

La desigualdad educativa, ¿son los programas de refuerzo la solución? Evidencia empírica del impacto a nivel intracentros

Educational inequality, are the reinforcement programs the solution? Empirical evidence of the impact at the intra center level

Adriano Villar-Aldonza* y Borja Gambau-Suelves**

*Departamento de Estructura e Historia Económica y Economía Pública. Facultad de Economía y Empresa. Universidad de Zaragoza (España)

**Instituto de Estudios Fiscales (España)

Resumen

El diferente ritmo de aprendizaje entre el alumnado es, y ha sido, un problema tanto para el profesor como para la comunidad educativa y académica en su conjunto. Entre las políticas públicas dirigidas a corregir esta problemática destacan dos: el agrupamiento del alumnado con base en sus capacidades-tracking- y, más recientemente, los programas de refuerzo extraescolares. En este trabajo se evalúa el impacto que, sobre la desigualdad educativa, tuvo un programa de refuerzo implantado de manera voluntaria en los institutos españoles entre 2005 y 2012: el plan PROA. Como medida de aproximación a la desigualdad educativa se ha empleado la brecha en términos de resultados existente entre los alumnos de un mismo centro, así como el índice de desigualdad de Gini. Los resultados obtenidos, a partir de la técnica de Propensity Score Matching, señalan que la aplicación del programa no redujo las diferencias en términos de puntuación entre el alumnado, de igual modo que no aumentaron las calificaciones medias en lectura en aquellos centros que decidieron participar en el programa.

Palabras clave: PROA; Propensity Score Matching; Desigualdad; Evaluación de impacto.

Correspondencia: Adriano Villar-Aldonza, adrianovillaraldonza@gmail.com, Paseo de la Gran Vía, 2, 50005 Zaragoza (España).

Abstract

The need to cater to the different learning pace of students still remains a frequent problem for both teachers and the academic community. Among the public policies aimed to solve this problem, two stand out: the grouping of students based on their abilities -tracking- and, more recently, the extracurricular reinforcement programs. This paper evaluates the impact that the PROA, a reinforcement program implemented on a voluntary basis by different High Schools in Spain between 2005 and 2012, had on educational inequality. To measure educational inequality, the gap in terms of results between the students of the same school has been used, as well as the Gini inequality index. The results obtained, which are based on the Propensity Score Matching technique, indicate that the implementation of this program did not bridge the gap in terms of marks between students. They also suggest that average reading scores did not increase in those schools that decided to participate in the program.

Keywords: PROA; Propensity Score Matching; Inequality; Impact evaluation

Introducción

Uno de los criterios básicos que conforman un sistema económico ideal, tal y como establecieron los representantes de la Economía del Bienestar en los años 30 del siglo pasado, es el principio de equidad social. La escuela, en su papel como férrea defensora de este precepto, ha tratado de minimizar el negativo impacto que las desigualdades asociadas al nivel de partida del individuo ejercen sobre su posterior logro académico. En este sentido, la Economía de la Educación ha centrado su interés en identificar los determinantes del fracaso escolar. La condición de inmigrante, un bajo nivel de estudios de los progenitores, el status socioeconómico, son, entre otras, causas señaladas por la literatura como responsables de los malos resultados obtenidos por los alumnos (Calero, Choi & Waisgrais, 2010; Gamoran, 2001; González-Rodríguez, Vieira & Vidal, 2019). Sin embargo, y pese a los esfuerzos dirigidos a comprender y remediar este tipo de desventajas, la presencia de “buenos” y “malos” alumnos dentro del aula continúa siendo una problemática común en todo sistema educativo.

El agrupamiento del alumnado, entendido como la agrupación del mismo en diferentes aulas o centros en función de sus capacidades, ha sido una práctica común por considerar que los grupos homogéneos permiten un mayor nivel de adaptabilidad de las técnicas de instrucción e incentivan la participación entre los estudiantes más rezagados, entre otros aspectos. Sin embargo, esta práctica ha demostrado no ser la solución. Slavin (1990), tras una revisión de literatura en la que comparó un total de 29 estudios previos, concluye que el impacto neto del agrupamiento es nulo, y es que, la mejora en los resultados experimentada entre los alumnos que partían de una situación de ventaja inicial, es decir, buenos alumnos, se ve compensada por el empeoramiento entre los malos, ampliando por tanto la brecha entre unos y otros.

Más adelante, en 2006, Hanushek evaluó los efectos del agrupamiento temprano mediante la comparación de los resultados obtenidos en la escuela primaria, recogidos en PIRLS 2001- *Progress in International Reading Literacy Study* (PIRLS, por sus siglas en inglés), y la secundaria, PISA 2003- *Programme for International Student Assessment* (PISA, por sus siglas en inglés). “Para ello empleó una muestra de más de 15 países,

demostrando que aquellos que apostaron por los programas de agrupamiento en lugar de una educación integral (no hay segmentación) vieron aumentada la desigualdad educativa. A diferencia de Slavin (1990), Hanushek (2006) señala que el empeoramiento tiene lugar entre todos los grupos afectados, siendo especialmente perjudicial entre aquellos que presentan mayores dificultades en el aprendizaje” (Villar-Aldonza & Gambau-Suelves, 2018, p.2)

Con relación a la segunda alternativa de política planteada, los programas de refuerzo, los estudios realizados hasta la fecha coinciden en señalar el positivo impacto de este tipo de prácticas sobre los resultados académicos de sus participantes (véase Briggs, 2001; De Paola & Scoppa, 2014). A modo de ejemplo, citaremos algunos de estos trabajos: Lavy y Schlosser (2005) determinaron, para el caso de Israel, que un programa de refuerzo escolar mejora la media, a nivel de centro, entre un 3-4 %; siendo la mejora para los estudiantes asistentes entre un 11-12% en sus resultados; si bien, como concluyen Banerjee, Cole, Duflo y Linden (2007), la mejora a largo plazo sobre los resultados tiende a disminuir. En otro estudio más reciente realizado en Italia, Battistin y Meroni (2016) analizan el impacto de un aumento en las horas de clase en los centros con peores resultados; evidenciándose un impacto positivo y significativo sobre las matemáticas y no significativo sobre lenguaje. En el panorama nacional destaca la publicación de García-Pérez e Hidalgo-Hidalgo (2017), donde se analiza el impacto del Plan de Acompañamiento Escolar (PAE de ahora en adelante). Las conclusiones del estudio señalan que la asistencia a clases de refuerzo contribuye a mejorar las calificaciones entre el alumnado, un 17,4% en términos de desviación estándar, a la par que reduce la probabilidad de estar en el cuartil inferior entre un 3,5% y 6,4%. Con todo, y como señalan los propios autores en su trabajo, la imposibilidad de conocer a partir de la base de datos original qué individuos asistieron al programa de refuerzo hace que sus resultados hayan de ser interpretados con cautela.

Sin embargo, no encontramos a fecha de hoy ningún trabajo que analice, de manera empírica, el efecto de los programas de refuerzo extraescolares sobre la desigualdad. Es por ello que la realización de estudios sobre este tema resulta de gran interés. Dar una respuesta racional a si los programas de refuerzo son la solución a la desigualdad entre buenos y malos alumnos requiere de la realización de estudios cuantitativos.

En este trabajo se ha evaluado el impacto que las clases de refuerzo educativo tuvieron sobre aquellos centros que decidieron participar en un programa de refuerzo implantado en España entre 2005 y 2012, el Plan de Refuerzo y Orientación Académica (PROA de ahora en adelante). En particular, se ha analizado una vertiente del mismo implantada en la secundaria como es el Plan de Acompañamiento Escolar. Para ello se ha hecho uso de la base de datos PISA 2012, ampliada por el Ministerio de Educación, Cultura y Deporte (MECD a partir de ahora) del gobierno español añadiéndole el número de años que los centros habían participado en dicho programa.

El trabajo realizado constituye una novedosa aportación, ya que se enfoca en la evaluación del impacto de un programa de refuerzo en términos de desigualdad frente a la aproximación más común centrada en la mejora de los resultados. Para ello se ha seguido el trabajo de Hanushek (2006), en el que evaluó, a nivel de país, el impacto que la agrupación del alumnado tuvo sobre la desigualdad. Como aproximación de

la desigualdad empleó tres medidas: la desviación estándar de los resultados de las pruebas dentro de cada país; la diferencia entre la puntuación del estudiante situado en el percentil 75 y el situado en el percentil 25; y la diferencia de rendimiento entre el percentil 95 y el 5. En nuestro caso, los datos disponibles, han permitido realizar este análisis a nivel de centro, por lo que la desigualdad queda definida como las diferencias entre las puntuaciones obtenidas por los estudiantes de un mismo centro (percentil 75- percentil 25 y percentil 90-percentil 10)

Los resultados obtenidos, a partir de la técnica Propensity Score Matching (Rosenbaum & Rubin, 1983), señalan que, en términos generales, la aplicación del programa no tuvo efecto a nivel de centro escolar. Aquellos centros que participaron en el programa no vieron reducidas las diferencias en la puntuación entre los mejores y peores alumnos con respecto a los centros en donde no se aplicó.

El trabajo se estructura de la forma que sigue. En el siguiente apartado se presenta el potencial de los programas de refuerzo y su auge entre los países del entorno europeo, prestando especial atención al caso español. En el tercer epígrafe se presenta la base de datos, PISA 2012, las variables del estudio y la metodología empleada. A continuación, se presentan los resultados y, finalmente, la sección de conclusiones.

El potencial de los programas de refuerzo

En 2009 el Consejo Europeo estableció el nuevo Marco Estratégico para la cooperación europea en el ámbito de la educación y la formación «ET 2020». Este Marco ha diseñado objetivos estratégicos comunes para los Estados miembros, e indicadores y puntos de referencia que sirven para controlar el progreso hacia su consecución (Informe Español, 2013, p.6).

Con base en estas propuestas, y en consonancia con los objetivos definidos para 2010, los países miembros pusieron en práctica una serie de planes dirigidos a reducir la tasa de abandono escolar. En Bélgica, por ejemplo, se puso en marcha el programa *Tutorat*, en el que estudiantes universitarios dan apoyo educativo extraescolar a grupos reducidos de alumnos inmigrantes o desfavorecidos de la enseñanza secundaria superior. Irlanda, en 2005, desarrolló el DEIS (*Delivering Equality of Opportunity in Schools*), el cual pretende cubrir las necesidades educativas de los niños pertenecientes a entornos más desfavorecidos, dotando para ello a los centros participantes de una serie de recursos adicionales que incluyen más personal y financiación. Alemania y Francia, pese a no seguir las recomendaciones propuestas en el marco europeo, también cuentan con programas dirigidos a corregir las desigualdades sociales entre el alumnado: la Estrategia de apoyo específico a los alumnos con peor rendimiento, en el primero de ellos y, las zonas de Educación Prioritaria en el segundo. *The Youth Guidance Centre* en Dinamarca o *The Tanoda Centres* en Hungría.

En esta línea, en 2005 se puso en marcha en España un ambicioso proyecto, el Plan de Refuerzo Orientación y Apoyo. Este proyecto, fruto de la cooperación territorial entre el Ministerio de Educación y Ciencia y las Comunidades Autónomas, ha tratado de unificar esfuerzos entre Administración, centro, claustro directivo y profesorado, además de la propia familia, para reducir el negativo impacto de los factores genera-

dores de la desigualdad y garantizar la atención a los colectivos más vulnerables (véase el informe publicado por el MECD en 2007: Plan de Refuerzo Orientación y Apoyo).

A su vez, este plan de refuerzo educativo dirigido a la etapa de educación secundaria consta de dos modalidades: el Programa de Apoyo y Refuerzo (PAR) y el Programa de Acompañamiento Escolar (PAE), el cual se ofrece también en centros de educación primaria. El primero de ellos, el PAR, trata de alcanzar los objetivos definidos en el plan PROA mediante la intervención directa en tres ámbitos como son: centro, familia y entorno. Así, aquellos centros que decidan participar se verán obligados a ofrecer una atención directa al alumnado, facilitar y fomentar la relación con las familias y mejorar el entorno educativo.

La segunda modalidad, el PAE, a diferencia del anterior, focaliza su atención en el propio alumnado. En concreto, este programa se dirige a estudiantes con dificultades educativas a través del trabajo o apoyo organizado en horario extraescolar durante, al menos, cuatro horas semanales. Sus alumnos asistentes, generalmente en grupos de entre 8 y 11, son aquellos cuyos profesores y, especialmente, sus tutores hayan decidido que presentan dificultades en el aprendizaje: retraso en el proceso de maduración personal, mala integración en el grupo y en el centro, ausencia de hábitos de trabajo, retraso en el proceso de aprendizaje y otras similares. "Para superar estos inconvenientes, los monitores acompañantes o profesores llevarán a cabo las funciones de guía y orientación proporcionando, en su caso, los materiales adecuados, resolviendo dudas y ayudando en el desarrollo de actitudes y hábitos de organización del tiempo, planificación del trabajo, concentración y constancia en su elaboración y calidad en la realización y expresión de los resultados" (PLAN PROA Plan de Refuerzo, Orientación y Apoyo, 2007, p.18). Mediante este proceso se pretende actuar no sólo sobre sus alumnos asistentes, sino que también se espera mejorar el clima del centro, así como las expectativas escolares futuras del alumnado. Aspecto clave éste último a la hora de explicar las diferencias entre buenos y malos estudiantes en situación de riesgo (Rodríguez & Rosquete, 2019).

La siguiente tabla (tabla 1) recoge el número de centros que participaron en el programa PAE entre los años 2005 y 2012, distinguiendo aquellos que además participaron en la edición de PISA 2012 y que, por tanto, compondrán nuestra muestra objeto de interés.

Tabla1

Centros que implantaron el programa PAE y fueron evaluados en PISA 2012

	Total centros	2008/2009		2009/2010		2010/2011	
	PISA 2012	PAE	PISA	PAE	PISA	PAE	PISA
Andalucía	52	200	11	320	16	350	16
Aragón	51	19	3	28	6	31	8
Asturias	56	11	5	11	6	11	6
Islas Baleares	54	10	6	15	10	15	14

	Total centros	2008/2009		2009/2010		2010/2011	
	PISA 2012	PAE	PISA	PAE	PISA	PAE	PISA
Cantabria	54	10	9	10	5	18	6
Castilla León	55	36	6	36	7	36	7
Cataluña	51	71	5	92	4	92	4
Extremadura	53	23	4	37	8	50	11
Galicia	56	40	8	45	4	45	8
La Rioja	54	13	10	12	15	15	19
Madrid	51	100	9	109	11	114	11
Murcia	52	28	10	39	14	51	19
Navarra	51	6	3	7	2	8	2
País Vasco	174	13	3	30	13	42	20
TOTAL	902	692	97	908	130	984	154

Fuente: Instituto Nacional de Evaluación Educativa

Método

Muestra

En este estudio se ha empleado la base de datos PISA 2012, en la cual se evalúa a niños de 15 años de todo el mundo en las competencias matemática, lectora y ciencias. Además, cada alumno responde a un conjunto de preguntas sobre sus antecedentes familiares, ambiente escolar y hábitos de estudio. Esta información, relativa al estudiante, se ve complementada por el cuestionario rellenado por el propio centro, en el que se ofrece información sobre el personal, el ambiente del centro, las prácticas pedagógicas y los recursos materiales y humanos disponibles. Además de esta información, el MECD incorporó, para cada centro de la muestra, los años de participación en el programa. Sin embargo, no se especificó que alumnos asistieron al programa y cuáles no, de tal modo que la propia idiosincrasia de la base de datos sugiere la utilización del propio centro como unidad muestral. Esta hipótesis se ve reforzada por el planteamiento del programa, el cual pretende involucrar a toda la comunidad educativa. La muestra final quedó integrada por 518 centros.

Variables e instrumentos

En este apartado se describen las diferentes aproximaciones empleadas en la medición de la desigualdad educativa. Tal y como señala Marchesi en el año 2000: “la igualdad de resultados supone que se encuentran rendimientos similares entre los alumnos procedentes de distintas clases sociales, culturas o sexos” (p.3). Esta interpretación

sugiere, por tanto, que la variable empleada como output tome en consideración las diferencias en las calificaciones obtenidas por los alumnos como aproximación de la desigualdad, siendo esta mayor cuantas mayores diferencias se encuentren.

Para ello se han tenido en cuenta las pautas seguidas por Hanushek (2006), en el que evaluó, para más de 15 países, el efecto que el agrupamiento del alumnado tuvo sobre la desigualdad educativa. Como medidas de aproximación a la desigualdad empleó tanto la diferencia en términos de resultado entre percentiles (95 versus 5; 75 versus 25) como la desviación estándar sobre los mismos. En nuestro caso en particular la unidad muestral considerada ha sido el propio centro, por lo que la primera aproximación (95-5) precisa de un rango mayor a fin de garantizar un mínimo de individuos en cada nivel. En esta línea, se ha empleado como aproximación a la desigualdad la diferencia entre percentiles (90-10 y 75-25) para cada uno de los 518 centros que componen la muestra.

$$\text{Brecha educativa } i = P_{s90} - P_{s10} \quad (1)$$

$$\text{Brecha educativa } i = P_{s75} - P_{s25} \quad (2)$$

Donde i representa el centro y s al alumnado, recogiendo por tanto la siguiente igualdad las diferencias entre “buenos” y “malos” alumnos en la competencia lectora para cada centro. El motivo por el cual se ha empleado la competencia lectora reside en la elevada tasa de participación de la población estudiantil de origen inmigrante en el programa, los cuales encuentran dificultades adicionales en esta competencia como consecuencia de la diferencia idiomática.

Además, como novedad a la hora de buscar una medida que aproxime la desigualdad educativa a nivel de centro se ha empleado el índice de Gini (utilizado habitualmente en el análisis redistributivo para medir la desigualdad de la renta). En este caso en particular se usará para evaluar la desigualdad en términos de puntuaciones dentro de cada centro, entendida ésta como la diferencia entre las calificaciones obtenidas por cada alumno con respecto a la equidistribución de las mismas. En este sentido, cuanto mayor sea la distancia —mayor sea el Índice de Gini— mayor será la desigualdad entre las notas de los alumnos de un mismo centro. Sus valores comprenden entre 0 (máxima igualdad) y 1 (máxima desigualdad).

$$G = \left| 1 - \sum_{k=1}^{n-1} (X_{k+1} - X_k) (Y_{k+1} - Y_k) \right| \quad (3)$$

Donde, X es la proporción acumulada de población; Y es la proporción acumulada de resultados.

Por otra parte, y dado que la decisión a nivel de centro de participar vendrá determinada por una serie de características internas tales como la motivación del equipo directivo, los recursos del centro o, muy especialmente, el perfil de su alumnado, resulta útil conocer las características que expliquen su participación en el programa de refuerzo, todas ellas recogidas en PISA

Además, a fin de dotar al estudio de la máxima información posible, se ha incorporado a la base de datos original la puntuación predicha en lectura de los alumnos a la edad de 10 años. Para ello se ha construido un pseudo panel a partir de PISA 2012 y PIRLS 2006. La fusión, a partir de estas dos bases de datos, es posible debido a que ambas bases evalúan a la misma cohorte poblacional en diferentes etapas educativas: PIRLS a la edad de 10 años y PISA a la edad de 15 años, siendo el grupo poblacional objeto de estudio aquellos alumnos nacidos en el año 1996 (Villar-Aldonza et al., 2016). La aplicación de la metodología de imputación múltiple, algoritmo EM, en la base de datos PIRLS 2006, junto al método seguido por Choi, Gil, Mediavilla y Valbuena (2018), ha permitido obtener el conjunto de información necesario para realizar el posterior análisis. El proceso seguido parte de la identificación de aquellas variables comunes en ambas bases de datos: género, status socioeconómico, status de inmigración, nivel educativo de los padres, entre otras. Esta serie de variables o características, van a servir para explicar a través de una regresión econométrica los resultados en lectura de los alumnos evaluados en PIRLS 2006. Los coeficientes obtenidos de la regresión representan el efecto de cada una de las variables -género, status socioeconómico.- sobre la variable a explicar (resultados en lectura en PIRLS 2006). Posteriormente, estos coeficientes se incorporan manualmente a sus variables correspondientes en una nueva regresión definida en PISA 2012. De esta manera se obtiene la predicción de la puntuación que hubiese obtenido un alumno evaluado en PISA (a los 15 años), a la edad de 10 años dadas sus características.

Propensity Score Matching

Dado que los centros que participaron en el programa decidieron hacerlo voluntariamente por los beneficios que éste les ofrecía, nos encontramos ante un caso de autoselección (véase Scott, 2019). Esta situación provoca que los centros evaluados no sean iguales en sus características observables (véase tabla 1 Apéndice). Esta consideración daría lugar a errores en la interpretación de los resultados, no pudiendo concluir si las diferencias que se van a observar después del tratamiento se deben sólo a éste o a otro tipo de tendencia que los grupos arrastrarían desde el momento previo a la aplicación del programa. Este problema, comúnmente conocido como sesgo de selección, se deriva de la endogeneidad del predictor (véase Heckman, 1979). “Como solución a esta problemática se han propuesto múltiples estrategias metodológicas tales como: el método bietápico (Heckman, 1979), el diseño de regresión discontinua (Thistlewaite & Campbell, 1960), el método de diferencias en diferencias (Card & Krueger, 1995) y el método de variables instrumentales (Wright, 1928), entre otros” (Villar-Aldonza & Gambau-Suelves, 2018). En nuestro caso se ha empleado el Propensity Score Matching, en adelante PSM, propuesto por Rosenbaum y Rubin (1983). Esta técnica bietápica tiene el potencial de medir las diferencias entre dos grupos teniendo en cuenta sus características observables, de tal modo que la única diferencia que puede atribuirse en sus resultados es el número de años de participación en el programa.

A continuación, y siguiendo las pautas definidas por Caliendo y Kopeinig (2008), se exponen las fases seguidas en la implementación del PSM. En primer lugar, para tratar

de neutralizar el sesgo de selección mencionado anteriormente, se tratará identificar a centros que sean lo más homogéneos posible en todas sus características excepto en una, la participación en el programa de refuerzo. Rosenbaun y Rubin (1983) resolvieron este problema proponiendo una magnitud única, el propensity score, que no es sino un indicador sintético de la información contenida en las variables X de control. En efecto, esa puntuación —score— mide la probabilidad que un centro tiene de participar en el programa de refuerzo en función de sus características observables. Esta primera fase se lleva a cabo mediante un modelo de regresión logística o similar, como es nuestro caso, en el que se han empleado los modelos probits, cuya especificación se detalla a continuación:

$$p(D_i = 1|X_i) = \frac{\exp(\gamma X_i)}{1 + \exp(\gamma X_i)}$$

Los score permiten resumir la información observable de cada centro en un único valor, que será la referencia por el cual, en una segunda fase, los algoritmos matemáticos emparejen a aquellos centros que fueron tratados respecto a aquellos que no impartieron el programa (control), usando para ello un criterio lógico: cuanto más similares sean los score de los centros, más similares serán los resultados de las variables output; de manera que, si se aprecia alguna diferencia significativa en esos resultados será totalmente atribuible al impacto del programa de refuerzo analizado.

Con el objetivo de garantizar la robustez de los resultados se emplean tres estimadores de los emparejamientos diferentes. El primero de ellos, el vecino más cercano simple (SNN), se caracteriza por realizar los emparejamientos entre las unidades de tratamiento y control más próximos en términos de su score. En segundo lugar, la modificación que proponen Abadie, Drukker, Herr e Imbens (2004) sobre el estimador anterior (vecino más cercano). En este caso los emparejamientos se realizan entre la unidad de tratamiento y las cuatro unidades de control más similares ($m=4$) \otimes MNN. Por último, se ha utilizado el método kernel, el cual permite emparejar los centros tratados con un promedio ponderado de todas las unidades de control, asignando pesos decrecientes a aquellas que se encuentran más lejos en términos de probabilidad. La ventaja de este último algoritmo, frente al vecino más cercano o la modificación de Abadie et al. (2004), estriba en que permite considerar toda la información muestral del grupo de control, por lo que la estimación de impacto suele ser más precisa.

La clave del funcionamiento del PSM reside, por tanto, en la realización de un buen matching, es decir en encontrar centros del GC con una ps altamente similar a la de los del GT. Una vez calculado el score y seleccionado el algoritmo de emparejamiento, la tercera fase propuesta por Caliendo y Kopeinig (2008) pasa por contrastar la bondad de ajuste de estos emparejamientos, asegurando que los modelos cumplan las propiedades de balanceo y soporte común.

Una vez seleccionada la submuestra de centros comparables, la siguiente fase del PSM pasa por calcular el impacto del programa., haciendo uso para ello del estimador del efecto medio del tratamiento sobre los tratados (ATT), el cual considera el efecto sobre aquellos centros que participaron en el programa.

Resultados

Como se ha detallado en el epígrafe anterior, la primera fase en la realización del PSM consiste en encontrar un grupo de centros que no participaron en el programa de refuerzo PAE que sean comparables con aquellos otros que sí lo hicieron. Para ello se debe estimar, en primer lugar, la probabilidad —score— que un centro tiene de participar en el programa en base a sus características observables. En la tabla 2 se recogen los resultados de los cuatro modelos probits empleados en el trabajo, los cuales reportan un valor del ps para cada uno de los centros de la muestra. Este score será lo que permitirá acometer la segunda fase del PSM: la búsqueda de los emparejamientos óptimos entre los centros del GT y GC.

Los resultados obtenidos señalan que las variables que mayor grado de influencia concentran en la probabilidad de participar en el programa son: la puntuación predicha en PIRLS, el porcentaje de madres con estudios superiores, el porcentaje de alumnado nativo, el porcentaje de repetidores, el grado de perseverancia medio en el centro, los problemas de conducta y, por último, la localización del centro. Estos resultados se muestran coherentes con los objetivos del programa, ya que está dirigido principalmente a aquellos centros que concentran un mayor porcentaje de alumnos en situación de desventaja educativa así como un peor ambiente escolar. Aunque el R^2 no supera el 20% para ninguno de los modelos estimados, en estos modelos es más importante el porcentaje de predicciones correctas, que en nuestro caso se sitúa en torno al 70%, considerado en la literatura un grado de fiabilidad aceptable.

Tabla 2.

Modelos probits correspondientes a las variables resultado.

VARIABLES INDEPENDIENTES(1)	PISA	BRECHA (90-10)	BRECHA (75-25)	GINI
PIRLS(2)	-1355.37*** (0.004)	-34.730 (0.362)	-29.334 (0.349)	-45.265* (0.142)
Padres_estudios_medios(%)	-0.157 (0.913)	-0.0665 (0.962)	0.3305 (0.815)	0.0583 (0.967)
Padres_estudios_superiores(%)	1.638 (0.255)	0.249 (0.860)	1.309 (0.360)	1.014 (0.465)
Madres_estudios_medios(%)	-1.536 (0.678)	-0.1164 (0.972)	-0.277 (0.934)	-3.313 (0.354)
Madres_estudios_superiores(%)	-5.544* (0.141)	-3.764 (0.272)	-3.517 (0.321)	-6.589** (0.077)
Alumnado_nativo(%)	-3.502*** (0.039)	-2.673*** (0.027)	-0.868 (0.430)	-1.292 (0.230)
Familias_status_socioec._alto(%)	1.837 (0.230)	2.032 (0.163)	1.628 (0.269)	1.963 (0.180)

VARIABLES INDEPENDIENTES(1)	PISA	BRECHA (90-10)	BRECHA (75-25)	GINI
Repetidores_secundaria(%)	3.702*** (0.000)	3.129*** (0.002)	3.077*** (0.002)	3.432*** (0.001)
Género(hombres)(%)	0.187 (0.889)	1.275 (0.313)	0.617 (0.619)	0.8184 (0.520)
Hogares_con_entre_101_y_200_libros(%)	-4.083 (0.501)	2.060 (0.698)	2.407 (0.646)	-1.178 (0.844)
Hogares_con_más_de_200 libros(%)	-2.055 (0.737)	3.688 (0.487)	4.05 (0.44)	0.638 (0.914)
Alumnos_asistieron_ISCED 0_más_de_un_año(%)	-0.402 (0.729)	-0.912 (0.417)	-0.675 (0.550)	-0.048 (0.968)
Hogares_con_posesiones_culturales_altas(%)	-0.859 (0.490)	-1.650 (0.178)	-1.571 (0.187)	-0.697 (0.566)
Madres_trabajan_tiempo_completo(%)	0.177 (0.875)	-0.781 (0.479)	-0.774 (0.468)	-0.493 (0.644)
Madres_trabajan_tiempo_parcial(%)	1.367 (0.318)	1.144 (0.372)	0.549 (0.668)	1.655 (0.178)
Nacimiento_primer_semestre(%)	-1.810 (0.164)	-1.242 (0.276)	-0.496 (0.664)	-0.822 (0.471)
Familia_monoparental(%)	2.054 (0.285)	1.598 (0.387)	2.454 (0.202)	2.562 (0.173)
Moral_profesor_alta	0.226 (0.374)	0.163 (0.509)	0.022 (0.926)	0.100 (0.681)
Cursos_profesorado	-0.074 (0.796)	0.0081 (0.976)	0.149 (0.575)	0.0754 (0.781)
Comportamiento_alumnado	1.770 (0.194)	0.758 (0.554)	1.186 (0.362)	1.394 (0.296)
Perseverancia	-2.317* (0.150)	-2.329* (0.134)	-2.348* (0.130)	-2.351* (0.140)
Autocontrol	0.229 (0.878)	0.1738 (0.906)	0.004 (0.998)	1.042 (0.483)
Localización (urbano)	-0.347 (0.237)	-0.5089** (0.077)	-0.549** (0.056)	-0.474** (0.100)
Índice_diversidad	0.078 (0.550)	0.0416 (0.747)	0.110 (0.387)	0.100(0.434)
Índice_comportamiento	0.071 (0.602)	0.090 (0.496)	0.066 (0.620)	0.111 (0.417)
Índice_absentismo	0.086 (0.548)	0.0638 (0.649)	0.048 (0.730)	0.103 (0.455)

VARIABLES INDEPENDIENTES(1)	PISA	BRECHA (90-10)	BRECHA (75-25)	GINI
Profesor_valora_resultados	0.104 (0.811)	0.0169 (0.966)	0.090 (0.824)	-0.128 (0.747)
Problemas_conducta	-1.121*** (0.009)	-0.602* (0.133)	-0.618* (0.128)	-1.160*** (0.004)
Log likelihood	-268,9	-280,53	-283,01	-279,17
LR Chi2	127.08 (0.000)	103.82 (0.000)	98.85(0.000)	106.55 (0.000)
PCP (Porcentaje de predicciones correctas)	71,83%	69,41%	71,02%	70.42%
Pseudo R2	0,191	0,156	0,148	0,160
Soporte_Común	[0.045-0.998]	[0.106-0.995]	[0.090-0.952]	[0.065-0.974]

(1) Los modelos probit incluyen la interacción independiente entre todas las variables para ajustar un score más representativo de las características particulares de cada centro. En total se especificaron 12 modelos con el objetivo de estimar el efecto aislado del tratamiento en función del número de años que el centro participó en el programa

(2) La variable PIRLS se configura en cada modelo probit en relación a su variable resultado en la que se va a medir el impacto. Por ejemplo, en el modelo probit para PISA, se utiliza la puntuación predicha en PIRLS. Para la BRECHA (90-10), se utiliza la diferencia entre el percentil 90 y 10 de PIRLS. De manera análoga para las otras variables.

Nota. Entre paréntesis se encuentran los p-valores. *significatividad estadística al 15%;** significatividad estadística al 10%; ***significatividad estadística al 1%

En este punto, resulta conveniente destacar que todos los modelos cumplen una serie de características deseables para garantizar el buen emparejamiento. En concreto, todos los modelos presentan una bondad de ajuste elevada, significatividad conjunta de los parámetros, zonas de soporte común amplias (entre el 0.1X y el 0.93X en media). Una vez estimado el ps, el siguiente paso será realizar el emparejamiento (matching) entre los centros tratados y no tratados. Los estimadores empleados, como ya se ha señalado en el epígrafe anterior, fueron: el vecino más cercano simple, el modificado por Abadie et al.(2004) y el Kernel.

Los resultados obtenidos, ofrecidos en la tabla 3, pese a no presentar un efecto estadísticamente significativo, muestran una clara evolución en función del número de años de aplicación del programa. En efecto, para el primer año se observa cómo disminuye la nota media -desde los 9.3 puntos calculados con el estimador de kernel hasta los 60.2 puntos obtenidos con Abadie, a la par que la desigualdad educativa no presenta un comportamiento claro. Para la brecha entre el percentil 90-10 se reduce en dos de las tres estimaciones realizadas, incrementándose para el percentil 75-25 y el índice de Gini en dos de sus tres estimaciones.

A los dos años de aplicación, sin embargo, la nota media aumenta tanto por el método del vecino más cercano simple como por el modificado por Abadie et al., disminuyendo por el algoritmo de Kernel. La desigualdad, por su parte, se incrementa si se consideran las diferencias entre las puntuaciones obtenidas por los peores y los mejores alumnos (brecha 90-10), mientras que se reducen si se amplía la muestra al 25% (brecha 75-25).

Finalmente, el tercer año de aplicación del programa presenta unos resultados mucho más esperanzadores. En efecto, a los tres años no sólo aumenta la nota media -desde 1.6 con Abadie et al. a 14.1 con el vecino más cercano-, sino que también se reducen las diferencias entre grupos (brecha 90-10 y brecha 75-25), resultando significativa la estimación mediante el vecino más cercano para la brecha 75-25. El índice de Gini, de igual modo que para el segundo año de aplicación del programa, presenta un signo negativo en dos de sus tres estimaciones.

Tabla 3

Estimación del impacto sobre los resultados en lectura y la desigualdad del programa de refuerzo educativo PAE para aquellos centros que lo impartieron 1, 2 o 3 años (ATT).

Número de años	Algoritmo	Puntuación Media enlectura	BRECHA EDUCATIVA (90-10)	BRECHA EDUCATIVA (75-25)	ÍNDICE DE GINI
1 AÑO	SNN	-10.41 (-0.908)	-33.77 (-1.809)	3.20 (+0.494)	0.001 (+0.154)
	MNN	-60.20 [0.249]	33.90 [0.598]	1.42 [0.764]	7.76E-11 [1.000]
	KERNEL	-9.34 (-1.17)	-19.54 (-1.312)	-1.21 [0.216]	-0.001 (-0.209)
2 AÑOS	SNN	4.99 (+0.507)	23.55 (+1.833)	-11.89 (-1.346)	0.007 (+0.971)
	MNN	3.99 [0.457]	2.41 [0.700]	-3.36 [0.521]	-6,00E-10 [1.000]
	KERNEL	-0.23 (-0.031)	9.56 (+0.949)	-7.43 (-1.178)	0.003 (+0.523)
3 AÑOS	SNN	14.11 (+1.192)	-22.29 (-1.312)	-20.36*** (-2.146)	0.003 (+0.46)
	MNN	1.63 [0.770]	-9.80 [0.064]	5.74 [0.234]	-1.5E-09 [1.000]
	KERNEL	4.81 (+0.614)	-16.22 (-1.128)	-7.02 (-1.064)	0.003 (+0.596)

Nota: Entre corchetes se encuentran los p-valores y entre paréntesis los valores de los t-ratio.

Discusión y conclusiones

Con este trabajo se ha tratado de contribuir a conocer un poco mejor un fenómeno en expansión como son los programas de refuerzo educativo. En efecto, su promoción y puesta en funcionamiento responde a las últimas directrices en materia de política educativa estipuladas en el marco europeo con el objetivo de reducir y prevenir el aban-

dono y fracaso escolar. Sin embargo, el novedoso carácter de este tipo de programas hace que, en la actualidad, sean necesarios más estudios que contribuyan a conocer y comprender mejor los efectos derivados de este tipo de prácticas sobre el alumnado. La novedad de este estudio, frente a otros trabajos presentados anteriormente, radica en la consideración de la desigualdad como variable resultado. Para ello se han seguido las pautas establecidas por Hanushek en su trabajo de 2006, donde evaluaba el impacto del agrupamiento *tracking* sobre la desigualdad tomando como muestra un conjunto de más de 15 países. En este caso en particular, la naturaleza de los datos ha permitido desarrollar el análisis tomando como conjunto muestral el propio centro. Como medidas de aproximación a la desigualdad se han empleado las brechas existentes entre el alumnado de un mismo centro (percentil 75-25; percentil 90-10) y, como novedad, se ha utilizado el índice de Gini. Estas alternativas en la medición de la desigualdad permiten capturar no sólo la brecha educativa entre los buenos y malos alumnos, sino que también recogen los cambios en la distribución de las notas del alumnado dentro de cada centro; calculada ésta última a partir del índice de Gini.

Los resultados obtenidos, a partir de la técnica PSM, estarían señalando que el programa de refuerzo no tuvo un impacto significativo en la reducción de la desigualdad, de igual modo que tampoco mejoraron las calificaciones medias entre aquellos centros que hubieron participado. Por otro lado, si se distingue en función del número de años de aplicación, nuestra recomendación en materia de política educativa sugiere, pese a no obtenerse unos resultados estadísticamente significativos, que la implementación de este tipo de programas se imponga con carácter obligatorio durante los tres primeros cursos de la secundaria. En efecto, la puesta en marcha del programa por un único año ha mostrado una reducción en la nota media en lectura así como un efecto nada claro sobre la desigualdad.

En su segundo año de aplicación, sin embargo, la nota media a nivel de centro aumenta para dos de las tres estimaciones empleadas, a la par que la desigualdad se ve reducida si se consideran las diferencias entre los peores y los mejores alumnos (brecha 90-10) y aumenta si se atiende a la brecha 75-25. Finalmente, en su tercer año de aplicación, la nota media a nivel de centro mejora para las tres estimaciones realizadas, reduciéndose las diferencias entre grupos (brecha 90-10 y brecha 75-25) hasta el punto de resultar significativa la estimación por el vecino más cercano simple para la brecha 75-25. Estos resultados estarían sugiriendo, por tanto, que una vez superado el coste de entrada asociado a su primer año de aplicación se obtendrían unos resultados más esperanzadores.

Estos resultados deben, en todo caso, ser interpretados con cautela. En efecto, aunque en el estudio no se han escatimado los esfuerzos por reducir al mínimo el sesgo de selección que potencialmente afecta a nuestros datos, no es posible asegurar que las estimaciones estén totalmente libres de este problema, que, por otra parte, se encuentra intrínseco en los estudios que evalúan las intervenciones educativas (Murnane & Willett, 2010). Por otro lado, se ha de señalar la necesidad de evaluaciones de este tipo en España que permitan conocer que alumnos asistieron a dicho programa, de tal modo que seamos capaces de evaluar su impacto en función de las características del propio alumnado. En este sentido, conocer al propio estudiante será la clave para que este tipo de prácticas alcancen con éxito sus objetivos.

Referencias

- Abadie, A., Drukker, D., Herr, J.L., & Imbens, G.W. (2004). Implementing matching estimators for average treatment effects in Stata. *Stata Journal* 4(3):290-311. DOI: <https://doi.org/10.1177/1536867X0400400307>
- Battistin, E., & Meroni, E. C. (2016). Should we increase instruction time in low achieving schools? Evidence from Southern Italy. *Economics of Education Review*, 55, 39-56. doi:<https://doi.org/10.1016/j.econedurev.2016.08.003>
- Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, 122(3), 1235-1264 <https://doi.org/10.1162/qjec.122.3.1235>
- Briggs, D. C. (2001). The effect of admissions test preparation: Evidence from NELS: 88. *Chance*, 14(1), 10-18. doi: <https://doi.org/10.1080/09332480.2001.10542245>
- Calero, J., Choi, A., & Waisgrais, S. (2010). Determinantes del riesgo de fracaso escolar en España: una aproximación a través de un análisis logístico multinivel aplicado a PISA-2006. *Revista de Educación, extraordinario 2010*, 225-256.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72. doi: <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Card, D., y Krueger, A. B. (1994). Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, Vol. 84 No. 4, 772-793.
- Choi, Á., Gil, M., Mediavilla, M., & Valbuena, J. (2018). The Evolution of Educational Inequalities in Spain: Dynamic Evidence from Repeated Cross-Sections. *Social Indicators Research*, 138(3), 853-872.
- De Paola, M., & Scoppa, V. (2014). The effectiveness of remedial courses in Italy: a fuzzy regression discontinuity design. *Journal of Population Economics*, 27(2), 365-386.
- Gamoran, A. (2001). American Schooling and Educational Inequality: A Forecast for the 21st century. *Sociology of Education*, 74, 135-153. DOI: 10.2307/2673258
- García-Pérez, J. I. & Hidalgo-Hidalgo (2017). No student left behind? Evidence from the Programme for School Guidance in Spain. *Economics of Education Review*, 60, 97-111. doi: <https://doi.org/10.1016/j.econedurev.2017.08.006>
- González-Rodríguez, D., Vieira, M.J. & Vidal, J. (2019). La percepción del profesorado de Educación Primaria y Educación Secundaria sobre las variables que influyen en el Abandono Escolar Temprano. *Revista de Investigación Educativa*, 37(1), 181-200. DOI: <http://dx.doi.org/10.6018/rie.37.1.343751>
- Hanushek, E. A. (2006). Does Educational Tracking Affect Performance and Inequality? Differences in Differences Evidence Across Countries. *The Economic Journal*, 116(510), C63-C76. doi: <https://doi.org/10.1111/j.1468-0297.2006.01076.x>
- Heckman J.J. (1979). Sample selection bias as a specification error. *Econometrica. Journal of the Econometric Society* 47 (1), págs.153-161.
- Lavy, V., & Schlosser, A. (2005). Targeted remedial education for underperforming teenagers: Costs and benefits. *Journal of Labor Economics*, 23(4), 839-874.
- Marchesi, A. (2000). Un sistema de indicadores de desigualdad educativa. *Revista Iberoamericana de Educación*, 23, pp. 135-163

- MECD (2007). Plan de Refuerzo Orientación y Apoyo. https://sede.educacion.gob.es/publiventa/descarga.action?f_codigo_agc=12206_19&f_cod_area=E&f_titulo=Plan+de+refuerzo+orientaci%C3%B3n+y+apoyo.+PROA+2007&f_extension=pdf&method:descargaFichero=Download+file
- Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.
- Rosenbaum, P. R. & Rubin, D.B. (1983). The central role of propensity score in observational studies for causal effects, *Biometrika*, 70. 41-55. doi: <https://doi.org/10.1093/biomet/70.1.41>
- Rodríguez, D. R., & Rosquete, R. G. (2019). Rendimiento académico de adolescentes declarados en situación de riesgo. *Revista de Investigación Educativa*, 37(1), 147-162. doi: <http://dx.doi.org/10.6018/rie.37.1.303391>
- Scott, P. W. (2019). Causal Inference Methods for selection on observed and unobserved factors: Propensity Score Matching, Heckit Models, and Instrumental Variable Estimation. *Practical Assessment, Research & Evaluation*, 24(3), 2.
- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of educational research*, 60(3), 471-499. <https://doi.org/10.3102/00346543060003471>
- Thistlewaite, D. & Campbell, D. (1960): "Regression-Discontinuity Analysis: An alternative to the ex post facto experiment". *Journal of Educational Psychology* 51. 309-317. DOI: <http://dx.doi.org/10.1037/h0044319>
- Villar-Aldonza, A. Mancebón, M.J, Castro, G. & Sancho, J. M.G (2016). Evaluación del programa de refuerzo PROA ¿Es realmente una medida eficiente?. *Investigaciones de Economía de la Educación volume 11*, 451-466.
- Villar-Aldonza & Gambau-Suelves (2018). El problema de la brecha educativa entre buenos y malos alumnos, ¿son los programas de refuerzo la solución? Recuperado de: <http://2018.economicsofeducation.com/user/pdfsесiones/204.pdf?PHPSESSID=83r9grpiasn1krigmjsubkg3l3>
- Wright, P. G (1928), The Tariff on Animal and Vegetable Oils, *The American Economic Review*, 19(1): 152-156.

Apéndice

Tabla 1

Prueba t- de igualdad de medias entre aquellos centros que participaron en el programa y aquellos otros que no lo hicieron.

Descripción variable	PAE	NO PAE	Diferencia	Prueba t de igualdad de medias
Madres_estudios_primarios(%)	0.409	0.343	-0.067	0.00
Madres_estudios_secundarios(%)	0.295	0.294	-0.001	0.90
Madres_estudios_universitarios(%)	0.297	0.350	0.067	0.00
Alumnado_nativo(%)	0.837	0.881	0.043	0.00
Familias_posesiones_culturales_altas(%)	0.329	0.357	0.028	0.017
Familias_status_socioeconómico_alto(%)	0.245	0.286	0.041	0.305
Índice_absentismo_escolar	0.305	0.36	0.055	0.000
Puntuación_predicha_10 años	535.69	539.33	3.633	0.00

Tabla 2

Estadísticos descriptivos de las variables explicativas de los modelos probit.

Descripción variable	N	Media	Desv. Estándar	Mínimo	Máximo
Padres_estudios_medios(%)	518	0.26	0.10	0.03	0.57
Padres_estudios_superiores(%)	518	0.33	0.16	0.00	0.76
Madres_estudios_primarios(%)	518	0.37	0.17	0.00	0.92
Madres_estudios_medios(%)	518	0.29	0.10	0.00	0.61
Madres_estudios_superiores(%)	518	0.34	0.16	0.04	0.81
Alumnado_nativo(%)	518	0.86	0.14	0.14	1.00
Familias_status_socioec._alto(%)	518	0.34	0.17	0.00	0.84
Repetidores_secundaria(%)	518	0.29	0.13	0.00	0.75
Género (hombres)(%)	518	0.50	0.09	0.07	0.83
Hogares_con_menos_de_100_libros(%)	518	0.56	0.16	0.09	1.00
Entre_101_y_200_libros(%)	518	0.20	0.09	0.00	0.46
Más_de_200_libros(%)	518	0.24	0.13	0.00	0.64
Alumnos_asistieron_ISCED0_más_de_un_año(%)	518	0.84	0.11	0.29	1.00

Descripción variable	N	Media	Desv. Estándar	Mínimo	Máximo
Hogares_posesiones_culturales_altas(%)	518	0.35	0.13	0.00	0.77
Madres_trabajan_tiempo_completo(%)	518	0.45	0.15	0.03	0.84
Madres_trabajan_tiempo_parcial(%)	518	0.21	0.10	0.00	0.83
Mes_denacimiento_primer_semestre(%)	518	0.50	0.10	0.08	0.77
Familia_monoparental(%)	518	0.09	0.06	0.00	0.43
Moral_profesor_alta	518	0.68	0.47	0.00	1.00
Cursos_profesorado	518	0.23	0.42	0.00	1.00
Comportamiento_alumnado(%)	518	0.46	0.11	0.07	0.67
Perseverancia(%)	518	0.12	0.08	0.00	0.45
Autocontrol(%)	518	0.81	0.08	0.52	1.00
Localización (urbano)	518	0.27	0.44	0.00	1.00
Índice_diversidad	503	0.25	0.97	-2.02	2.11
Índice_comportamiento	500	0.29	0.95	-1.59	4.10
Índic_absentismo	500	0.32	0.96	-1.48	3.39
Profesor_valora_resultados	518	0.90	0.30	0.00	1.00
Problemas_conducta	518	0.90	0.30	0.00	1.00

Tabla 3

Test de diferencia de medias para las variables objeto de estudio antes y después del PSM.

Variable	Muestra	Tratados	Controles	Diferencia	S.E.	T-stat
PISA	Unmatched	474.71	484.79	10.08	3.50	2.87
	ATT	475.20	482.66	7.46	3.57	2.08
PERCENTIL	Unmatched	206.24	203.56	-2.68	-9.54	-0.76
	ATT	206.34	205.16	-1.17	-8.28	-0.32
RANGO IQ	Unmatched	108.91	108.29	-0.61	-5.54	-0.24
	ATT	109.25	108.77	-0.47	-5.61	-0.18
GINI	Unmatched	0.095	0.091	-0.003	-0.007	-2.03
	ATT	0.095	0.092	-0.003	-0.006	-1.68

Fecha de recepción: 2 de septiembre de 2019.

Fecha de revisión: 10 de octubre de 2019.

Fecha de aceptación: 9 de enero de 2020.