez años para distintos países; y Li (2008) que analiza las variaciones interregionales de productividad en el sector industrial y manufacturero Chino.

En esta metodología se supone que el valor de una variable económica en una determinada región A es el resultado de calcular la media de los valores $Y_{j,A}$ que presentan cada uno de los sectores económicos j de la economía, ponderados por el peso de cada uno de ellos $\theta_{j,A}$. De manera que podemos expresar el valor de la variable Y en los países A y B de la siguiente forma:

$$Y_A = \sum_{j=1}^J \theta_{j,A} Y_{j,A} \quad (5)$$

$$Y_B = \sum_{j=1}^J \theta_{j,B} Y_{j,B} \quad (6)$$

A partir de aquí se pueden descomponer las diferencias entre dos regiones A y B como sigue:

$$Y_B - Y_A = \underbrace{\sum_{j=1}^J \theta_{j,A} (Y_{j,B} - Y_{j,A})}_{\text{Efecto región}} + \underbrace{\sum_{j=1}^J (\theta_{j,B} - \theta_{j,A}) Y_{j,A}}_{\text{Efecto especialización}} + \underbrace{\sum_{j=1}^J (\theta_{j,B} - \theta_{j,A}) (Y_{j,B} - Y_{j,A})}_{\text{Efecto interacción}} \quad (7)$$

El efecto región muestra las diferencias debidas a que, para un sector económico dado, la variable agregada difiere entre regiones. El efecto especialización recoge las diferencias en las medias debidas a la diferente especialización sectorial de cada región, una región puede especializarse en sectores que tienen mayor o menor valor del agregado económico a utilizar. Por último, el efecto interacción se interpreta como la medida en que una región se especializa más o menos que otra región en un sec-

tor económico que tiene mayor o menor valor del agregado que en esa otra región.

La descomposición anteriormente vista para cualquier variable objeto de estudio puede aplicarse al actual trabajo. El análisis Shift-Share permite hacer una descomposición de las diferencias regionales en resultado educativo a partir de la utilización del modelo regional que estimaremos. Si se expresa la nota media de dos regiones, a y b respectivamente, de la siguiente forma:

$$\bar{y}^a = c^a + \sum_{h=1}^n \beta_h^a \bar{x}_h^a \quad (8)$$

$$\bar{y}^b = c^b + \sum_{h=1}^n \beta_h^b \bar{x}_h^b \quad (9)$$

Donde: \bar{y}^a e \bar{y}^b , son la nota media de ambas regiones respectivamente; c^a y c^b , son las constantes del modelo de las mismas; β_h^a y β_h^b , son los coefi-

cientes para cada región; y, \bar{x}_h^a y \bar{x}_h^b , los valores medios de las variables utilizadas en el modelo en cada territorio.

Entonces se puede hacer una descomposición de las diferencias regionales en:

$$\bar{y}^b - \bar{y}^a = \underbrace{(c^b - c^a) + \sum_{h=1}^n \bar{x}_h^a (\beta_h^b - \beta_h^a)}_{\text{Diferencias en Rendimientos}} + \underbrace{\sum_{h=1}^n \beta_h^a (\bar{x}_h^b - \bar{x}_h^a)}_{\text{Diferencias en Características}} + \underbrace{\sum_{h=1}^n (\bar{x}_h^b - \bar{x}_h^a) (\beta_h^b - \beta_h^a)}_{\text{Componente Mixto}} \quad (10)$$

Para interpretar esta descomposición hay que tomar el punto de vista de la región a . Así, las *diferencias en rendimientos* hay que entenderlas como la variación que se produciría en la nota media de la región a si pasara a tener los coeficientes de la región b y se corresponde con el efecto región anteriormente visto. En otras palabras, este componente contabiliza las diferencias debidas a que los alumnos en a obtienen distintos resultados educativos a los de b a igualdad de características (o a igualdad de sector económico, en el caso de la descomposición de una variables agregada), por tanto es la parte de las diferencias regionales que no se explican por las características que cada Comunidad Autónoma presenta en relación a las variables explicativas tenidas en cuenta en nuestro modelo. En cuanto a las *diferencias en características*, vienen a interpretarse como la variación que se produciría en la nota media de la región a si sus alumnos pasaran a tener las características que tienen los alumnos de b y se corresponde con el efecto especialización del análisis *Shift-Share* clásico. En otros términos, este componente contabiliza las diferencias debidas a que los alumnos de a presentan características distintas (o especialización en sectores económicos distintos en el caso de variables agregadas) a los de b (por ejemplo, diferen-

en nivel educativo y ocupacional de los padres, entorno cultural del hogar, y actitud). Hay un tercer componente denominado *componente mixto* (efecto interacción) que surge de manera matemática o residual debido al tipo de descomposición por el que se opta, ya que tanto las *diferencias en rendimientos* como en *características* se calculan manteniendo constantes las diferencias en el componente complementario. Ambos componentes contabilizan los efectos de cambiar única y exclusivamente los coeficientes y las características, respectivamente. Sin embargo, la diferencia total entre regiones se debe a tener coeficientes y características distintas. No obstante, en el análisis realizado, este componente mixto no suele explicar, de media, más del 10% de las diferencias totales entre dos regiones, por lo que en la descripción de los resultados se prescinde del mismo.

4. Análisis bivariente



4. Análisis bivalente

4.1. Variables relacionadas con el resultado educativo

Para explicar las diferencias regionales encontradas, en este apartado se estudian que variables o factores tienen un mayor efecto, consideradas de manera aislada al resto, sobre el rendimiento educativo de un alumno. Estos resultados serán la base para la selección de las variables que se utilizarán posteriormente en el **análisis multivariante**. En la tabla A2 se cruzan las variables de resultado educativo en las tres disciplinas analizadas con distintas variables que se tratan en este trabajo como variables explicativas. En dicha tabla también se recoge el número de observaciones de la muestra utilizada que hay en cada una de las categorías de las variables expuestas.

La media del rendimiento educativo de los alumnos se sitúa, para el conjunto de España en 488 puntos para Ciencias, 480 en Matemáticas y 461 en Lectura frente a una media para el conjunto de la OCDE de 500 puntos en cada disciplina.¹¹ Sin embargo, existen diferencias

importantes entre regiones, cuando controlamos por distintas variables explicativas y esto es lo que se explica en lo que sigue.

Como se observa en la Tabla A2, Andalucía es, entre las regiones con muestra ampliada, la región española con el rendimiento educativo más bajo. En el otro extremo tenemos a La Rioja que presenta la característica inversa: el rendimiento es 32 puntos superior a Andalucía en ciencias, 46 puntos en matemáticas y 31 en lectura. Asimismo hay comunidades que se sitúan en lugares de bajo rendimiento educativo como es el caso de Cataluña así como el resto de regiones sin muestra ampliada que aparecen agrupadas en el grupo “resto de España”. El resto de comunidades con muestra ampliada se sitúan en una zona intermedia. Hay que destacar, sin embargo, el caso del País Vasco que en ciencias tiene un rendimiento medio-bajo, en matemáticas se sitúa en el centro, pero en lectura es la segunda región de mayor rendimiento.

Según los datos mostrados en la tabla, se observa que hay un efecto positivo en el rendimiento del alumno tanto de la

11 Es importante notar que no se puede comparar la nota de Ciencias con la nota de otras disciplinas. Decir que los estudiantes españoles sacan mejor nota en Ciencias que en Matemáticas

o Lectura es erróneo, dado que estas puntuaciones se realizan en cada disciplina respecto a la media de países participantes. Por tanto, sólo son comparables las medias de distintos grupos, individuales o regionales, para una misma disciplina.

educación de la madre como de la del padre. Estas dos variables están muy correlacionadas (el coeficiente de correlación entre ambas es de 0.53), esto es, ambos progenitores suelen tener un nivel educativo bastante similar. Debido a esto, puede ser difícil separar los efectos aislados de la educación del padre y de la madre. Para evitar este problema, se utilizan como variables explicativas diversas combinaciones del nivel educativo del padre y de la madre. Para simplificar el análisis, se clasificó el nivel educativo de cada progenitor en alto/bajo, donde alto incluye educación secundaria y superior. Por tanto, el nivel educativo de los padres tiene cuatro categorías, según cada cónyuge tenga nivel educativo alto o bajo.

Las categorías profesionales del ISCO se han reagrupado en categorías ocupacionales más amplias tal y como se observa en la tabla A2. Como es de esperar, los hijos de padres con ocupación alta sacan mejores notas que los de aquéllos con una ocupación de nivel bajo o aquéllos cuya madre es ama de casa.

También se muestran en esta tabla las notas de PISA de los alumnos según el número de posesiones de riqueza en el hogar, para cada una de las materias estudiadas. No se puede encontrar un patrón de comportamiento por el que se pueda llegar a una conclusión clara sobre la influencia de la riqueza familiar en el rendimiento educativo. Por ejemplo, observamos en ciencias que los individuos que no poseen ningún bien de los seleccionados en esta variable han obtenido 485 puntos y los que poseen los cuatro bienes obtienen una puntuación de 478. Para matemáticas los alumnos que poseen los cuatro bienes considerados han obtenido tres puntos menos que los estudiantes que no poseen ninguno.

Tal y como era de esperar se observa que a mayor índice cultural-educativo del hogar, medido según se explicó antes en función del número de posesiones culturales y educativas del mismo, mayor es el rendimiento educativo. Los efectos de esta variable se dan en todas las disciplinas. Para ciencias un alumno de cultura educativa alta obtiene 80 puntos más que un alumno de cultura educativa baja. Para matemáticas y lectura dicha cifra es 75 y 72 respectivamente.

Se puede ver que los alumnos nativos obtienen mayor nota que los alumnos inmigrantes en cualquiera de las tres materias estudiadas. Cabe preguntarse si este dato tiene su origen en el hecho de que ser inmigrante tiene una naturaleza propia que puede afectar a la educación del alumno o si, por otro lado, la diferencia de notas tiene su causa en las características que suelen presentar los inmigrantes en cuanto a otros factores influyentes como la ocupación y educación de sus padres, el nivel cultural del hogar, el nivel de riqueza familiar o el interés por los estudios. El análisis multivariante nos permitirá responder a esta cuestión.

Para el total de España los niños sacan mejor nota en matemáticas y en ciencias, que las niñas, aunque para esta última disciplina la diferencia no se puede considerar significativa. Por el contrario, las niñas presentan un mejor comportamiento en lectura que los niños.

Según el idioma hablado en el hogar por el estudiante se pueden observar algunas diferencias interesantes en las comunidades bilingües. Para la Comunidad Autónoma de Cataluña se observan diferencias notables en el rendimiento entre los estudiantes que hablan catalán en casa y los que hablan castellano. Los primeros obtienen 40 puntos más que los segundos. En cuanto a la Comu-

nidad de Galicia existen igualmente diferencias aunque menos importantes que en el caso de la Comunidad catalana y esta vez la ventaja es a favor de los que hablan castellano que obtienen 19 puntos más que los que hablan gallego. Los navarros que hablan Eusquera obtienen 40 puntos menos que los estudiantes navarros que hablan castellano en casa. Sin embargo, este dato hay que tomarlo con mucha cautela puesto que el tamaño del grupo de habla regional es muy reducido en esta región (solo son 92 las observaciones en este caso). Por último, el análisis bivalente muestra que no hay diferencias relevantes entre los alumnos del País Vasco que hablan Eusquera y los que hablan castellano. El rendimiento obtenido por los estudiantes que hablan en casa una lengua distinta a la regional o la nacional es considerablemente inferior a los otros grupos. No obstante, este hecho puede estar muy relacionado con el estatus inmigrante de los alumnos. Además, hay que tener cautela a la hora de obtener conclusiones de dicho resultado puesto que de nuevo las observaciones para esos grupos son muy pequeñas.

A continuación en la misma Tabla A2 se muestra el rendimiento educativo según el índice de actitud de los alumnos hacia las ciencias y el estudio. Las notas obtenidas por los alumnos que presentan un índice alto de actitud son considerablemente mejores que aquéllos estudiantes que tienen un índice medio y bajo. Las diferencias entre los dos grupos extremos (índice alto y bajo) son de 90, 77 y 60 puntos para ciencias, matemáticas y lectura respectivamente.

En cuanto a los alumnos repetidores disminuyen su rendimiento en gran medida respecto a los alumnos que no repiten curso. Las diferencias entre los alumnos que han repetido un curso y los que se encuentran matri-

culados en su correspondiente curso son 89, 90 y 85 para ciencias, matemáticas y lectura respectivamente. Estas diferencias son mayores con los alumnos que han repetido dos cursos o más: 144, 149 y 144 respectivamente para las mismas asignaturas.

Una vez vistas todas las características individuales, tanto las referidas a dotaciones iniciales como a las que miden el nivel de esfuerzo del alumno, se procede a realizar el análisis descriptivo en relación a las variables de grupo, que en este caso se refieren básicamente a las características del colegio del alumno. Estos factores son idénticos para todos los alumnos que pertenecen a un mismo centro. No obstante, en cada una de las tablas la columna de frecuencia (n) refleja el número de alumnos que va a un colegio que presenta la característica correspondiente.

Por lo que respecta a las variables de colegio, la descripción de los factores que parecen influir favorablemente en las notas de acuerdo al análisis bivalente se puede resumir en:

1. Mayor proporción en el número de chicas: cuantas más chicas, mejores notas.
2. Propiedad del colegio: los alumnos de colegios privados sacan las mejores notas y lo de los públicos las peores.
3. Porcentaje de financiación pública: cuanto mayor es el porcentaje de financiación pública, peores notas. Como es obvio, para los colegios públicos no tiene mucho sentido esta variable ya que la mayoría de los colegios públicos manifiestan tener, según su director, más de un 90% de financiación pública.

4. Porcentaje de repetidores en el colegio: cuantos más repetidores hay en el colegio, peores notas sacan los alumnos. Esta variable puede considerarse como una variable de entorno o contexto educativo en el que están inmersos los estudiantes. A mayor porcentaje de repetidores se entendería que peor es el contexto educativo del alumno y esto parece estar relacionado con un peor rendimiento académico.
5. Nivel educativo de los padres del colegio: aquellos alumnos que asisten a colegios donde hay un alto nivel educativo del conjunto de padres sacan mejor nota.
6. Política de agrupación de los alumnos: aquellos alumnos que van a colegios donde agrupan a los alumnos en función de su rendimiento en todas las asignaturas, sacan mejores notas.
7. Política de información del centro: los alumnos que asisten a colegios que difunden información sobre los datos académicos de los alumnos y la usan para tomar decisiones sacan mejores notas.
8. Presión de los padres: los alumnos que asisten a colegios donde los padres ejercen mucha presión sobre el centro, sacan mejores notas.
9. Competencia: los alumnos que asisten a colegios que tienen más competencia en su zona de influencia, sacan mejores notas.
10. Selección: los alumnos que asisten a colegios que seleccionan más a los alumnos, sacan mejores notas.
11. Porcentaje de inmigrantes: los alumnos que van a colegios con un porcentaje menor de inmigrantes, sacan mejores notas.

Llegados a este punto, es oportuno hacer dos comentarios. En primer lugar, estos resultados hasta aquí presentados tienen que ser confirmados por el análisis multivariante, ya que las relaciones aquí expuestas pueden responder a efectos espurios. Por ejemplo, se observa que los alumnos que van a un colegio privado o concertado tienen un rendimiento educativo superior al de los que van a un colegio público. Dado que las características de las familias y de los alumnos de estos colegios son distintas de las de los colegios públicos, esta aparente relación entre propiedad del colegio y nota media del alumno ha de interpretarse con precaución. El análisis multivariante nos permitirá obtener los efectos de la propiedad después de eliminar el efecto de otras covariables. En segundo lugar, llama la atención el hecho de que el número de alumnos por profesor tenga una relación contraria a la esperada, esto es, cuantos más alumnos por profesor, peor nota. En realidad, este efecto se debe a que el número de alumnos por profesor está relacionado con la propiedad del colegio. En los colegios privados y concertados, el número de alumnos por profesor es mayor que en los públicos¹². Dado que en los privados y concertados los alumnos sacan mejor nota, se obtiene que el ratio alumno-profesor tiene una relación contraria a la esperada. Esto ocurre también con el número de ordenadores por alumno (variable no mostrada). El ratio de alumnos por ordenador está inversamente relacionado con la nota. De nuevo, la explicación es que en los públicos hay menos alumnos por ordenador. Por esta razón, las variables que podemos denominar de “recursos” (profesores, ordenadores) no se han incluido en el análisis multivariante, ya que pue-

12 El número de alumnos por profesor en los colegios privados y concertados es aproximadamente 17, cuando el mismo ratio para los colegios públicos es de 10.

de estar relacionada con factores inobservables por los que no se puede controlar.

El análisis bivalente realizado permite, no obstante, tener una primera hipótesis sobre las variables que pueden influir en el rendimiento educativo. Dado que, como ya se ha mostrado, muchas de estas variables están correlacionadas con dicho rendimiento, se procederá a estimar un modelo econométrico que nos permita recoger la influencia de cada variable de manera aislada.

4.2. Características de las Comunidades Autónomas

Si se argumenta que las diferencias entre Andalucía y otras CCAA se deben en gran medida a las diferencias en características esto sólo puede ser debido a que la primera tiene una composición sociodemográfica “peor” en relación a un cierto número de las variables que afectan el resultado educativo. Para tener una aproximación de la importancia de las características en la explicación de las diferencias entre Andalucía y el resto de Comunidades Autónomas se presenta en el anexo una serie de gráficos en los que se compara la situación de las distintas CCAA.

En cuanto al nivel educativo de los padres, se observa en el Gráfico A1 que Andalucía está claramente peor, ya que es la región con menos padres de nivel educativo alto (secundaria o universitaria). El 60% de los alumnos andaluces tienen padre con educación secundaria o universitaria, cuando en la siguiente región con muestra ampliada que menos padres educados tiene ese mismo

porcentaje es del 76%. La región que más porcentajes de padres educados tiene es Asturias donde el 88% de los alumnos tiene padre con educación secundaria o universitaria. El patrón es muy parecido para el nivel educativo de la madre, la región que menos madres educadas tiene es Andalucía y la que más Cantabria. No obstante, se observa un hecho relevante que diferencia a Andalucía del resto de CCAA, es la única donde el porcentaje de madres con nivel educativo alto es ligeramente superior al porcentaje de padres con el mismo nivel educativo, sin embargo, en el resto de CCAA los padres están considerablemente más educados que las madres.

Otra variable que afecta al rendimiento educativo es la ocupación de los padres. Los alumnos con padre o madre de ocupación alta tienen un mayor rendimiento que aquéllos con padre o madre de ocupación baja, y éstos últimos, a su vez, también presentan mayor rendimiento que los de madres amas de casa. En el Gráfico A2 se observa la composición de las distintas CCAA en relación a la ocupación de la madre. En primer lugar, el porcentaje de alumnos andaluces con madre de ocupación alta (20%) es menor que en cualquier otra comunidad autónoma, siendo Cataluña la región en donde dicho porcentaje es mayor (38%). Andalucía es la región que más porcentaje de madres cuya ocupación es ama de casa (37%), mientras que en el otro extremo tenemos de nuevo a Cataluña con sólo un 13% de alumnos cuyas madres no participan en el mercado laboral. Por lo que respecta al porcentaje de madres con ocupación baja Andalucía presenta un patrón similar al del resto de regiones.

La cultura educativa del hogar es de las variables que más influyen en el resultado. El hecho de que un alum-

no pertenezca a un hogar con cultura educativa alta es incluso más importante que el hecho de que un alumno tenga padres con educación alta. Por ello, el tener en cuenta las características que presenten los alumnos en relación a esta variable aumenta la capacidad predictiva del resultado medio de una determinada región. Así, en el gráfico A3 se observa que Andalucía es la región que menos porcentaje de alumnos pertenecen a un hogar con una cultura educativa alta (un 29%) mientras que dicho porcentaje es de casi el 50% en Aragón, la región con más alumnos con cultura educativa alta.

Los inmigrantes obtienen considerablemente peor nota que los alumnos nativos por lo que el hecho de que en una región haya un porcentaje elevado de inmigrantes puede hacer que la nota media de dicha región disminuya considerablemente. A la vista del gráfico A4 se deduce que la peor nota media de Andalucía no se explica por este hecho. Mas bien al contrario, Andalucía es la segunda región que menos inmigrantes tiene, poco más del 2%. Por tanto, atendiendo únicamente a este factor se esperaría que la nota media de Andalucía fuese mejor que en el resto de CCAA. El caso contrario lo presentan Cataluña y Resto de España cuyo porcentaje de alumnos inmigrantes es aproximadamente 9% y 10%.

Las variables que más importancia tienen en la predicción del resultado educativo son las variables de carácter individual. Los alumnos con actitud hacia el estudio alta han obtenido un resultado educativo mayor en las pruebas. Igualmente el hecho de que un alumno repita supone que su nota disminuya en gran medida. En los siguientes gráficos A5 y A6 se observa claramente que Andalucía es la región que peores características presenta en relación a estas dos variables. El porcentaje de alumnos con actitud alta es de casi el 24%

mientras que dicho porcentaje llega a ser el 30% para Castilla y León, la región con más alumnos de actitud alta. En relación a los alumnos 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%. Andalucía es la región con mayor porcentaje de repetidores con un 47% frente al 23% del País Vasco, la región que menos repetidores tiene.

5. Modelo para el total de España



5. Modelo para el total de España

En esta sección se tratará de confirmar los efectos principales encontrados en el análisis bivalente previo mediante un análisis de regresión que busque los efectos condicionales de cada variable explicativa, manteniendo constantes el del resto de variables que pueden estar influyendo el resultado del alumno. El objetivo último es ver como quedan las diferencias en resultado entre comunidades autónomas en España cuando todos estos condicionantes son adecuadamente tenidos en cuenta.

Algunas de las variables que se usan en este análisis condicional son las que aparecen en la Tabla A3 del apéndice, donde se recoge la definición de algunas de ellas, por haber tenido un tratamiento especial respecto al análisis previo. La búsqueda de la mejor especificación para el modelo econométrico ha obligado a definir algunas variables de manera más agrupada, esto es, permitiendo menos categorías. A continuación se explican los principales cambios realizados de cara a estimar econométricamente el modelo propuesto.

En la regresión se incluyen variables de región que vienen a recoger el efecto de pertenecer a una comunidad autónoma o región determinada. En el modelo aparecen todas las regiones excepto Andalucía que está en el término independiente de la regresión, es decir, to-

das las comparaciones se harán con el individuo medio residente en Andalucía. 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. En el análisis bivalente parece haber diferencias entre los alumnos gallegos según la lengua que hablan en casa, sin embargo estas diferencias desaparecen en todas las especificaciones probadas en el proceso de estimación por lo que para esta comunidad autónoma no se ha diferenciado por la lengua que se habla en casa.

Finalmente, y en base a lo obtenido en el análisis bivalente, se incluyen variables que miden el nivel cultural en el hogar y algunas variables que definen características del colegio. Nótese que todas las variables son binarias, esto es solo pueden tomar valor 1 ó 0, a excepción de la variable que refleja la edad del estudiante (*Mes de nacimiento*) que debería interpretarse de distinta manera a las demás.

5.1. Especificación del modelo

El modelo multivariante desarrollado y estimado es el resultado de un proceso de búsqueda de la especificación más adecuada en términos de (1) variables que se deben incluir en el modelo en base a la evidencia empírica previa; (2) nivel de significatividad de las mismas; y (3) poder predictivo del modelo. Concretamente, a partir de una regresión inicial con todas las variables que en la sección anterior han resultado ser las más importantes de cara a explicar las diferencias observadas en resultado educativo, se han llevado a cabo una serie de especificaciones alternativas donde se han testado distintas combinaciones de las variables explicativas. Estas combinaciones se diferencian tanto por el número de variables explicativas incluidas en el modelo como por la categorización de las mismas o la introducción de interacciones entre ellas. La validez de cada una de las especificaciones ha sido testada a través de los correspondientes contrastes de significatividad de los coeficientes de las variables, lo que ha servido de base para rechazar los modelos sucesivos hasta encontrar la mejor especificación. Las variables que se incluyen en la especificación final pueden clasificarse en tres grupos. En primer lugar, hay variables que son significativas en todas las disciplinas estudiadas (ciencias, matemáticas y lectura). Hay algunas que sólo son significativas en la regresión para alguna disciplina. Y por último, hay algunas variables que no son significativas para ninguna de las tres regresiones pero que se mantienen en la especificación final porque han sido relevantes en otros estudios anteriores realizados con datos de PISA.

5.2. Resultados del modelo estimado

Los resultados de la estimación del modelo de regresión, estimado en primer lugar sin controlar por efectos inobservables de colegio, esto es, estimado por Mínimos Cuadrados Ordinarios e incluyendo tanto variables regionales, como individuales, variables relacionadas con las dotaciones iniciales de la familia y variables de colegio, se muestran en la Tabla A4. En ella se presentan los resultados para cada una de las disciplinas analizadas en PISA: ciencias, matemáticas y lectura. Esta tabla presenta para cada regresión el coeficiente estimado y una medida de la significatividad del coeficiente: el estadístico t de Student. Así, según los grados de libertad calculados para la estimación de cada coeficiente, cuando este estadístico es lo suficientemente alto se puede confiar que el coeficiente estimado es significativamente distinto de cero.

En todos estos modelos se está considerando como individuo de referencia, cuyo resultado aparece estimado en el término constante de cada regresión, a un alumno no repetidor, con residencia en Andalucía, nativo, de sexo masculino, con padre y madre de nivel educativo y ocupación baja, cuya madre trabaja, con una actitud ante la ciencia de nivel bajo y que vive en un hogar con una dotación cultural baja.¹³ Para este individuo de re-

13 La constante también incluye la categoría omitida en las variables de colegio. Concretamente recoge alumnos que estudian en un colegio público con un nivel educativo de los padres bajo, con un porcentaje de inmigrantes y repetidores bajo (menor del 3% para inmigrantes y 15% para repetidores), con mayoría de chicos en la clase. También se ha controlado en estas regresiones por el tamaño del municipio donde está el colegio, por el nivel de competencia del colegio en su zona de influencia y por

ferencia, el modelo predice un resultado medio de 478 puntos en ciencias, 485 puntos en matemáticas y 438 puntos en lectura.

En primer lugar, se presenta en la Tabla A4 las diferencias estimadas entre las regiones españolas, manteniendo constantes el resto de variables explicativas. Así se observa que, frente a Andalucía, el mismo individuo tomado como referencia obtiene entre 16.3 y 17.7 puntos más en ciencias si reside en La Rioja, Castilla León ó Galicia. Estas diferencias se obtienen, gracias a nuestro modelo multivariante, una vez que se han tenido en cuenta el resto de diferencias en las características observables tenidas en cuenta en dicho modelo. Las diferencias regionales en Matemáticas son aun mayores: van desde los 17.9 puntos con Cantabria y llegan hasta los 45.7 puntos para los Navarros que hablan una lengua distinta del castellano y Euskera. Por último, las diferencias en Lectura van desde los 10.9 puntos de distancia entre el alumno medio andaluz y el de Castilla León y los 24.2 puntos de diferencia con los alumnos de Galicia. Así por tanto, se encuentra que las diferencias de los alumnos andaluces existen, aunque como después comentaremos, son sensiblemente menores a las que se encuentran en el análisis bivalente donde no se controla por ninguna característica observable de los alumnos, los padres o los colegios. Llama la atención que, sin embargo, las diferencias son nulas con Asturias, Cataluña y con el resto de regiones españolas que no tienen muestra ampliada en PISA (Madrid, Comunidad Valenciana, Castilla la Mancha, Extremadura,

Murcia, Baleares y Canarias). También son mínimas con el País Vasco y con Navarra. En estas dos regiones solo parece que sacan mejores notas los alumnos que no hablan Euskera, aunque las diferencias solo son significativas para Ciencias y Lectura, en el caso del País Vasco, y Matemáticas, para el caso de Navarra.

El análisis realizado permite separar la influencia de las distintas variables explicativas a la hora de explicar las diferencias observadas entre alumnos de distintas regiones. Así, en el segundo panel de la Tabla A4 se tiene el efecto sobre el resultado educativo en las tres disciplinas estudiadas de las dotaciones iniciales con las que cuenta el hogar del alumno. En primer lugar se observa el **fuerte efecto que tiene el nivel educativo de los padres, así como el tipo de ocupación de los mismos**. Se encuentra que los hijos de padres con nivel educativo alto tienen, controlando por todo lo demás, un resultado entre 11 y 18 puntos superior a los hijos de padres con nivel educativo bajo. Es especialmente interesante que, en hogares donde solo uno de los padres tiene nivel educativo bajo, es el de la madre el que resulta fundamental para aumentar el resultado del alumno: si la madre tiene un nivel educativo superior al obligatorio (educación secundaria) la nota del alumno es entre 11.7 y 18 puntos superior mientras que si el de educación alta es el padre el resultado sube 11.7 puntos en Lectura, siendo no significativas las diferencias en las otras dos disciplinas. En cuanto al tipo de ocupación de los padres, tener un puesto de trabajo con cualificación requerida de nivel alto, sube el resultado del alumno entre 6 y 7 puntos. Es marcadamente interesante el resultado para la ocupación de las madres: el resultado de los alumnos cuyas madres no trabajan y se autodeclaran como amas de casa es entre 9.6 y 12.7 puntos inferior. Como este resultado se obtiene habien-

el tamaño de la clase donde está el alumno. Ninguna de estas variables resultan significativas en las especificaciones mostradas pero se mantiene su control por aumentar la generalidad descriptiva del modelo.

do controlado ya por el nivel educativo de la madre, nos estaría indicando que la inserción laboral de la madre puede ayudar a un mayor resultado del alumno, frente a la opción de madres que no trabajan. Estos resultados son especialmente importantes desde el punto de vista de la igualdad de oportunidades, en la medida en que el estatus educativo y ocupacional de los padres tengan una gran influencia en el resultado educativo mayor es la desigualdad de oportunidades dado que los alumnos no tienen forma de cambiar el estatus socioeconómico de sus familias. En un trabajo realizado por Checchi y Peragine (2005) para Italia con los datos de PISA se pone de manifiesto que las regiones del sur, en las que la educación de los padres tienen una mayor influencia en el resultado educativo que en la regiones del norte, también existe una mayor desigualdad de oportunidades medida a través de un índice de desigualdad.

Vemos como el mes de nacimiento tiene cierto efecto sobre el resultado. Así la distancia entre el alumno que nace en enero, que es el individuo de referencia considerado en el término independiente de la regresión, y el que nace en diciembre, que muestra un valor 11 en la variable “mes de nacimiento” es en Matemáticas de casi 10 puntos lo que está denotando, posiblemente, las diferencias en el nivel de madurez de los alumnos aun a estas edades. Las distancias son menores y poco significativas en Lectura (6.1 puntos) e intermedias en Ciencias (7 puntos). En cuanto al género del alumno se encuentra que los niños tienen un mayor resultado, *ceteris paribus*, que las niñas tanto en Ciencias como en Matemáticas mientras que ocurre lo contrario en Lectura, donde las niñas tienen un resultado superior en 23.1 puntos al de los niños. Finalmente, **los alumnos inmigrantes tienen un marcado menor resultado que los alumnos nativos**: su resultado medio en lectura es

19 puntos inferior al de los nativos, siendo las distancias máximas en ciencias donde el inmigrante medio tiene un resultado 30 puntos inferior al de un alumno nativo.

La variable compuesta que mide el nivel cultural de la familia ha emergido como una de las más importantes. **Los hogares con un nivel cultural alto suponen para el alumno medio un resultado hasta de 18.8 puntos más en Matemáticas, 17.8 puntos más en Ciencias y 15.8 en Lectura.** La importancia de esta variable es mayor que la de tener ambos padres con nivel educativo alto.

El siguiente panel de la Tabla A4 recoge una variable de decisión o responsabilidad individual del alumno que mide la actitud del mismo ante la ciencia (ver descripción en las secciones anteriores). Los resultados muestran que la influencia de esta variable sobre el resultado del alumno es muy importante. El coeficiente para la variable *Actitud alta* es incluso mayor que la suma del coeficiente de las variables *Madre y padre educación alta*, *Padre de ocupación alta* y *Madre de ocupación alta*. Esto significa que **el hecho de que un alumno tenga una buena actitud ante la ciencia, es más importante que el hecho de que ambos padres tengan educación y ocupación alta. El efecto predicho llega a aumentar el resultado en más de 53.8 puntos en Ciencias, 40.3 en Matemáticas y casi 28.1 puntos en Lectura.**

Asimismo, tal y como se indicaba anteriormente, **el hecho de repetir curso hace que el alumno tenga un resultado educativo sensiblemente inferior.** Así, este resultado es de hasta 67.5 puntos inferior en Matemáticas para alumnos que han repetido un curso y llega

a casi duplicarse para los alumnos que han repetido más de un curso (122.1 puntos). Las diferencias en Ciencias y Lectura son igual de significativas y amplias que lo son en Matemáticas. Hay que tomar con precaución, no obstante, el efecto de esta variable porque puede estar determinada por las mismas variables inobservables que afectan al resultado educativo del alumno. Por ello, se ha estimado el modelo también sin incluir esta variable en la especificación, obteniéndose los mismos resultados para el resto de variables explicativas (aunque con un nivel de diferencias superior al mostrado aquí). Asimismo, aunque no se muestran los resultados de éstas,¹⁴ también se ha realizado una estimación separada para los alumnos que no repiten y para los que repiten algún curso. Los efectos de las variables explicativas sobre el resultado de alumnos repetidores y no repetidores son básicamente los mismos aunque se encuentran algunas diferencias en la intensidad del efecto de algunas variables. Por ejemplo, el efecto de la edad del alumno, así como el del estatus laboral de la madre es más intenso para alumnos repetidores (los alumnos repetidores hijos de amas de casa o nacidos en los últimos meses del año tienen sensiblemente peores resultados que el resto). Por el contrario la actitud del alumno ante la ciencia afecta de manera más positiva a los alumnos no repetidores que a los que repiten algún curso.

Para las variables de colegio se encuentra evidencia de que **el nivel educativo de los padres de los compañeros del alumno influye de forma muy significativa sobre el resultado del mismo**. También parece afectar el porcentaje de niñas en el colegio que

aumenta algo, entre 6.8 y 7.7 puntos, el resultado en Matemáticas y Lectura. Sin embargo, el resto de variables de colegio tienen una significatividad más reducida, siendo estas variables significativas solo para algunas disciplinas y no para otras. Básicamente lo único que se encuentra es que **las diferencias entre colegio público, privado y concertado apenas existe, una vez que controlamos por otras características del colegio, de la clase en la que está el alumno y, sobre todo de las características individuales y de dotaciones iniciales del alumno**. Solo parece significativamente afectar algo el hecho de que el alumno esté en un colegio privado cuando los padres, según declara el director del colegio, ejercen un nivel de presión alto sobre la calidad del colegio. En estos casos se obtiene que el resultado en ciencias aumenta en 18 puntos.

A la vista de los resultados **hay, por tanto, tres Comunidades Autónomas que obtienen significativamente mejores resultados que la Comunidad Andaluza en todas las asignaturas una vez tenidas en cuenta las variables explicativas. Estas Comunidades son Castilla y León, Galicia y La Rioja**. Además obtenemos evidencia de un mayor resultado en Ciencias de los alumnos de Cantabria, en Matemáticas de los alumnos de Aragón, Cantabria y Navarra (si no hablan Euskera), y un mayor resultado en Lectura de los alumnos de Aragón, Cataluña (si hablan catalán) y del País Vasco que hablan en casa castellano.

Para ver en qué medida las variables explicativas incluidas en el modelo son la causa de las diferencias interregionales observadas se presentan los Gráficos A7, A8 y A9. En él se comparan las diferencias regionales para ciencias, matemáticas y lectura, con las mismas diferencias una vez tenidas en cuenta todas las variables

14 Aunque están a disposición del lector interesado.

explicativas de las tres especificaciones del modelo econométrico. En este último caso, solo se muestran las diferencias que han resultado ser estadísticamente significativas. En la medida en que las diferencias entre regiones se hayan reducido en las regresiones frente a las observadas a nivel descriptivo, se tiene un test de cuanto explica nuestro modelo de dichas diferencias regionales. Si una diferencia regional emerge como no significativa, se puede interpretar que la totalidad de las diferencias observadas a nivel descriptivo son explicadas por las variables incluidas en el modelo estimado, y por tanto, podemos concluir que no existen diferencias en las notas obtenidas por los alumnos de ambas regiones.

En estos gráficos se observa que las diferencias controlando por las variables explicativas del modelo econométrico se han reducido bastante para la gran mayoría de regiones. Asimismo, se observa que existen varias regiones donde las diferencias observadas a nivel descriptivo emergen como no significativas una vez controladas las diferencias observables entre regiones. Así, por ejemplo, las diferencias entre Andalucía y Asturias o las de Andalucía con el resto de regiones españolas que no tienen muestra ampliada en PISA son estadísticamente no significativas. Sin embargo, **las diferencias con Castilla León, Galicia y La Rioja son estadísticamente significativas para cualquiera de las tres disciplinas estudiadas y muy importantes, aunque se reducen también de manera considerable frente a las observadas a nivel descriptivo** (de más de 40 puntos a entre 10 y 30 puntos, dependiendo de la región y la disciplina).

Estas diferencias entre regiones que persisten una vez controladas las variables explicativas del modelo no

pueden ser atribuidas más que a otras variables no observables y que podrían estar relacionadas con el sistema educativo de estas regiones. También podría ser que algunas características del estudiante y que determinan su esfuerzo no hayan sido tenidas en cuenta y que las mismas se distribuyan de manera distinta entre las distintas regiones estudiadas. Así, a modo de ilustración, La Rioja que tiene una puntuación mayor a Andalucía en 46, 63 y 47, en ciencias, matemáticas y lectura respectivamente cuando no se condiciona por ninguna variable explicativa, muestra unas diferencias en el modelo condicional que siguen siendo altas aunque considerablemente menores: 16, 39 y 21 puntos respectivamente para las mismas asignaturas.

Otro resultado relevante y que merece la pena ser destacado es que existen regiones para las que las diferencias han cambiado de signo, frente a las diferencias observadas a nivel descriptivo. Para los vascos las diferencias en ciencias cambian de signo, siendo los alumnos andaluces, más exitosos en su resultado que los vascos en esta materia. Finalmente, en el caso de Cataluña, los alumnos que no hablan catalán obtienen un peor resultado en Ciencias que los andaluces, una vez se controlan todas las variables explicativas que influyen en el mismo.

Alternativamente, en la Tabla A5 se presentan los resultados de la estimación de los coeficientes de las variables de dotaciones iniciales y de esfuerzo individual cuando se controla por la presencia de efectos fijos inobservables de colegio. Se puede observar, que en general los coeficientes son bastante parecidos a los estimados con Mínimos Cuadrados Ordinarios, no obstante se pueden destacar algunas diferencias. La variable que incluye a los alumnos con madre de educación baja y

padre alta parece que gana significatividad en las tres disciplinas, al igual que las variables de ocupación alta de los padres. Los coeficientes estimados por efectos fijos de las variables de cultura educativa media y alta (8,32 y 13,53 respectivamente) son algo menores en el caso de la lectura que los estimados para MCO (10,27 y 15,81 respectivamente). Igualmente para la variable actitud alta se estima un coeficiente por efectos fijos (25,45) menor que por MCO (28,15). Estos resultados se deben interpretar como que, para el caso concreto de la especificación utilizada y los datos del total de España, la estimación por MCO no cambia en gran medida respecto al estimador intragrupos, sin embargo, el estimador intragrupos permite controlar en todo caso por variables inobservables por lo que será la estrategia de estimación para el modelo regional.

Finalmente, en los Gráficos A10, A11 y A12 se muestran las diferencias estimadas mediante la especificación alternativa del Modelo España separando los alumnos que repiten algún curso de los que no lo hacen. Con este ejercicio **se pretende contrastar si las diferencias regionales descritas anteriormente se distribuyen de manera distinta entre alumnos repetidores y no repetidores. Y efectivamente, parece que las diferencias son considerablemente distintas entre ambos colectivos.** En primer lugar se ve en este gráfico como las diferencias en Ciencias son mayores entre repetidores que lo que son entre los que no repiten para todas las regiones excepto para los alumnos de Cantabria, los Navarros y los Vascos que no hablan Eusquera. Estos dos últimos grupos parecen hacerlo sensiblemente mejor y peor, respectivamente, que los andaluces que no repiten. Los resultados son, sin embargo, mucho menos claros para Lectura y Matemáticas. En el caso de Matemáticas, parece que las diferen-

cias con respecto a Andalucía son mayores en el caso de los repetidores para Asturias, Castilla León, Galicia y La Rioja. Sin embargo para el resto de regiones, las diferencias con los andaluces se manifiestan de manera más intensa entre los no repetidores. Finalmente, en el caso del resultado en Lectura son solo los repetidores en Castilla León, en Galicia y La Rioja los que, en el caso de repetir, lo hacen sensiblemente mejor que los andaluces en esta situación. Los alumnos vascos que hablan Eusquera, en caso de repetir, sacan considerablemente peor nota que los andaluces. En el resto de casos, las diferencias también son mayores entre los alumnos que no repiten curso.

6. Modelos regionales



6. Modelos regionales

El análisis realizado en el anterior apartado supone que los parámetros a estimar son idénticos en cada una de las CCAA. Sin embargo, los coeficientes de cada una de las variables explicativas pueden diferir entre sistemas educativos. Esto es, es posible que, por ejemplo, el efecto de la educación de los padres sea distinto en diversas regiones. Para contemplar esta posibilidad hay que estimar una regresión para cada Comunidad Autónoma, ejercicio que se presenta en esta sección.

6.1. Resultados de la estimación

Una vez visto las variables que influyen en el resultado educativo en el modelo nacional se procede a estimar los correspondientes coeficientes en cada una de las CCAA para ver si estos difieren entre los distintos sistemas educativos regionales.

La tabla A6 muestra una regresión explicativa del resultado en ciencias para cada Comunidad Autónoma. Las tablas A7 y A8 muestran la misma información para matemáticas y lectura, respectivamente. En la constante se incluyen los hombres nativos que tienen padres de educación y ocupación baja, pertenecen a hogares

con cultura educativa baja, tienen una actitud hacia el estudio baja y no han repetido curso. Por ejemplo, en la regresión para ciencias de Andalucía los alumnos que están en la constante tienen un resultado estimado de 491,03. Las diferencias en las constantes de las distintas regiones implican diferencias en el resultado de los alumnos con estas características que se deben a efectos no observados y no controlados en la regresión. A modo de ejemplo, Aragón tiene una constante para la regresión de ciencias de 500,73 lo que implica que los alumnos aragoneses con estas características obtienen por término medio casi 10 puntos más que los alumnos andaluces.

Por otro lado, para cada región se han calculado unos coeficientes distintos. Puesto que todas las variables que se incluyen en el modelo son categóricas, excepto la variable edad, estos coeficientes se interpretan como el efecto que tiene en el resultado educativo cada categoría respecto de la que está en la constante. Por ejemplo, la variable **Madre y padre educación alta** tiene un coeficiente de 20,6 en la regresión para Andalucía, lo que quiere decir que el pasar de tener madre y padre de educación baja (categoría de los individuos de la constante) a tener ambos progenitores con educación alta supone un incremento del resultado educativo, *ceteris paribus*, de esa magnitud. También, se pueden observar

diferencias en los coeficientes de las variables entre distintas regiones lo cual muestra que estos efectos son distintos en cada Comunidad Autónoma. Por el ejemplo, la variable **Madre y padre educación alta** en la regresión para Aragón tiene un coeficiente de 8,05, mucho menor que en Andalucía, por lo que el efecto que tiene un cambio en la educación de los padres de los alumnos es menor en Aragón que en Andalucía. En la medida en que los coeficientes estimados para las distintas CCAA sean similares, las diferencias regionales se van a deber sobre todo a las características. Y, en la medida en que los coeficientes sean distintos, las diferencias en rendimientos serán más importantes.

A continuación, se procede a comentar los resultados para la asignatura de ciencias. No se harán comentarios para matemáticas y lectura ya que el lector puede trasladar fácilmente los mismos a estas otras materias.

La estimación del resultado en ciencias para los alumnos andaluces que presentan las características de la constante es de 491. Para el resto de regiones la misma estimación está entre los 475 puntos de País Vasco y los 526 de Galicia. Se observa un resultado sensiblemente mayor en 7 regiones: Aragón, Cantabria, Castilla y León, Cataluña, Galicia, La Rioja, y Navarra. Las regiones con peor resultados son País Vasco y Resto de España. Si bien, se estima una constante similar para Asturias.

Las variables que indican **la educación de los padres** tienen, en Andalucía, unos coeficientes estimados de 23,65 para los individuos con madre de educación alta y padre de educación baja, 17,37 para los individuos de madre de educación baja y padre alta, y 20,6 para los individuos con ambos progenitores con educación

alta. Estos coeficientes son mayores que los que se estiman para cualquiera del resto de las regiones, lo cual indica de que en la región andaluza el efecto de la educación de los padres es superior al que se estima para el resto de regiones. Otro aspecto a destacar es que Andalucía es la única región en la que las tres categorías¹⁵ de la variable ahora analizada son significativas al 1%. Esto es un aspecto destacable puesto que para la mayoría de las regiones ninguna de estas categorías es significativa o como mucho tan sólo tener madre y padre con educación alta tiene un efecto significativo (País Vasco).

Por lo que respecta a la **ocupación de los padres** para Andalucía la única variable que es significativa es **Madre ama de casa** cuyo coeficiente es negativo lo cual quiere decir que existe un diferencial positivo para los alumnos cuyas madres participan en el mercado laboral. Este hecho contrasta con otras regiones para las que no se encuentra un resultado significativamente distinto entre los alumnos con madres amas de casa y aquéllos con madres de ocupación baja, entre éstas están: Aragón, Castilla y León, Galicia, La Rioja, Navarra y Resto de España. Para Andalucía, no se encuentra que tenga efecto en las notas de los alumnos el hecho de tener padre o madre con ocupación alta, cuando estas variables si son significativas para otras regiones como: Aragón, Asturias, Cantabria, La Rioja y Navarra.

La **cultura educativa del hogar** de pertenencia tiene un efecto positivo para el resultado del alumno. A pesar de ello para Andalucía no ha podido encontrarse evidencia de efecto diferencial de los individuos que se inclu-

15 Madre educación alta y padre baja, Madre educación baja y padre alta y Madre y padre educación alta.

yen en el grupo de **cultura educativa media** respecto a los individuos de la constante. En otras regiones comparadas se observa claramente que tanto los individuos de cultura educativa media como los de alta obtienen un resultado significativamente distinto a los estudiantes de la constante: para Castilla y León los coeficientes de estas dos variables binarias para ciencias han sido 17,55 y 25,74 respectivamente, una cuantía bastante mayor a lo que podemos observar para Andalucía con 5,82 y 15,98 respectivamente. En el País Vasco, las dos variables son significativas, aunque en este caso, la importancia de los coeficientes es mucho menor y parecida a los coeficientes de Andalucía, con 8,82 y 14,37 para los alumnos de cultura educativa media y alta respectivamente. Para la mayoría de las regiones tan sólo es significativo el coeficiente de los individuos de cultura educativa alta y ese coeficiente es mayor que en Andalucía.

En las distintas regresiones las variables individuales, la actitud y la repetición de curso, tienen un efecto estimado muy importante en el resultado educativo. La variable de **actitud** es significativa en todas las regiones y tan sólo se encuentran diferencias en la cuantía de los coeficientes estimados. En cuanto al grupo de **actitud media**, el coeficiente va desde los 32,68 de Aragón a los 14,32 de Cantabria. El coeficiente para el grupo de **actitud alta** en Andalucía es el menor de todos con 46,66, por el contrario el coeficiente de la región Navarra es el mayor con 68,95. La variable que indica si el **estudiante repite** es sin duda la más importante a la hora de explicar el resultado educativo. Esta variable es significativa para todas las regiones sin excepción. Por lo que respecta a la cuantía de los coeficientes, Andalucía es la cuarta región donde más efecto negativo tiene el **repetir un curso**, con un coeficiente de -68,46, 12

puntos más que en Aragón, la región que menor efecto tiene, y 6 puntos menos que Navarra, la región que mayor efecto tiene. El efecto negativo en Andalucía para los alumnos que **han repetido dos o más cursos** es de -114,40, casi veinte puntos más que en Castilla y León, la región que menos efecto tiene, y 15 puntos menos que Galicia, la región que más efecto tiene. Por tanto, el efecto negativo de repetir en Andalucía está en torno a la media, situándose entre Aragón y Castilla y León, regiones en las que el efecto es menor, y Navarra y Galicia, regiones en las que el efecto es mayor. Para todas las regiones, los alumnos que han repetido una vez tienen una nota predicha bastante menor que los que no han repetido. Este efecto negativo suele ser duplicado en todas las regiones si el alumno ha repetido dos o más veces.

Por lo que respecta al **mes de nacimiento** y **sexo** del alumno son dos variables que merecen un comentario común. Ambas son variables que se espera que se distribuyan uniformemente entre todas las regiones. Por tanto, no se puede esperar que las diferencias entre regiones se expliquen por diferencias respecto a la composición por edad o por sexo. Sin embargo, sí puede haber diferencias en resultados entre los individuos de igual edad y género de distintas regiones. El **mes de nacimiento** no parece ser una variable relevante en Andalucía ya que el coeficiente estimado no es significativo. Esto diferencia a esta región del resto, ya que para todas las Comunidades Autónomas la variable edad tiene un coeficiente significativo al menos para alguna disciplina concreta. Para ciencias la región donde el coeficiente es más significativo es Aragón con -1,76, lo que se interpreta cómo el decremento de la nota por cada mes menos de edad de los alumnos. El **sexo del alumno** es una variable que sale significativa para to-

das las regiones estimándose un efecto positivo para los hombres en ciencias. Para la disciplina de ciencias Galicia y Andalucía son las regiones que mayor diferencial de notas se estima entre hombres y mujeres siendo los coeficientes de -23,27 y -22,29 respectivamente. Y la región que tiene el diferencial más bajo estimado es Aragón con -7,34.

Considerando la variable **inmigrante**, Andalucía, junto con Asturias y Galicia, son las únicas regiones en las que no se encuentra evidencia suficiente sobre el efecto de ser inmigrante en el resultado educativo a pesar se haberse estimado unos coeficientes muy elevados para cada una de ellas. El ser inmigrante tienen un efecto negativo significativamente distinto de cero para el resto de regiones. Además este efecto es muy importante y para darse cuenta de ello basta con citar que en Castilla y León un alumno inmigrante obtiene 47,81 puntos menos que un alumno nativo medio.

6.2. Diferencias en resultados para distintos grupos de estudiantes

En la tabla A9 se muestran las notas predichas para los estudiantes de cada una de las categorías de las variables que se tienen en cuenta en el modelo. Los resultados se muestran para cada una de las disciplinas analizadas comparándose las predicciones de, por un lado Andalucía; y por otro, el total del resto de CCAA. De nuevo, se comentarán los resultados para ciencias, dejando al lector el análisis de matemáticas y lectura.

La primera lectura de la tabla debe centrarse en las columnas (1) y (2) de cada una de las disciplinas: ciencias, matemáticas y lectura. Estas columnas contienen las *notas predichas* de Andalucía y del total del resto de CCAA, respectivamente. Cada una de las predicciones que se muestran se han calculado teniendo en cuenta los coeficientes estimados para los modelos de las Tablas A6, A7 y A8. La predicción para los estudiantes que pertenecen a una categoría concreta de una variable determinada se ha hecho suponiendo que dichos individuos presentan los valores medios del total de España en relación al resto de variables. Por ejemplo, para la *nota predicha* de las mujeres, tanto en Andalucía como en el resto de CCAA, se supone que el resto de variables distintas al género presentan el valor medio del total del país. Este supuesto se hace para que la diferencia entre las notas predichas para Andalucía con el resto de CCAA se deba exclusivamente al efecto rendimiento y no al efecto composición. Una segunda lectura de la tabla debe recaer en la columna (3), que nos indica las diferencias en rendimiento predicho entre Andalucía y el resto de CCAA para cada una de las categorías (a partir de ahora *diferencias en rendimientos*). Esta co-

lumna permite identificar cuáles son las categorías de alumnos andaluces que mayores o menores *diferencias en rendimientos* presentan con el resto de regiones. Por último, para cada variable y cada región comparada se puede obtener la diferencia entre la *nota predicha* de la mejor categoría (la que obtiene mejor nota en PISA) y la peor categoría (la que obtiene peor nota) para captar la variación de la nota para cada variable en las dos regiones comparadas (que en la tabla se especifica como *diferencias entre categorías*).

En cuanto a la asignatura de ciencias, para todas las categorías de cada una de las variables, se obtiene que la *nota predicha* para Andalucía está por debajo de la *nota predicha* para el resto de CCAA. Para la mayoría de las categorías la *diferencia en rendimientos* se sitúa alrededor de 11 puntos. Otro rasgo general es que hay determinadas categorías que presentan una *diferencia en rendimientos* superior al resto de categorías.

En el caso de la variable de educación de los padres la categoría de *Madre y Padre de educación baja* es la de mayor *diferencia en rendimientos* con 24,31 puntos en contraposición al resto de categorías para las que dicha cifra se encuentra entre 8,87 y 11,04. Este resultado se justifica por el hecho de que las *diferencias entre categorías* son bastante mayores en Andalucía. La diferencia que hay entre la *nota predicha* de un alumno con ambos padres de educación alta y la de un alumno de ambos padres de educación baja es de 20,6 en Andalucía y de 7,24 en el total de CCAA restantes. Esto indica que los alumnos con padres de nivel educativo bajo tienen un rendimiento educativo relativamente peor en Andalucía que en el resto de España. O lo que es lo mismo los alumnos con padres de mayor nivel educativo se encuentran en una situación relativa mejor en Andalucía que en el resto del país.

Otra variable que parece marcar diferencias entre Andalucía y el resto es la ocupación de la madre del alumno. Mientras que las categorías de madre con ocupación baja y alta tienen igual *diferencias en rendimientos*, entorno a 9,5, la misma cifra para los alumnos de madre ama de casa llega a ser de 20,35, lo cual se debe a que las diferencias entre los alumnos con madre ama de casa y los alumnos con madre de otro nivel ocupacional son mayores en Andalucía que en el resto de CCAA.

Para los inmigrantes se ha estimado unas *diferencias en rendimientos* de 25,05, bastante superior a la cifra para los nativos de 11,45. Parece que los inmigrantes tienen una situación relativa peor en Andalucía.

Igualmente el género marca otra diferencia entre Andalucía y el resto. Las *diferencias en rendimientos* estimadas para las mujeres es de 15,37 lo que supone casi el doble de la misma magnitud para los hombres, 8,58. Este hecho se debe a que la *diferencia entre categorías* (entre hombres y mujeres) es de 22,29 en la región andaluza frente a 15,5 en el resto. Por tanto, aunque los hombres tengan mayores notas predichas en las dos regiones comparadas, éstos gozan de una mejor situación relativa en Andalucía porque éstos se parecen más a sus homólogos del resto de CCAA de lo que se parecen las mujeres andaluzas a las mujeres del resto del país.

Por lo que respecta a la variable de actitud de los estudiantes, los alumnos con actitud alta son bastante mejores en términos relativos en el resto de España. La *diferencia de rendimiento* para dicha categoría es de 20,01 a favor del resto de CCAA mientras que la misma cifra para los alumnos con actitud alta y media está entre 8,44 y 9,34. Este resultado se debe a que

las *diferencias entre categorías* de actitud son bastante menores en Andalucía, lo que quiere decir que el tener una actitud alta supone menos incremento en la nota de PISA que en el resto de CCAA.

En cuanto a la variable repite la *diferencia de rendimiento* de los alumnos que han repetido un curso (15,37) es mayor que la misma magnitud para los alumnos que no han repetido curso (10,93) y los alumnos que han repetido al menos dos (11,94).

También es de señalar que no se encuentran diferencias entre las categorías de las variables ocupación del padre y cultura educativa.

Tras este ejercicio que ha servido para visualizar las diferencias en resultados educativos entre Andalucía y el resto de CCAA para alumnos homogéneos, se procede a calcular cual es la importancia de las *diferencias de rendimientos* y en *características* para explicar las diferencias medias entre la región andaluza y cada una del resto de Comunidades Autónomas.

7. Análisis *Shift-Share*



7. Análisis *Shift-Share*

En este apartado se va a hacer una descomposición de las diferencias observadas entre Andalucía y las otras regiones. Se va a calcular qué parte de estas diferencias se deben a *características* y qué parte a *rendimientos*. La diferencia en características se obtiene calculando la variación que se produciría en la puntuación de Andalucía si se cambiara su composición sociodemográfica por la de otra región determinada. Por lo que respecta a la diferencia en rendimientos, se obtiene calculando la variación que se produciría en la puntuación de Andalucía si se cambiara la nota en PISA de todos los grupos de alumnos con diferentes características por las notas que obtienen los grupos de alumnos análogos en otra región. Se supone, por tanto, en este ejercicio que cambian los rendimientos de los estudiantes andaluces por los rendimientos de los estudiantes de iguales características en otra región.

En los gráficos A13, A14 y A15 se pueden observar las diferencias entre Andalucía y otras regiones debidas a *rendimientos* y a *características* para ciencias, matemáticas y lectura respectivamente. En el eje vertical de cada gráfico se representan las diferencias en rendimientos y en el eje horizontal las diferencias en características, de modo que cada región está identificada por un punto según las coordenadas que le correspondan.

Por ejemplo, Andalucía tiene con Aragón, para la disciplina de ciencias, unas diferencias en características de 19,44 y unas diferencias en rendimientos de 22,85 y por eso esta representado con las coordenadas (19,44; 22,85). Por otro lado la línea azul del gráfico une a los puntos en los que las diferencias en rendimientos son iguales a las diferencias en características de tal manera que las regiones que están por encima de esta línea tienen mayores diferencias en rendimientos que en características, las regiones por debajo de la línea se diferencian más en características que en rendimientos, y las regiones que están sobre la línea son para las que las diferencias en características y rendimientos son iguales.

También se presentan resultados de descomposición en las tablas A10, A11 y A12, en las que se presenta la nota observada para cada región y la nota predicha para Andalucía en caso en que tuviera las características de las distintas regiones. En estas tablas también se establece una columna que indica el porcentaje de las diferencias regionales explicadas por *características*.

A la vista de los tres gráficos se observa que cinco regiones se sitúan por encima de la línea azul. Éstas son La Rioja, Castilla y León, Aragón, Galicia y Cantabria. Esto quiere decir que para todas las disciplinas las

regiones citadas presentan mayores diferencias con Andalucía en rendimientos que en características. De todas las anteriores regiones la que más diferencias en rendimientos presenta es La Rioja. Si Andalucía pasara a tener los rendimientos de esta región las diferencias de puntuación en PISA entre ambas se reducirían en un 66,48%, 77,03%, y 66,35% para ciencias, matemáticas y lectura respectivamente. No obstante, las diferencias en características siguen siendo importantes para explicar las diferencias observadas entre la región andaluza y las mismas regiones anteriormente citadas. Por ejemplo, si Andalucía llegara a tener la composición sociodemográfica de Galicia (las características de Galicia) las diferencias de puntuación en PISA entre ambas se reducirían en un 48,60%, un 42,50%, y en un 33,23% para la disciplina de ciencias, matemáticas, y lectura respectivamente.

Para todas las asignaturas hay cuatro regiones que se sitúan debajo de la línea azul. Éstas son Resto de España, Cataluña, País Vasco, y Asturias. Se puede decir que las diferencias observadas entre estas regiones y Andalucía se deben sobre todo a las diferencias en características y no tanto en rendimientos. Las diferencias en características son muy importantes a la hora de explicar las diferencias con País Vasco, Cataluña y Resto de España. Si Andalucía tuviera las características de País Vasco o de Cataluña, su puntuación en ciencias aumentaría en 31,67 o 22,53 puntos respectivamente llegando a superar la nota observada de ambas regiones. De hecho, el valor de los rendimientos estimados en ciencias para País Vasco es de -4,9 puntos. No ocurre igual para las asignaturas de matemáticas y lectura, ya que las diferencias observadas con Cataluña y País Vasco no vienen explicadas tan sólo por las características ya que éstas sólo suponen el 90%, y 68% de las diferen-

cias totales con Cataluña, para matemáticas y lectura respectivamente, y un 74% y 58% con País Vasco, para las mismas asignaturas respectivamente. En cuanto a Asturias, las diferencias observadas también se explican en gran medida por los rendimientos. De hecho si los alumnos andaluces llegaran a tener los rendimientos de los asturianos las diferencias se acortarían en 16, 15, y 11 puntos para ciencias, matemáticas y lectura respectivamente.

Navarra presenta un caso peculiar puesto que la estimación para rendimientos y características depende de la asignatura de la que se esté hablando. Para el caso de ciencias se encuentra por encima de la línea azul lo que significa que las diferencias se deben sobre todo a características (24,47) y no a rendimientos (18,06). Para el caso de matemáticas ocurre lo contrario, y se estima que la diferencia en características es menor que en rendimientos, con 23,28 y 31,15 respectivamente. Por último, en relación a la disciplina de lectura se observa a Navarra muy próxima a la línea azul por lo que las diferencias en características son muy similares a las diferencias en rendimientos, 20,97 y 19,72 respectivamente.

8. Conclusiones



8. Conclusiones

El análisis realizado con el *Modelo para el Total de España* ha permitido comparar el rendimiento educativo de estudiantes de diversas comunidades autónomas una vez ajustamos por diversas variables que pueden estar influyendo en dicho rendimiento. El hecho de que estudiantes de distintas regiones con iguales características difieran en el rendimiento educativo, indica que hay 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 elementos relacionados con el sistema educativo regional, o incluso variables de los estudiantes o sus colegios que no son observables en este trabajo. Los alumnos de la Rioja, Castilla León y Galicia presentan un rendimiento educativo sensiblemente superior al mostrado por los alumnos andaluces para todas las asignaturas e incluso cuando analizamos de manera separada a los alumnos repetidores y no repetidores. También los alumnos no repetidores de Cantabria lo hacen significativamente mejor para todas las asignaturas comparándolos con sus homólogos andaluces. El resto de regiones se sitúan en una posición u otra dependiendo de la disciplina y el grupo de alumnos (repetidores o no) que tengamos en cuenta. Por ejemplo, los alumnos Navarros que hablan castellano son significativamente mejo-

res en ciencias y matemáticas que los andaluces, pero no ocurre lo mismo para lectura. También se observa que los alumnos Vascos que hablan Eusquera parecen hacerlo peor que los andaluces en ciencias.

Lo importante, no obstante, de los resultados presentados con el *Modelo para España* es que gran parte de las diferencias existentes entre alumnos de Andalucía y el resto de regiones se han conseguido explicar con las variables incluidas en el modelo, lo cual demuestra que dichas variables (especialmente las relacionadas con el interés por el estudio del alumno, su carácter de repetidor de curso o no y las características socioculturales de su familia) son en gran medida la causa de los diferentes resultados que se obtuvo en el examen de PISA según la región del estudiante.

El análisis *Shift-Share* ha permitido descomponer las diferencias regionales entre Andalucía y el resto en *características y rendimientos*. Para ello, se ha estimado un modelo distinto para cada comunidad autónoma con muestra ampliada a través del *Modelo Regional*. Los resultados muestran que cada región presenta rasgos diferenciales. Los coeficientes estimados en cada región son diferentes en importancia y en significatividad. Entre los rasgos diferenciales más importantes de la

región andaluza están: un mayor efecto de la educación de los padres que en la mayoría del resto de regiones; un mayor efecto negativo para los alumnos cuyas madres son amas de casa; un mayor efecto positivo para los hombres en ciencias y matemáticas, pero un menor efecto positivo para las mujeres en lectura; y un menor efecto positivo de la actitud de los alumnos en ciencias y matemáticas.

Otro de los resultados encontrados con la descomposición de las diferencias regionales es que a pesar de que los alumnos andaluces tienen claras desventajas de partida respecto al resto de Comunidades Autónomas debido al nivel educativo y ocupacional de los padres, y la cultura educativa del hogar, las diferencias entre Andalucía y cada una de las regiones con las que se compara no se deben únicamente a diferencias en características, a excepción de los casos de Cataluña y País Vasco para la disciplina de ciencias. El modelo estimado permite predecir que si los alumnos andaluces llegaran a tener las características de los alumnos vascos o catalanes, la nota media de Andalucía superaría la nota observada para País Vasco y Cataluña en ciencias, y se recortaría la mayoría de las diferencias en matemáticas. No obstante, este resultado sólo se daría si los coeficientes estimados para Andalucía permanecieran inalterados ante un cambio sociodemográfico. Es probable que si se produjera un cambio en la composición socioeconómica de las familias andaluzas también cambiaran los rendimientos que las variables socioeconómicas tienen sobre el resultado educativo, en cuyo caso los resultados serían diferentes.

La descomposición de las diferencias regionales en *características* y en *rendimientos* es diferente dependiendo de la región que se compara con Andalucía. Así,

parece que las *características* explican gran parte de las diferencias con País Vasco, Cataluña, Resto de España, y Asturias. Por lo que respecta a los *rendimientos*, explican mejor las diferencias con Castilla y León, La Rioja, Aragón, Cantabria, y Galicia. Con Navarra las *diferencias en características* parecen ser similares a las *diferencias en rendimientos*.

Diferenciando por categorías o grupos de estudiantes también hay diversidad en cuanto a las *diferencias en rendimientos*. Parece que en Andalucía los alumnos que tienen ambos padres con educación baja están en una situación de rendimiento relativo peor, respecto al resto de grupos de educación de los padres, debido a que los alumnos andaluces con padres de baja educación se distancian de sus homólogos en otras regiones bastante más de la distancia existente entre los alumnos con padres de educación alta entre Andalucía y otras regiones. Igualmente, los alumnos que tienen madre con ocupación ama de casa están en una situación de rendimiento relativo peor en Andalucía que en el resto de las Comunidades Autónomas. Otros grupos que tienen una situación relativa peor en Andalucía respecto al resto de regiones estudiadas son: las mujeres, los inmigrantes y los alumnos de actitud alta.

Bibliografía



Bibliografía

- (1) Abburrà, L. (2005), "As good as the others. Northern Italian students and their peers in other European regions". Istituto Ricerche Economico Sociali del Piemonte. November.
- (2) Bielik, P. y Rajcaniova, M. (2008), "Shift-share analysis of employment growth: the case A of the V4 countries", *Agricultural Economics-Zemedelska Ekonomika*, 54: 347-351.
- (3) 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.
- (4) Carabaña, J. (2008), "Las diferencias entre países y regiones en las pruebas PISA". Mimeo. [www.collegeemeritos.es/docs/repositorio/es_ES/documentos/pisa_carabana_\(vf\).pdf](http://www.collegeemeritos.es/docs/repositorio/es_ES/documentos/pisa_carabana_(vf).pdf)
- (5) Castaldi, C. (2009) "The relative weight of manufacturing and services in Europe: An innovation perspective", *Technological Forecasting and Social Change*, 76: 709-722.
- (6) Checchi, D. y Peragine, V. (2005) "Regional Disparities and Inequality of Opportunity: The Case of Italy", *IZA Discussion Paper Series*, 1874.
- (7) Ciccone, A. y Garcia-Fontes, W. (2008) , "The quality of the Catalan and Spanish Education Systems: A perspective from PISA". Mimeo. www.antonio-ciccone.eu/wp-content/uploads/2008/11/cc-quality-of-education-system-9-with-tables.pdf
- (8) 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.
- (9) Coleman, J. S., T. Hoffer (1987), "Public and private high schools. The impact of communities. New York: Basic Books.
- (10) 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.
- (11) 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.

- (12) Feinstein, L. y J. Symons (1999), "Attainment in secondary education." *Oxford Economic Papers*, 51: 300-321.
- (13) Fertig, M. (2003), "Who's to Blame? The Determinants of German Students' Achievement in the PISA 2000 Study." *IZA Discussion Paper Series*, 739.
- (14) Hanushek, E. A. (1986), "The economics of schooling", *Journal of Economic Literature*, 24, pp. 1141-1177.
- (15) Henderson, V., P. Mieszkowski y Y. Sauvageau (1978), "Peer group effects in educational production functions." *Journal of Public Economics*, 10 (1): 97-106.
- (16) Instituto de Evaluación. Ministerio de Educación y Ciencia. "PISA 2006. (2007) Programa para la Evaluación Internacional de Alumnos de la OCDE. Informe Español". www.institutodeevaluacion.mec.es/contenidos/internacional/pisainforme2006.pdf
- (17) Junta de Andalucía. Consejería de Educación. (2007) "Informe PISA 2006. Resultados en Andalucía".
- (18) Li, D. (2008), "Interregional productivity variations in Chinese manufacturing and mining industry", *Applied Economics Letters*, 15: 1073-1077.
- (19) Noell, J. (1982) "Public and Catholic schools: A re-analysis of public and private schools." *Sociology of Education*, 55: 123-132.
- (20) OCDE. (2005) "PISA 2003 Technical Report."
- (21) OCDE. (2006) "Assessing Scientific, Reading and Mathematical Literacy: A framework for PISA 2006"
- (22) OCDE. (2008) "Informe PISA 2006. Competencias científicas para el mundo del mañana"
- (23) OCDE. (2009) "PISA 2006 Technical Report."
- (24) Opdenakker, M. C. y J. Van Damme (2006), "Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and Catholic schools and types of schools." *School Effectiveness and School Improvement*, 17 (1): 87-117.
- (25) Smith, J. y Naylor (2005), "Schooling effects on subsequent university performance: evidence for the UK university population," *Economics of Education Review*, 24: 549-562.
- (26) Sprietma, M. (2006), "Regional School comparison and school choice: how do they relate to student performance?", *Département des Sciences Économiques de l'Université catholique de Louvain. Discussion paper*, 2006-2.
- (27) Summers, A. A. y B. L. Wolfe (1977), "Do schools make a difference?" *American Economic Review*, 67 (4): 639-652.
- (28) Zinovyeva, N., Felgueroso, F. y Vazquez, P. (2008). "Immigration and Students' Achievement in Spain," *Working Papers 2008-37*, FEDEA.

Apéndices



Apéndices

Tabla A1. Número de observaciones

Región	Muestra
Andalucía	1.463
Aragón	1.526
Asturias	1.579
Cantabria	1.496
Castilla y León	1.512
Cataluña	1.527
Galicia	1.573
La Rioja	1.333
Navarra	1.590
País Vasco	3.929
Resto de España	2.076
Total España	19.604

Tabla A2. Resultado educativo según distintas variables de la base de datos

Variable	Categoría	n	Resultado en ciencias	Resultado en matemáticas	Resultado en Lectura
REGIONES					
Región	La Rioja	1.333	520	526	492
	Castilla y León	1.512	520	515	478
	Aragón	1.526	513	513	483
	Navarra	1.590	511	515	481
	Cantabria	1.496	509	502	475
	Asturias	1.579	508	497	477
	Galicia	1.573	505	494	479
	País Vasco	3.929	495	501	487
	Cataluña	1.527	491	488	477
	Resto de España	2.076	484	473	453
	Andalucía	1.463	474	463	445
DOTACIONES INICIALES					
Nivel educativo de los padres	Madre educación baja y padre baja	1.819	442	436	416
	Madre educación alta y padre baja	1.226	470	463	447
	Madre educación baja y padre alta	1.542	481	472	455
	Madre y padre educación alta	13.947	508	498	479
Ocupación de la madre	Alta ocupación	6.313	528	517	494
	Baja ocupación	8.060	475	467	453
	Ama de casa	4.492	470	462	441
Ocupación del padre	Alta ocupación	7.160	524	512	492
	Baja ocupación	11.026	473	466	447
Número de posesiones riqueza	0	1.356	485	475	462
	1	3.554	494	483	470
	2	5.327	491	483	463
	3	5.501	493	485	463
	4	3.773	478	471	449
Cultura educativa	Baja cultura educativa	4.445	446	441	422
	Media cultura educativa	6.482	486	477	460
	Alta cultura educativa	8.327	526	516	494

Variable	Categoría		n	Resultado en ciencias	Resultado en matemáticas	Resultado en Lectura
Estatus inmigrante	Nativo		18.456	494	485	465
	Inmigrante		911	434	429	415
Género	Femenino		9.801	486	476	479
	Masculino		9.803	491	484	443
Idioma hablado en las Comunidades bilingües	Cataluña	Castellano	856	481	474	469
		Catalán	532	521	520	505
		Otra	56	437	450	422
	Galicia	Castellano	821	515	505	492
		Gallego	703	496	484	468
		Otra	15	457	456	418
	Navarra	Castellano	1.407	516	518	484
		Eusquera	92	476	505	473
		Otra	42	488	487	443
	País Vasco	Castellano	2.998	498	503	493
		Eusquera	737	496	509	484
		Otra	58	426	414	419
VARIABLES INDIVIDUALES						
Índice de Actitud	alto		5.432	541	526	498
	medio		8.815	484	475	457
	bajo		5.247	450	448	437
Curso académico	No repetidor		13.515	528	520	499
	Repite un curso		5.009	439	430	414
	Repite 2 o más		1.080	384	371	355
VARIABLES DE COLEGIO						
Porcentaje de chicas	menos del 50%		9.973	485	478	458
	más del 50%		9.608	492	482	464
Propiedad del colegio	Privado		1.321	537	527	505
	Concertado		6.846	503	495	482
	Público		11.053	475	466	446
Porcentaje de financiación pública	Colegios privados hasta el 29%		627	550	540	519
	Colegios privados más del 29%		694	516	507	485
	Colegios privados hasta el 80%		3.531	512	504	487
	Colegios privados más del 80%		3.315	492	485	476

Variable	Categoría	n	Resultado en ciencias	Resultado en matemáticas	Resultado en Lectura
Porcentaje de repetidores en ESO	Menos del 15%	10.201	508	501	483
	15% o más	8.341	473	463	444
Nivel educativo de los padres de los compañeros	Bajo	9.781	470	463	444
	Alto	9.823	524	514	493
Comunidad en la que se encuentra el centro	de menos de 100.000 habitantes	11.709	482	474	454
	de más de 100.000	7.872	499	489	471
Agrupación dentro de las clases	Para todas las materias	1.519	507	492	471
	Para algunas o ninguna materia	15.811	486	478	460
Número total de alumnos por profesor en el centro	<= 9,24	5.935	472	466	446
	9,25 - 14,52	5.868	486	478	458
	14,53+	5.916	512	504	487
Información	Bajo	12.162	486	476	455
	Alto	6.510	494	486	471
Presión	Muchos padres	1.540	525	517	499
	Minoría de padres	17.835	485	477	457
Competencia	dos colegios o más	13.122	494	485	467
	uno o ninguno	6.224	479	471	451
Selección	Bajo nivel de selección	13.965	482	475	455
	Alto nivel de selección	4.969	513	501	486
Porcentaje de inmigrantes	3% o menos	9.832	494	488	467
	más de 3%	9.772	485	475	457

Gráfico A1. Porcentaje de padres educados
por CCAA

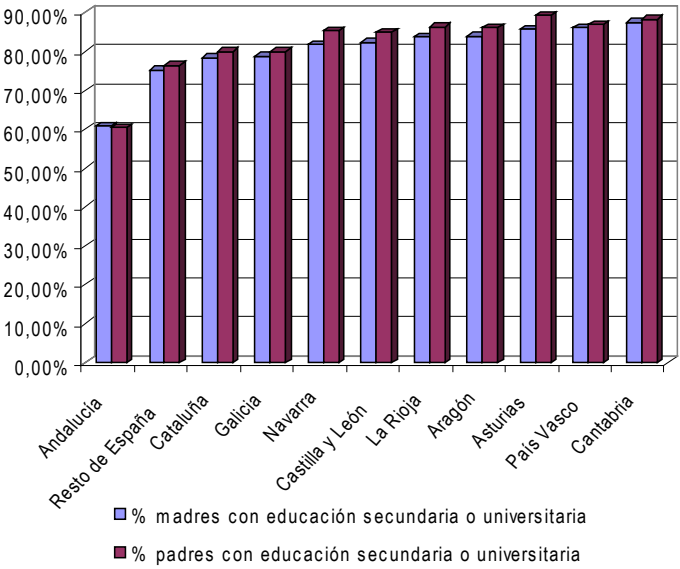


Gráfico A2. Ocupación de las madres por CCAA

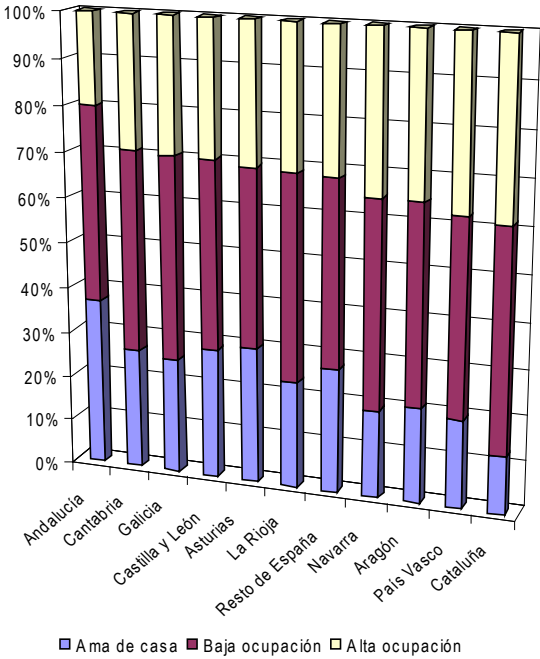


Gráfico A3. Porcentaje de alumnos con hogar de
cultura educativa alta en cada CA

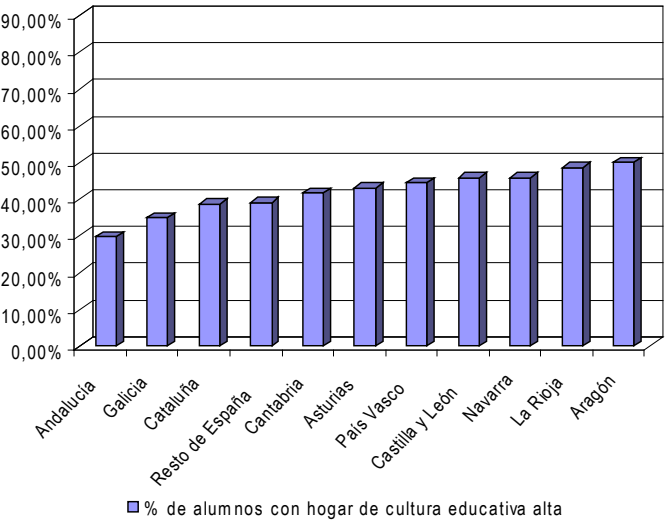


Gráfico A4. Porcentaje de alumnos inmigrantes
en cada CA

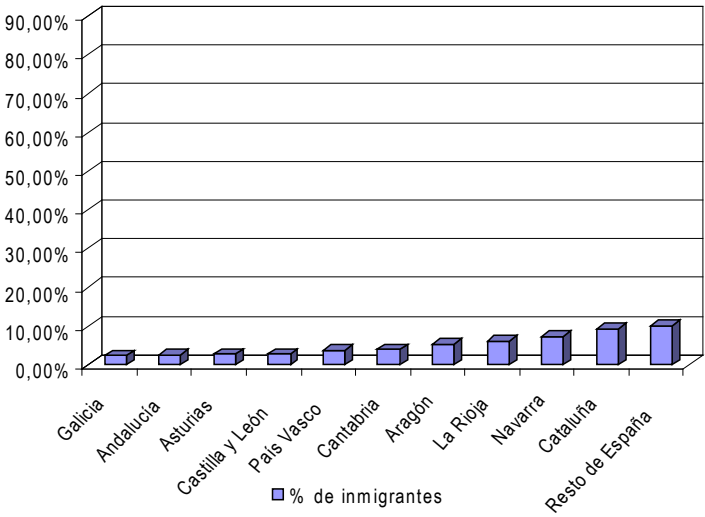


Gráfico A5. Porcentaje de alumnos con actitud alta en cada CA

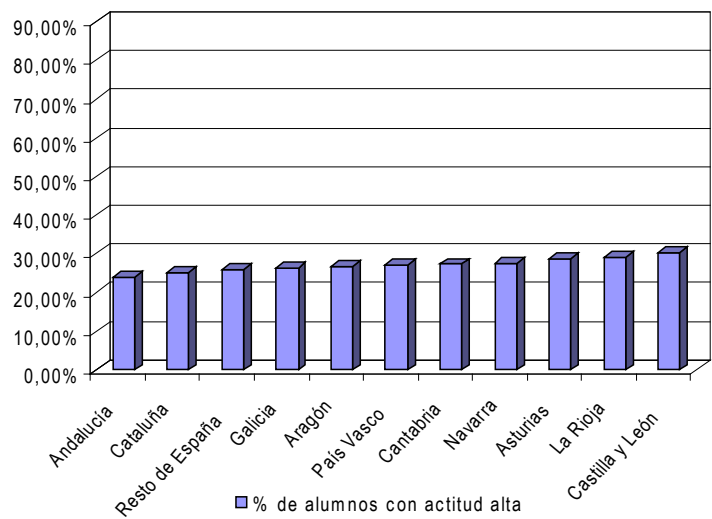


Gráfico A6. Porcentaje de repetidores en cada CA

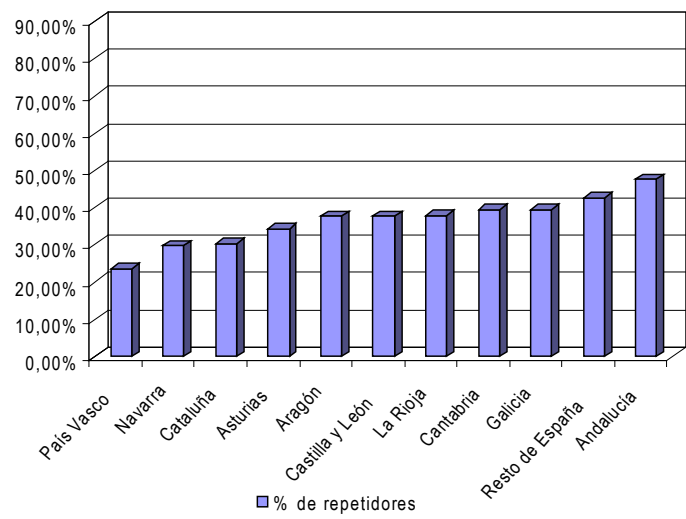


Tabla A3. Variables del modelo

Concepto y Variables	n	Explicación
REGIONES		
Cataluña hablan catalán	532	Variable que toma valor 1 para estudiantes de Cataluña que hablan catalán en su casa
Cataluña hablan castellano	856	Variable que toma valor 1 para estudiantes de Cataluña que hablan castellano en casa
Cataluña hablan otra lengua	56	Variable que toma valor 1 para estudiantes de Cataluña que hablan otra lengua distinta a las anteriores en casa
Navarra que hablan Eusquera	92	Variable que toma valor 1 para estudiantes de Navarra que hablan Eusquera en casa
Navarra que hablan castellano	1407	Variable que toma valor 1 para estudiantes de Navarra que hablan castellano en casa
Navarra que hablan otra lengua	42	Variable que toma valor 1 para estudiantes de Navarra que hablan otra lengua en casa
País Vasco que hablan Eusquera	737	Variable que toma valor 1 para estudiantes del País Vasco que hablan Eusquera en casa
País Vasco que hablan castellano	2998	Variable que toma valor 1 para estudiantes del P. Vasco que hablan castellano en casa
País Vasco que hablan otra lengua	58	Variable que toma valor 1 para estudiantes del País Vasco que hablan otra lengua distinta en casa
DOTACIONES INICIALES		
Mes de nacimiento	18101	Mes en que nació el estudiante distinto de enero. La variable va del valor 0, si nació en Febrero, a 11, si nació en Diciembre.
EDUCACIÓN DE LOS PADRES		
Madre educación alta y padre baja	1542	Variable que toma valor 1 si el alumno es de madre con educación alta y padre baja
Madre educación baja y padre alta	1226	Variable que toma valor 1 si el alumno es de madre con educación baja y padre alta
Madre y padre educación alta	13947	Variable que toma valor 1 si el alumno es de madre y padre con educación alta
OCUPACIÓN DE LOS PADRES		
Padre de ocupación alta	7160	Variable que toma valor 1 para alumnos que tienen padre con ocupación alta
Madre de ocupación alta	6313	Variable que toma valor 1 para alumnos que tienen madre con una ocupación alta
Madre ama de casa	4492	Variable que toma valor 1 para alumnos que tienen madre ama de casa
NIVEL CULTURAL EDUCATIVO		
Cultura educativa media	6482	Variable que toma valor 1 para alumnos cuyo hogar tienen un nivel cultural-educativo medio.
Cultura educativa alta	8327	Variable que toma valor 1 para alumnos cuyo hogar tiene un nivel cultural-educativo alto
VARIABLES DE COLEGIO		
Mayoría de chicas en el colegio	9608	Variable que toma valor 1 para estudiantes cuyo colegio tiene más de un 50% de chicas
Porcentaje de repetidores alto	8341	Variable que toma valor 1 para estudiantes que asisten a colegios con un porcentaje de repetidores superior al 15%
Colegio concertado con mucha presión	860	Variable que toma el valor 1 para los alumnos que asisten a colegios concertados con mucha presión.
Colegio privado con mucha presión	331	Variable que toma valor 1 para alumnos que asisten a colegios privados con mucha presión.
Padres de los compañeros con nivel educativo alto	9823	Es una variable que toma el valor 1 para los alumnos que asisten a un colegio donde el nivel educativo medio de los padres de los estudiantes es alto

Tabla A4. Resultados del Análisis de Regresión. Modelo España. MCO.

VARIABLE	CIENCIAS		MATEMÁTICAS		LECTURA	
	COEF.	t	COEF.	T	COEF.	t
Constante	477,93***	71,90	485,04***	72,74	438,05***	59,40
REGIONES						
Aragón	9,11	1,66	23,86***	4,18	11,34*	1,67
Asturias	3,12	0,60	7,04	1,39	5,36	1,04
Cantabria	11,67**	2,04	17,90***	3,86	9,44	1,65
Castilla y León	17,10***	3,35	27,78***	5,76	10,93**	2,20
Cataluña hablan Catalán	-3,46	-0,47	11,09	1,58	13,84*	1,81
Cataluña hablan Castellano	-12,03*	-1,91	-6,13	-1,14	3,46	0,54
Cataluña hablan otra lengua	-13,78	-1,20	12,25	0,97	-0,05	0,00
Galicia	17,72***	3,56	18,45***	3,43	24,25***	5,08
La Rioja	16,33***	3,48	39,58***	9,52	21,28***	4,23
Navarra que hablan Eusquera	-28,34***	-4,06	12,38	1,39	-0,15	-0,02
Navarra que hablan castellano	11,17**	2,30	26,81***	5,41	8,51	1,54
Navarra que hablan otra lengua	28,76	1,61	45,66***	2,98	8,64	0,48
País Vasco que hablan Eusquera	-20,70**	-2,59	3,40	0,46	2,98	0,34
País Vasco que hablan castellano	-9,90*	-1,95	6,84	1,31	14,44**	2,60
País Vasco que hablan otra lengua	-11,87	-0,76	-12,05	-0,74	5,46	0,26
resto	-3,48	-0,81	-1,40	-0,34	-5,94	-1,15
DOTACIONES INICIALES						
Madre educación alta y padre baja	16,07***	3,70	11,75***	3,01	18,01***	4,40
Madre educación baja y padre alta	6,50	1,51	5,90	1,42	11,76***	2,84
Madre y padre educación alta	13,97***	4,35	11,24***	3,78	17,99***	6,42
Padre de ocupación alta	6,27***	2,75	2,68	1,38	7,39***	3,29
Madre de ocupación alta	5,93**	2,10	4,84*	1,74	-0,74	-0,28
Madre ama de casa	-10,95***	-4,17	-9,59***	-3,61	-12,72***	-5,05
Sexo del alumno (mujer)	-15,87***	-9,57	-21,40***	-13,14	23,08***	13,07
Mes de nacimiento del alumno	-0,64**	-2,50	-0,89***	-2,99	-0,56*	-2,03
País de nacimiento (inmigrante)	-30,11***	-6,32	-23,35***	-5,35	-19,00***	-3,67
Cultura educativa media	8,24***	3,31	7,24***	2,90	10,27***	4,35
Cultura educativa alta	17,82***	5,40	18,81***	6,16	15,81***	5,07
VARIABLES INDIVIDUALES						
Actitud media	21,07***	9,06	14,68***	6,31	9,43***	3,68
Actitud alta	53,79***	21,28	40,28***	15,79	28,15***	10,92
Ha repetido un curso	-63,66***	-24,01	-67,50***	-25,70	-60,56***	-21,51
Ha repetido dos o más cursos	-111,31***	-27,14	-122,06***	-26,29	-104,86***	-20,42

VARIABLE	CIENCIAS		MATEMÁTICAS		LECTURA	
	COEF.	t	COEF.	T	COEF.	t
VARIABLES DE COLEGIO						
Mayoría de chicas en el colegio	6,07*	1,75	6,87**	2,06	7,66**	2,27
Colegio privado	-1,78	-0,17	-4,71	-0,46	-10,47	-0,97
Colegio concertado	-10,18	-1,23	-7,83	-1,26	11,18	0,92
Mucha presión en colegios públicos	-5,35	-0,28	9,39	0,42	7,41	0,23
Mucha presión en C. concertados	6,96	0,98	8,03	1,44	4,02	0,55
Mucha presión en colegios privados	18,08**	2,02	6,05	0,54	16,35	1,42
Nivel educativo de los padres alto	18,51***	4,69	12,62***	3,38	11,62**	2,55

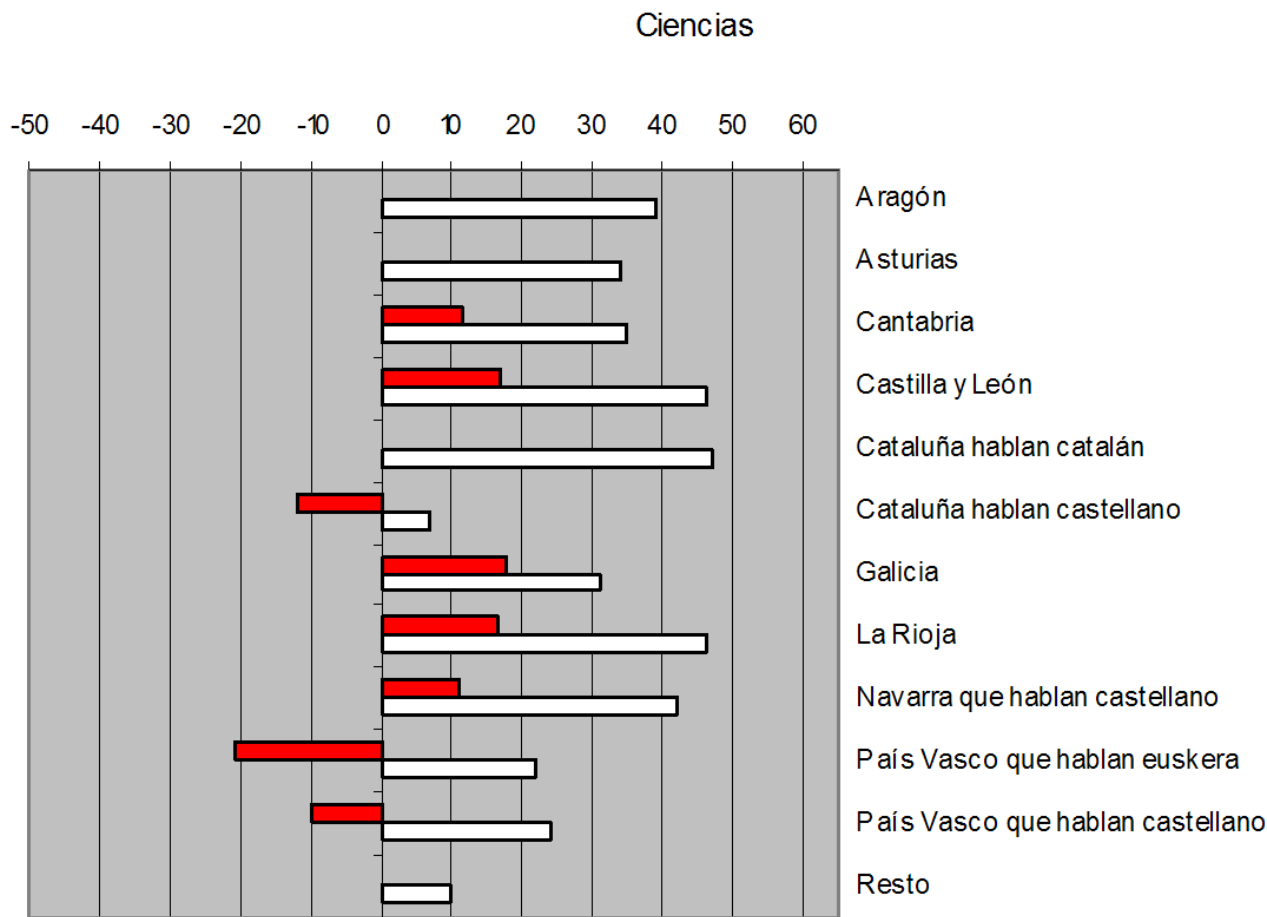
Notas: *, ** y *** significa que el coeficiente es significativo al 10%, al 5% o al 1%, respectivamente. El R² medio es 0,44 para ciencias; 0,44 para matemáticas; y 0,40 para lectura y el número de observaciones es 15134 para las tres regresiones. Se ha controlado por otras variables de colegio aunque no se muestran los resultados al no ser significativas.

Tabla A5. Resultados del Análisis de Regresión. Modelo España. Efectos fijos.

VARIABLE	CIENCIAS		MATEMÁTICAS		LECTURA	
	COEF.	t	COEF.	T	COEF.	t
DOTACIONES INICIALES						
Madre educación alta y padre baja	15,56***	3,78	11,49***	3,08	17,27***	4,61
Madre educación baja y padre alta	7,93*	1,84	6,57*	1,73	12,01***	3,18
Madre y padre educación alta	13,57***	4,37	10,64***	4,00	17,60***	7,23
Padre de ocupación alta	6,16***	3,01	3,11*	1,73	7,00***	3,39
Madre de ocupación alta	5,36**	2,06	3,16	1,32	-0,02	-0,01
Madre ama de casa	-10,94***	-4,52	-8,89***	-3,72	-13,52***	-5,42
Sexo del alumno (mujer)	-16,85***	-11,15	-22,79***	-15,51	21,75***	13,69
Mes de nacimiento del alumno	-0,71**	-2,60	-0,81***	-2,99	-0,47*	-1,72
País de nacimiento (inmigrante)	-28,75***	-5,86	-21,57***	-4,34	-17,03***	-3,15
Cultura educativa media	7,11***	3,26	6,76***	2,93	8,38***	3,85
Cultura educativa alta	16,36***	5,64	17,24***	6,54	13,53***	5,00
VARIABLES INDIVIDUALES						
Actitud media	20,49***	9,93	14,52***	6,73	9,44***	3,93
Actitud alta	52,05***	22,91	38,47***	16,58	25,45***	11,33
Ha repetido un curso	-63,71***	-29,64	-67,80***	-29,75	-61,05***	-25,54
Ha repetido dos o más cursos	-112,50***	-29,46	-121,65***	-26,95	-106,80***	-21,96

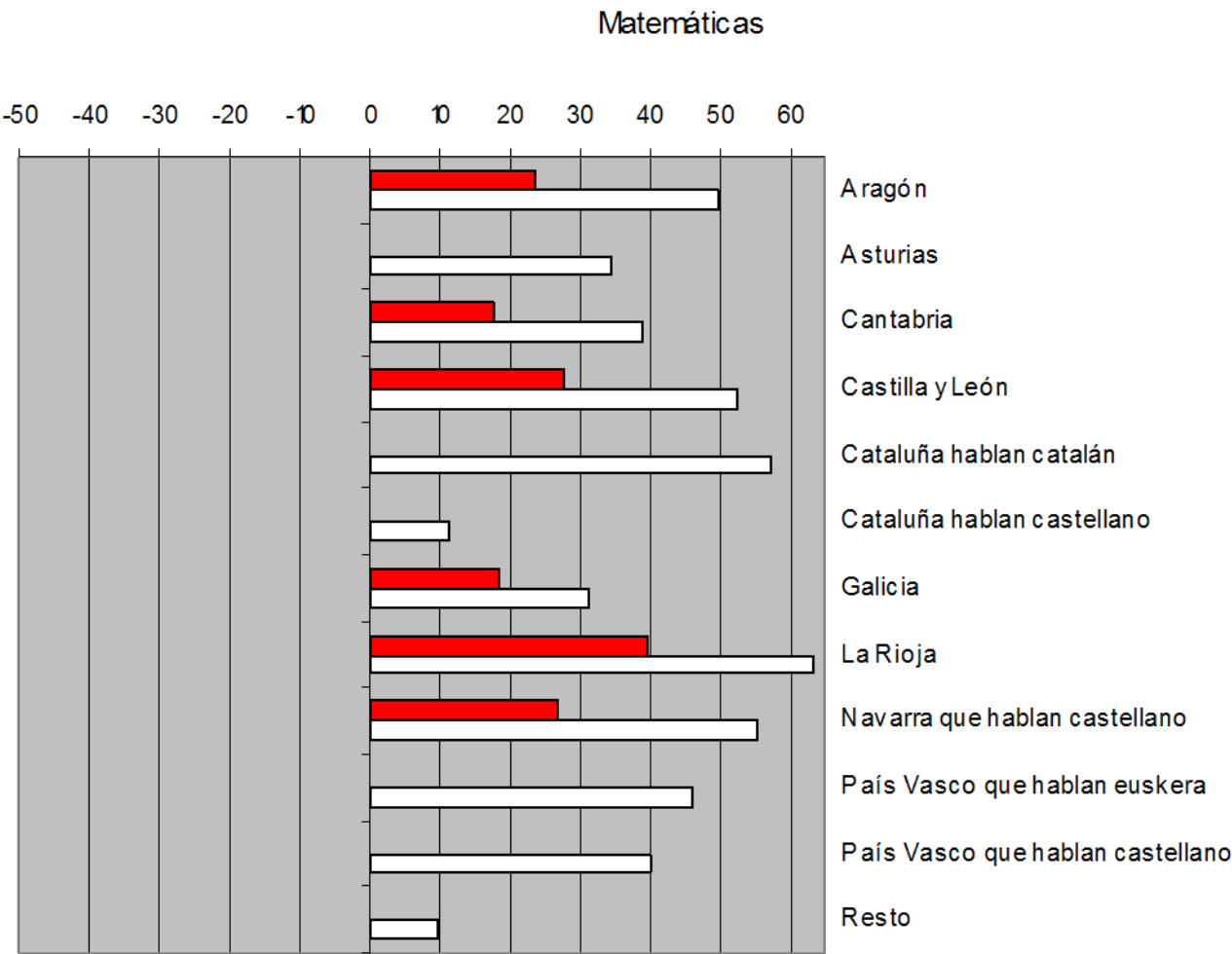
Notas: *, ** y *** significa que el coeficiente es significativo al 10%, al 5% o al 1%, respectivamente. El número de observaciones es 15134 para las tres regresiones.

Gráfico A7. Diferencias regionales observadas y predichas. Ciencias



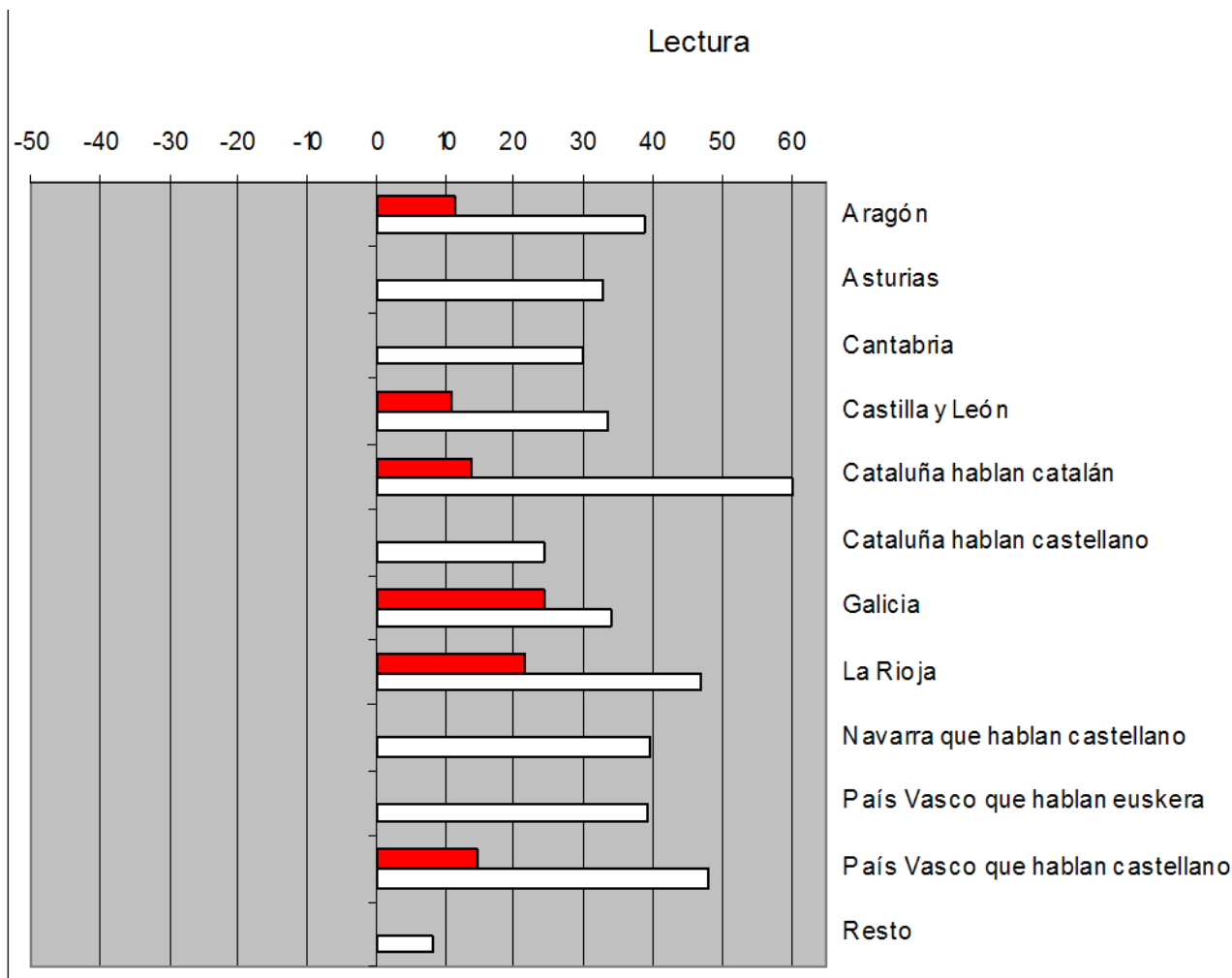
Nota: Las barras blancas son las diferencias observadas, las barras rojas son las diferencias predichas que resultan significativas. Las regiones que no tienen barra roja no son significativamente distintas a Andalucía según los resultados del análisis multivariante.

Gráfico A8. Diferencias regionales observadas y predichas. Matemáticas



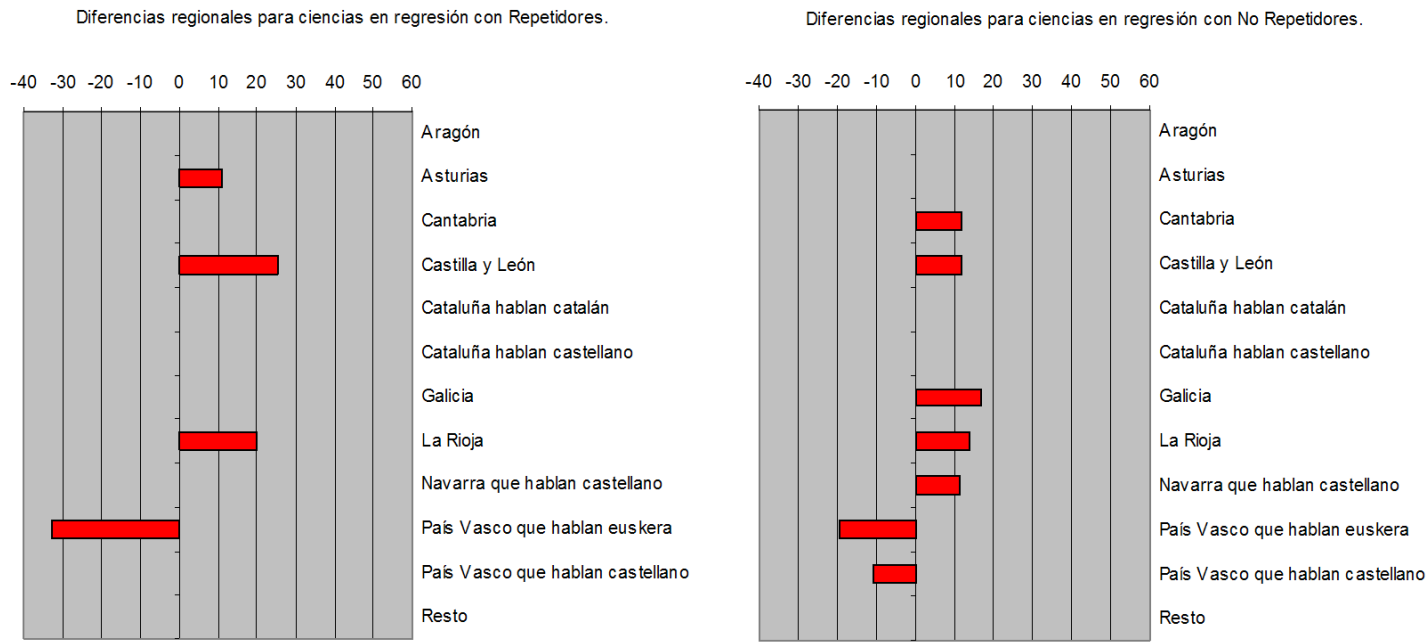
Nota: Las barras blancas son las diferencias observadas, las barras rojas son las diferencias predichas que resultan significativas. Las regiones que no tienen barra roja no son significativamente distintas a Andalucía según los resultados del análisis multivariante.

Gráfico Ag. Diferencias regionales observadas y predichas. Lectura



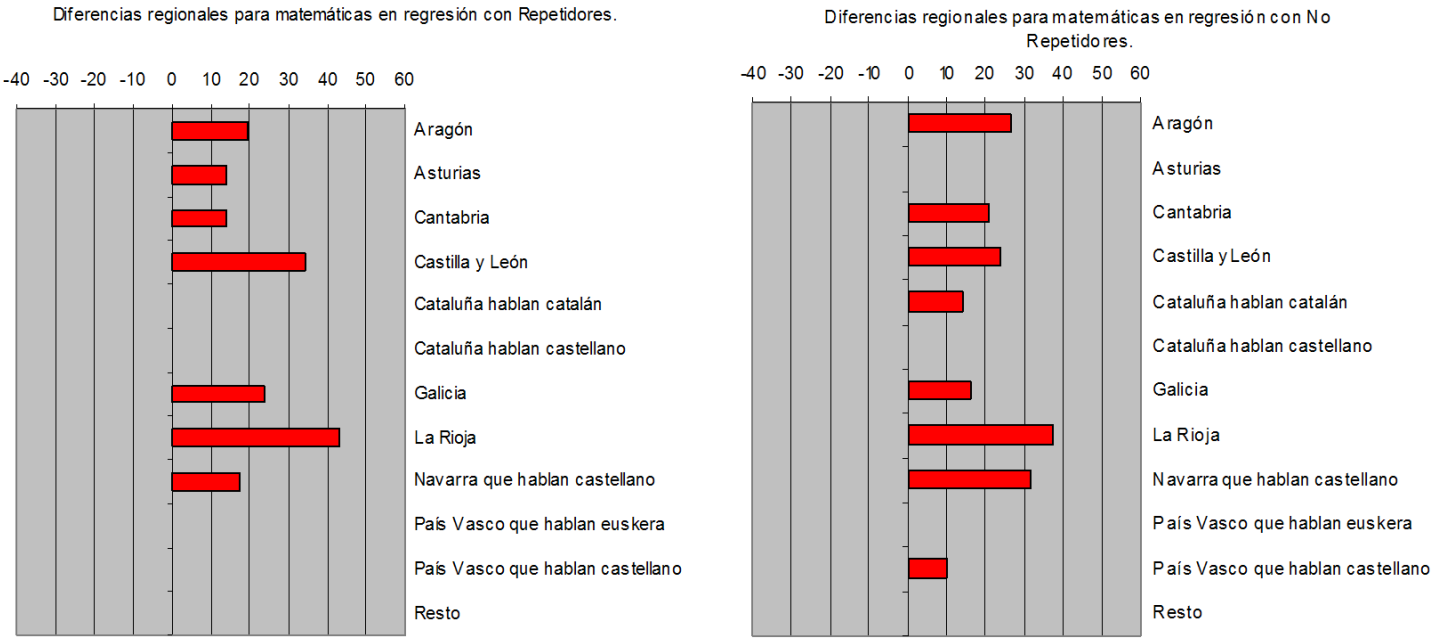
Nota: Las barras blancas son las diferencias observadas, las barras rojas son las diferencias predichas que resultan significativas. Las regiones que no tienen barra roja no son significativamente distintas a Andalucía según los resultados del análisis multivariante.

Gráfico A10. Diferencias regionales distinguiendo entre Repetidores y no Repetidores. Ciencias



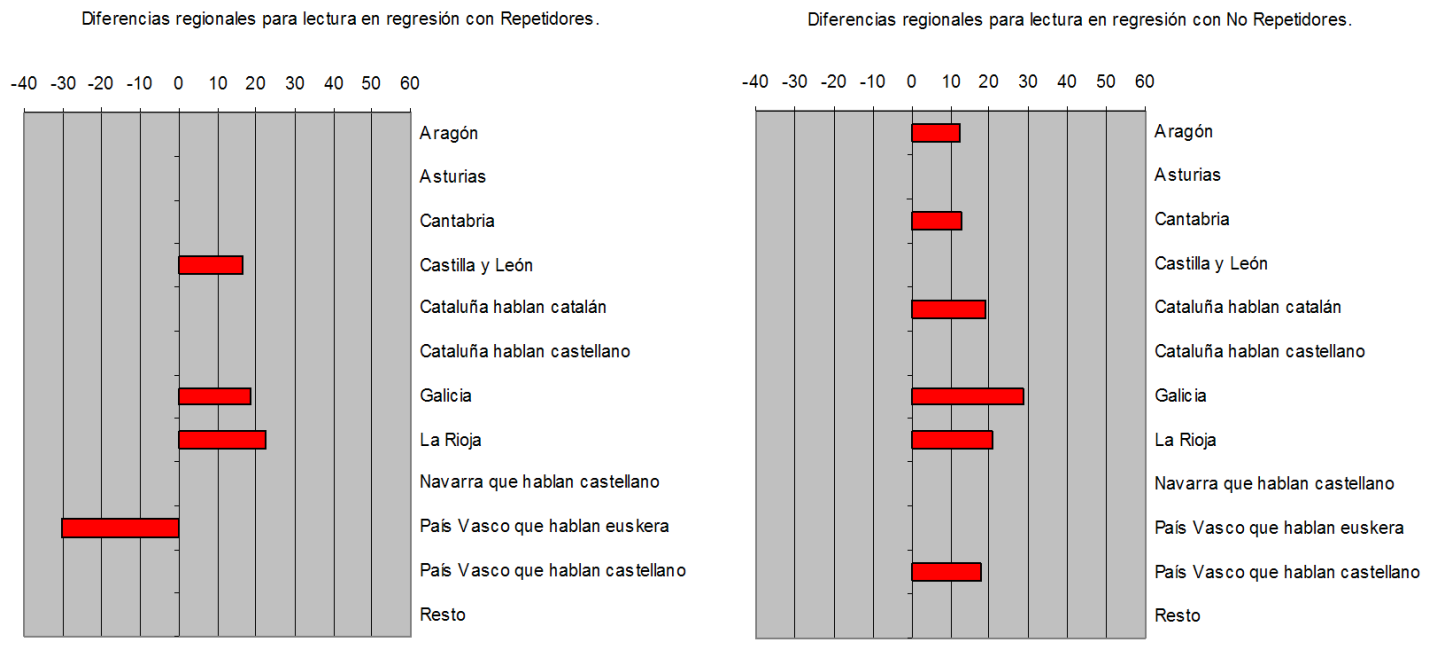
Nota: Las barras rojas son las diferencias predichas que resultan significativas. Las regiones que no tienen barra roja no son significativamente distintas a Andalucía según los resultados del análisis multivariante.

Gráfico A11. Diferencias regionales distinguiendo entre Repetidores y no Repetidores. Matemáticas



Nota: Las barras rojas son las diferencias predichas que resultan significativas. Las regiones que no tienen barra roja no son significativamente distintas a Andalucía según los resultados del análisis multivariante.

Gráfico A12. Diferencias regionales distinguiendo entre Repetidores y no Repetidores. Lectura



Nota: Las barras rojas son las diferencias predichas que resultan significativas. Las regiones que no tienen barra roja no son significativamente distintas a Andalucía según los resultados del análisis multivariante.

Tabla A6. Resultado de la estimación del modelo de regresión multivariante. Modelo Regional. Ciencias

Resultado en ciencias	Andalucía		Aragón		Asturias		Cantabria		Castilla y León		Cataluña	
R2	0.45		0.42		0.37		0.47		0.40		0.32	
nº observaciones	1300		1334		1332		1306		1300		1351	
Variable	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)
Madre educación alta y padre baja	23,65***	3,20	12,07	1,42	19,53**	2,45	6,57	0,62	-0,09	-0,01	7,75	1,02
Madre educación baja y padre alta	17,37***	2,71	8,12	0,75	11,07	0,74	-1,77	-0,17	2,69	0,25	-0,80	-0,09
Madre y padre educación alta	20,60***	4,11	8,05	1,14	14,16**	2,07	6,03	1,01	3,64	0,55	5,59	0,97
Padre de ocupación alta	5,91	1,22	13,43***	3,17	14,83***	2,83	4,95	1,17	7,12	1,24	3,56	0,88
Madre de ocupación alta	5,92	0,90	13,02***	3,08	4,70	0,97	11,96***	2,60	4,15	0,83	3,93	0,81
Madre ama de casa	-19,54***	-4,29	-8,22	-1,61	-10,01**	-2,39	-8,35*	-1,86	-1,69	-0,35	-14,16**	-2,14
Cultura educativa media	5,82	1,40	3,45	0,58	7,77	1,34	8,31**	2,17	17,55***	3,60	8,22*	1,72
Cultura educativa alta	15,98***	2,57	14,97**	2,00	18,92***	3,44	21,99***	4,95	25,74***	4,93	16,35***	3,01
Mes de nacimiento del alumno inmigrante	-0,88	-1,42	-1,76***	-2,83	-0,24	-0,46	-1,44**	-2,53	-0,11	-0,20	-0,76	-1,23
Sexo del alumno (mujer)	-38,75	-1,56	-29,32***	-2,69	1,57	0,10	-30,47**	-2,55	47,81***	-3,42	-37,95***	-4,04
Actitud media	-22,29***	-6,63	-7,34*	-1,76	-18,36***	-4,87	-19,51***	-5,75	-14,33***	-4,44	-20,94***	-5,18
Actitud alta	23,42***	5,48	32,68***	7,41	22,17***	4,19	14,32***	3,15	21,09***	4,32	15,06***	2,81
Ha repetido un curso	46,66***	7,24	61,95***	11,95	50,74***	7,60	50,52***	12,38	56,87***	11,07	59,86***	11,58
Ha repetido dos o más cursos	-68,46***	-15,26	-56,07***	-11,53	-59,01***	-12,64	-70,71***	-18,19	-56,42***	-15,45	-58,20***	-10,55
Constante	-114,40***	-14,94	-119,96***	-13,96	-102,16***	-9,59	-111,53***	-12,94	-95,70***	-9,82	-113,54***	-9,33
	491,03***	83,12	500,73***	78,20	492,22***	83,87	521,46***	89,98	501,27***	84,83	495,39***	85,44

Resultado en ciencias	Galicia		La Rioja		Navarra		País Vasco		Resto	
R2	0.42		0.42		0.44		0.34		0.35	
nº observaciones	1293		1183		1372		3237		1816	
Variable	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)
Madre educación alta y padre baja	4,21	0,45	5,37	0,50	5,81	0,57	6,91	0,93	13,98*	1,72
Madre educación baja y padre alta	2,99	0,32	13,87	1,10	-7,23	-0,61	8,53	1,26	3,46	0,42
Madre y padre educación alta	7,16	1,04	6,58	0,83	-4,29	-0,51	13,50**	2,39	11,93**	2,05
Padre de ocupación alta	4,36	0,94	6,30	1,21	17,32***	3,19	2,76	0,86	6,18	1,38
Madre de ocupación alta	0,70	0,14	12,64**	2,18	3,67	0,87	3,86	1,19	6,58	1,21
Madre ama de casa	-8,82	-1,50	-7,40	-1,27	-8,39	-1,57	-10,90**	-2,57	-6,02	-1,35
Cultura educativa media	4,14	0,91	4,79	0,84	4,97	0,81	8,82**	2,17	7,60	1,59
Cultura educativa alta	8,56	1,46	13,45**	2,12	15,48**	2,19	14,37***	3,70	17,37***	3,06
Mes de nacimiento del alumno inmigrante	-1,20**	-2,00	-1,03*	-1,73	-1,12**	-2,19	-0,72**	-2,08	-0,55	-1,21
Sexo del alumno (mujer)	-17,87	-1,17	-29,10***	-3,08	-19,40*	-1,92	-16,63	-1,57	-24,52***	-3,82
Actitud media	-23,27***	-6,75	-12,48**	-2,48	-15,89***	-3,86	-10,15***	-4,34	-12,75***	-4,02
Actitud alta	21,74***	4,37	24,41***	3,94	27,95***	5,74	26,03***	9,19	19,76***	4,63
Ha repetido un curso	56,59***	11,46	56,24***	7,85	68,95***	11,51	62,93***	17,11	48,64***	10,87
Ha repetido dos o más cursos	-72,85***	-16,60	-67,31***	-10,12	-74,55***	-13,44	-62,87***	-14,69	-62,24***	-14,43
Constante	-129,72***	-18,33	-112,78***	-9,56	-124,39***	-11,62	-113,96***	-14,42	-110,17***	-15,46
	525,60***	74,79	518,31***	76,26	512,74***	85,10	475,16***	92,61	486,64***	83,12

Notas: *, ** y *** significa que el coeficiente es significativo al 10%, al 5% o al 1%, respectivamente. (1) es el coeficiente estimado para cada variable y (2) el estadístico t de student.

Tabla A7. Resultado de la estimación del modelo de regresión multivariante. Modelo Regional. Matemáticas

Resultado en matemáticas	Andalucía		Aragón		Asturias		Cantabria		Castilla y León		Cataluña	
R2	0.44		0.40		0.39		0.48		0.40		0.34	
nº observaciones	1300		1334		1332		1306		1300		1351	
Variable	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)
Madre educación alta y padre baja	16,59**	2,29	10,80	0,84	13,40	1,56	4,13	0,41	4,71	0,43	6,03	0,78
Madre educación baja y padre alta	12,91*	1,95	11,11	0,74	10,57	0,84	8,31	0,72	6,04	0,72	-7,42	-0,74
Madre y padre educación alta	12,79***	2,75	7,73	0,84	10,25	1,41	17,11	2,47	8,90	1,31	6,69	1,07
Padre de ocupación alta	3,46	0,66	7,13	1,40	10,33**	2,15	-1,58	-0,42	-0,24	-0,05	4,31	1,12
Madre de ocupación alta	6,68	1,05	8,61	1,58	2,93	0,63	19,65***	4,12	8,94*	1,81	-0,84	-0,19
Madre ama de casa	-17,08***	-4,05	-7,62	-1,26	-10,29**	-2,44	1,65	0,37	1,25	0,20	-10,80*	-1,79
Cultura educativa media	4,17	0,98	10,18	1,44	7,19	1,54	5,13	1,16	10,51**	2,56	3,86	0,78
Cultura educativa alta	14,36**	2,65	22,04***	2,74	14,23***	2,58	16,30***	3,41	18,84***	3,28	12,70**	2,39
Mes de nacimiento del alumno	-0,45	-0,69	-0,95	-1,43	-0,68	-1,11	-1,15**	-2,03	0,96*	1,90	-0,97*	-1,70
inmigrante	-23,45	-1,08	-21,38*	-1,87	-5,70	-0,44	-13,37	-1,25	-66,81***	-5,06	-27,66***	-2,70
Sexo del alumno (mujer)	-25,67***	-7,13	-31,27***	-6,54	-23,79***	-6,88	-23,58***	-6,15	-21,23***	-7,08	-24,35***	-6,67
Actitud media	14,74***	3,09	32,98***	6,58	17,24***	3,05	14,88***	3,04	20,66***	4,03	13,14***	2,86
Actitud alta	28,30***	5,24	57,96***	8,51	46,98***	6,28	41,07***	9,22	59,49***	10,56	48,16***	8,66
Ha repetido un curso	-70,62***	-14,99	-70,77***	-14,26	-59,85***	-12,20	-74,59***	-21,14	-56,66***	-16,07	-69,35***	-11,25
Ha repetido dos o más cursos	-124,97***	-14,13	-142,31***	-13,86	-126,23***	-13,51	-139,67***	-16,70	-104,12***	-11,08	-123,90***	-10,09
Constante	495,02***	100,17	511,92***	64,21	498,41***	75,01	508,46***	78,49	496,37***	60,82	504,75***	89,62

Resultado en matemáticas	Galicia		La Rioja		Navarra		País Vasco		Resto	
R2	0.42		0.41		0.45		0.34		0.35	
nº observaciones	1293		1183		1372		3237		1816	
Variable	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)
Madre educación alta y padre baja	10,58	1,17	4,60	0,39	4,83	0,45	2,48	0,37	10,52	1,46
Madre educación baja y padre alta	-9,15	-1,10	22,03*	1,80	2,85	0,22	2,09	0,33	7,50	1,09
Madre y padre educación alta	4,06	0,66	8,80	0,97	3,91	0,39	6,48	1,28	10,93**	2,29
Padre de ocupación alta	3,48	0,98	1,72	0,34	8,54*	1,66	-0,84	-0,32	2,77	0,65
Madre de ocupación alta	-0,76	-0,15	5,29	1,04	5,44	1,25	7,86**	2,43	2,56	0,47
Madre ama de casa	-8,63	-1,63	-8,04	-1,58	1,12	0,20	-2,07	-0,54	-5,02	-1,23
Cultura educativa media	3,12	0,65	2,20	0,30	2,53	0,41	8,32*	2,01	10,11**	2,08
Cultura educativa alta	8,10	1,41	15,87**	2,20	16,32**	2,20	12,05***	2,79	22,14***	3,98
Mes de nacimiento del alumno	-1,03**	-2,09	-0,85	-1,59	-0,04	-0,09	-0,44	-1,16	-1,15***	-2,69
inmigrante	-13,07	-1,03	-38,33***	-3,36	-21,06**	-2,04	-22,69*	-1,98	-17,34**	-2,39
Sexo del alumno (mujer)	-25,99***	-7,59	-23,34***	-4,60	-20,98***	-4,71	-15,83***	-6,32	-20,34***	-6,43
Actitud media	19,69***	4,67	22,38***	3,89	21,04***	3,82	20,65***	7,43	11,81***	2,69
Actitud alta	51,02***	13,75	47,65***	7,16	61,43***	9,80	51,74***	15,12	32,54***	7,28
Ha repetido un curso	-68,82***	-14,56	-68,70***	-11,44	-82,66***	-15,38	-72,04***	-17,42	-65,69***	-14,93
Ha repetido dos o más cursos	-121,94***	-20,45	-115,05***	-9,13	-138,52***	-10,35	-123,52***	-15,64	-117,73***	-13,05
Constante	518,89***	82,10	535,14***	66,09	514,53***	76,67	496,22***	113,03	491,36***	98,02

Notas: *, ** y *** significa que el coeficiente es significativo al 10%, al 5% o al 1%, respectivamente. (1) es el coeficiente estimado para cada variable y (2) el estadístico t de student.

Tabla A8. Resultado de la estimación del modelo de regresión multivariante. Modelo Regional. Lectura

Resultado en lectura	Andalucía		Aragón		Asturias		Cantabria		Castilla y León		Cataluña	
R2	0.38		0.39		0.35		0.39		0.33		0.30	
nº observaciones	1300		1334		1332		1306		1300		1351	
Variable	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)
Madre educación alta y padre baja	21,06***	4,02	-6,49	-0,99	7,27	0,60	-10,00	-0,90	7,34	1,02	-0,96	-0,10
Madre educación baja y padre alta	14,17**	2,18	11,25	1,33	30,06*	1,96	11,76	1,27	-6,47	-0,89	15,05	1,71
Madre y padre educación alta	17,56***	3,60	1,48	0,27	27,17***	2,85	6,50	0,71	-5,66	-0,97	20,56***	4,12
Padre de ocupación alta	3,36	0,68	2,56	0,59	6,24	1,47	3,09	0,60	3,48	0,77	0,39	0,11
Madre de ocupación alta	3,78	0,64	8,29	1,61	3,93	0,80	7,80	1,51	-2,80	-0,66	5,78	1,19
Madre ama de casa	-15,17***	-3,26	-10,88**	-2,32	-18,20***	-4,16	2,53	0,49	-7,15	-1,55	-20,39***	-2,90
Cultura educativa media	5,82	1,17	7,65	1,28	10,84	1,61	3,85	0,98	17,69***	3,61	8,00	1,47
Cultura educativa alta	11,72	1,74	19,56***	3,47	19,45***	3,24	12,55***	2,80	20,33***	3,66	14,36**	2,35
Mes de nacimiento del alumno	-0,15	-0,23	-1,00*	-1,66	-1,16**	-2,11	-0,93	-1,50	0,83	1,56	-1,35**	-2,29
inmigrante	-6,71	-0,30	-11,51	-0,91	13,88	0,79	-15,85	-1,52	-46,77***	-2,76	-30,48***	-3,03
Sexo del alumno (mujer)	12,51***	3,43	34,69***	8,88	16,98***	5,11	22,99***	6,24	15,73***	4,91	13,68***	2,98
Actitud media	18,51***	4,07	21,41***	5,33	7,93	1,34	11,39**	2,09	12,69***	2,97	2,75	0,53
Actitud alta	30,80***	4,83	36,90***	7,40	19,85***	2,90	36,48***	5,91	37,38***	9,52	27,21***	5,44
Ha repetido un curso	-59,54***	-14,64	-52,84***	-13,29	-58,92***	-11,47	-64,35***	-13,73	-48,32***	-13,81	-62,45***	-8,99
Ha repetido dos o más cursos	-110,07***	-13,34	-119,49***	-12,67	-110,86***	-11,64	-114,67***	-10,28	-98,67***	-11,99	-112,77***	-7,52
Constante	446,37***	88,50	465,34***	100,21	458,00***	57,86	469,45***	47,54	462,90***	77,02	472,49***	101,96

Resultado en LECTURA	Galicia		La Rioja		Navarra		País Vasco		Resto	
R2	0.42		0.36		0.36		0.31		0.34	
nº observaciones	1293		1183		1372		3237		1816	
Variable	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)	coef. (1)	t (2)
Madre educación alta y padre baja	-9,80	-0,97	21,35*	1,84	4,73	0,53	6,22	0,96	25,77***	3,55
Madre educación baja y padre alta	0,44	0,06	23,28**	2,05	11,25	1,05	8,05	1,18	12,64*	1,82
Madre y padre educación alta	-6,94	-0,97	21,76**	2,40	5,78	0,62	14,07***	3,01	21,99***	5,04
Padre de ocupación alta	0,01	0,00	8,57*	1,89	9,73**	2,52	3,40	1,26	14,08***	3,02
Madre de ocupación alta	6,80	1,52	1,23	0,21	-2,48	-0,48	1,50	0,48	-6,65	-1,37
Madre ama de casa	-18,78***	-3,29	-8,27	-1,45	5,64	1,05	-5,40	-1,36	-13,34***	-2,77
Cultura educativa media	2,67	0,53	11,46	1,61	4,10	0,65	6,16	1,42	10,16**	2,26
Cultura educativa alta	6,48	1,03	20,31***	3,07	13,57*	1,92	9,98**	2,31	14,43***	2,86
Mes de nacimiento del alumno	-0,87	-1,56	-0,40	-0,57	-0,03	-0,06	-0,40	-0,99	-0,35	-0,77
inmigrante	-4,91	-0,29	-24,81**	-2,45	-10,50	-1,13	11,33	0,96	-13,01*	-1,72
Sexo del alumno (mujer)	23,07***	6,05	25,12***	5,35	25,69***	5,86	22,28***	8,31	29,51***	9,53
Actitud media	11,65**	2,21	10,70*	1,83	17,75***	3,47	20,55***	7,59	4,91	1,13
Actitud alta	37,94***	8,51	20,36***	2,94	43,69***	9,66	45,04***	12,98	16,00***	4,16
Ha repetido un curso	-70,18***	-16,35	-58,94***	-9,26	-63,88***	-11,26	-64,34***	-17,13	-61,69***	-13,15
Ha repetido dos o más cursos	-131,37***	-20,72	-92,32***	-7,67	-95,54***	-7,02	-120,02***	-16,42	-98,99***	-10,67
Constante	500,96***	72,54	464,37***	50,33	455,97***	69,34	461,01***	93,72	440,74***	94,26

Notas: *, ** y *** significa que el coeficiente es significativo al 10%, al 5% o al 1%, respectivamente. (1) es el coeficiente estimado para cada variable y (2) el estadístico t de student.

Tabla Ag. Predicción de la nota para las distintas categorías de las variables

Variable	Categoría	CIENCIAS			MATEMÁTICAS			LECTURA		
		(1)	(2)	(2)-(1)	(1)	(2)	(2)-(1)	(1)	(2)	(2)-(1)
Educación padres	a) Madre y Padre de educación baja	484,37	508,69	24,31	478,02	503,71	25,69	455,77	478,45	22,69
	b) Madre educación alta y padre baja	508,02	516,9	8,87	494,61	510,92	16,31	476,83	482,99	6,17
	c) Madre educación baja y padre alta	501,74	512,78	11,04	490,93	509,1	18,17	469,94	490,18	20,25
	d) Madre y padre educación alta	504,97	515,92	10,95	490,81	512,2	21,38	473,33	489,12	15,8
	Diferencias entre categorías (d-a)	20,6	7,24	-	12,79	8,49	-	17,56	10,67	-
Ocupación padre	a) Padre con ocupación baja	500,72	511,89	11,18	488,55	509,67	21,11	470,4	485,62	15,22
	b) Padre con ocupación alta	506,63	519,97	13,35	492,01	513,23	21,22	473,76	490,78	17,02
	Diferencias entre categorías (b-a)	5,91	8,08	-	3,46	3,56	-	3,36	5,16	-
Ocupación madre	a) Madre de ocupación alta	511,57	521,38	9,81	498,35	516,15	17,8	477,8	491,43	13,64
	b) Madre de ocupación baja	505,65	514,86	9,21	491,67	510,18	18,52	474,02	489,09	15,08
	c) Madre ama de casa	486,11	506,46	20,35	474,59	505,34	30,75	458,85	479,67	20,82
	Diferencias entre categorías (a-b)	5,92	6,52	-	6,68	5,97	-	3,78	2,34	-
	Diferencias entre categorías (b-c)	19,54	8,4	-	17,08	4,84	-	15,17	9,42	-
Cultura Educativa	a) Hogar con Cultura educativa baja	494	505,14	11,13	482,13	501,91	19,78	464,57	478,19	13,62
	b) Hogar con cultura educativa media	499,82	512,7	12,88	486,3	508,22	21,92	470,39	486,45	16,06
	c) Hogar con cultura educativa alta	509,98	521,86	11,87	496,49	517,77	21,27	476,29	493,29	17
	Diferencias entre categorías (a-c)	15,98	16,72	-	14,36	15,86	-	11,72	15,1	-
Inmigrante	a) Nativo	504,75	516,2	11,45	490,95	512,16	21,21	472,03	488,25	16,22
	b) Inmigrante	466	491,05	25,05	467,5	487,42	19,92	465,32	474,99	9,67
	Diferencias entre categorías (a-b)	38,75	25,15	-	23,45	24,74	-	6,71	13,26	-
Género	a) Hombre	514,43	523,01	8,58	503,01	522,84	19,83	465,37	475,97	10,61
	b) Mujer	492,14	507,51	15,37	477,34	499,77	22,43	477,88	498,95	21,07
	Diferencias entre categorías (a-b)	22,29	15,5	-	25,67	23,07	-	-12,51	-22,97	-
Actitud	a) Actitud hacia el estudio baja	478,95	488,29	9,34	475,06	487,84	12,78	454,44	472,86	18,41
	b) Actitud hacia el estudio media	502,37	510,81	8,44	489,8	507,29	17,48	472,95	485,03	12,08
	c) Actitud hacia el estudio alta	525,61	545,62	20,01	503,36	537,64	34,28	485,24	504,94	19,7
	Diferencias entre categorías (a-c)	46,66	57,33	-	28,3	49,8	-	30,8	32,09	-
Repite	a) No es repetidor	524,79	535,72	10,93	512,66	533,42	20,76	491,12	507,28	16,16
	b) Ha repetido un curso	456,33	471,7	15,37	442,04	464,51	22,47	431,58	446,69	15,11
	c) Ha repetido al menos dos curso	410,39	422,33	11,94	387,69	408,12	20,43	381,05	397,81	16,76
	Diferencias entre categorías (a-b)	68,46	64,02	-	70,62	68,91	-	59,54	60,59	-
	Diferencias entre categorías (b-c)	45,94	49,37	-	54,35	56,39	-	50,53	48,88	-

NOTAS: (1) nota predicha para Andalucía. (2) Nota predicha para el resto de CCAA menos Andalucía. (2)-(1) Diferencia entre la predicción para Andalucía y para el resto de CCAA (diferencias en rendimientos).

Gráfico A13. Diferencias en rendimientos y características de Andalucía con el resto de regiones. Ciencias

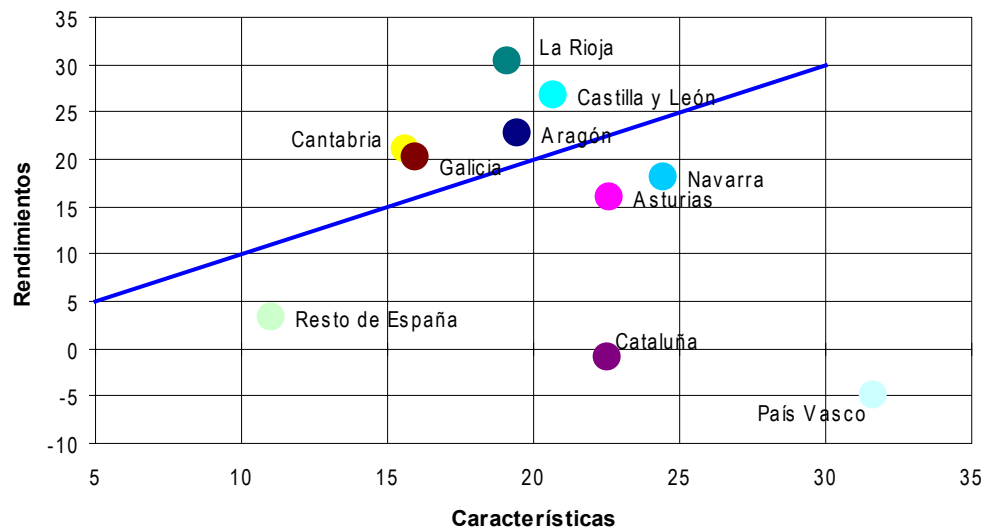


Gráfico A14. Diferencias en rendimientos y características de Andalucía con el resto de regiones. Matemáticas

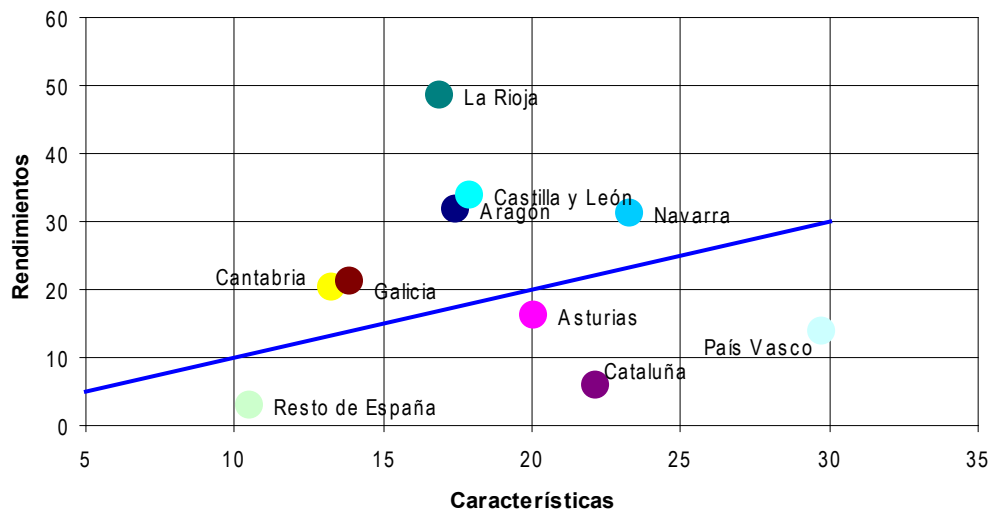


Gráfico A15. Diferencias en rendimientos y características de Andalucía con el resto de regiones. Lectura

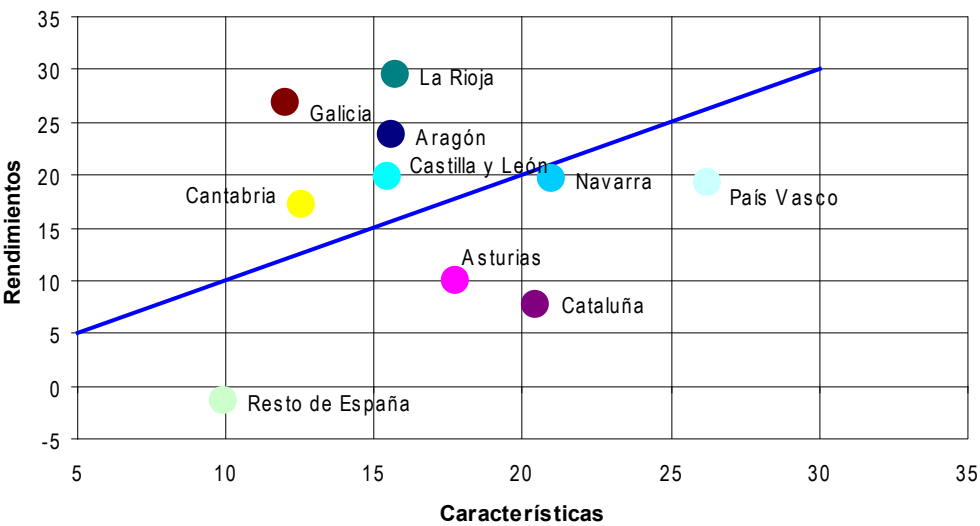


Tabla A10. Nota observada para cada región y nota predicha para Andalucía con las características de cada región. Ciencias

Región	Ciencias		
	Nota observada	Nota predicha para Andalucía	% Características
Aragón	518,2	498,44	49,59%
Asturias	514,73	501,63	63,34%
Cantabria	513,69	494,61	45,00%
Castilla y León	521,68	499,68	48,45%
Cataluña	495,63	501,53	135,48%
Galicia	511,86	494,97	48,60%
La Rioja	524,73	498,09	41,75%
Navarra	517,54	503,47	63,49%
País Vasco	501,66	510,67	139,76%
Resto de España	491,7	490	86,61%
Andalucía	479	479	-

Tabla A11. Nota observada para cada región y nota predicha para Andalucía con las características de cada región. Matemáticas

Región	Matemáticas		
	Nota observada	Nota predicha para Andalucía	%
Aragón	517,4	485,45	35,32%
Asturias	502,88	488,1	57,63%
Cantabria	504,28	481,28	36,60%
Castilla y León	518,26	485,91	35,63%
Cataluña	492,54	490,16	90,30%
Galicia	500,68	481,89	42,50%
La Rioja	531,12	484,9	26,77%
Navarra	521,12	491,28	43,83%
País Vasco	508,18	497,73	73,99%
Resto de España	480,5	478,54	84,32%
Andalucía	468	468	-

Tabla A12. Nota observada para cada región y nota predicha para Andalucía con las características de cada región. Lectura

Región	Lectura		
	Nota observada	Nota predicha para Andalucía	%
Aragón	488,94	467,56	42,12%
Asturias	483,43	469,77	56,54%
Cantabria	478,94	464,54	46,55%
Castilla y León	481,22	467,46	52,91%
Cataluña	481,96	472,42	68,16%
Galicia	488,05	463,98	33,23%
La Rioja	496,54	467,69	35,23%
Navarra	486,09	472,97	61,51%
País Vasco	497,06	478,26	58,28%
Resto de España	459,69	461,94	129,26%
Andalucía	452	452	-

