

# EVALUACIÓN DE POLÍTICAS EDUCATIVAS EN ESPAÑA

**Francisco Pedraja Chaparro**

## **Introducción**

Este trabajo sobre la evaluación de las políticas públicas en el campo de la educación tiene su origen en mi experiencia investigadora más reciente en el ámbito de la Economía Pública.

La justificación de la evaluación de las políticas públicas puede hacerse, al menos, desde tres dimensiones. En primer lugar, la rendición de cuentas a los ciudadanos sobre la utilización de los fondos públicos por parte de sus gobernantes es una pieza fundamental en toda democracia. En segundo lugar, los principios de eficacia y eficiencia en el uso de los recursos públicos aparecen recogidos en nuestras normas de mayor rango, Constitución Española (artículos 31.2 y 103.1), Ley Orgánica de Estabilidad Presupuestaria y Sostenibilidad Financiera (artículo 7) y Ley General Presupuestaria (artículo 26), por poner tres ejemplos relevantes. Por último, desde una perspectiva más pragmática, el diseño correcto de la política pública, esto es, conocer si alcanza sus objetivos y, en caso contrario, proponer su modificación o supresión exige su evaluación de manera que los programas públicos se basen en resultados y no en buenas intenciones o meras intuiciones.

Por otra parte, el gasto en educación, uno de los programas fundamentales del Estado del Bienestar, muestra efectos positivos en los dos objetivos básicos de la Economía Pública, eficiencia (crecimiento) y equidad (igualdad de oportunidades) además de generar otras consecuencias favorables para la sociedad como son la mejora en los niveles de salud, la reducción de la actividad delictiva o la mejora de su calidad democrática<sup>1</sup>. De modo más específico, si acotamos el ámbito de análisis a los resultados escolares, dada la amplitud de su contenido y el espacio del que dispongo, existen pruebas evidentes de la relación de estos últimos con el crecimiento económico. Según Hanushek y Woessman (2011), medidas cualitativas de capital humano, es decir, resultados escolares en pruebas internacionales estandarizadas, explican una parte importante de la variación interanual en las tasas de crecimiento del PIB entre países. Con resultados del informe PISA, aproximadamente tres cuartas partes de la variación de las tasas de crecimiento entre los

---

<sup>1</sup> Para una revisión exhaustiva de los favorables efectos de la educación tanto individuales como sociales puede consultarse el trabajo de Wolfe y Zuvekas (1997).

países puede explicarse por dos variables: el nivel inicial de ingresos y los resultados académicos de la población.

El desarrollo de la evaluación de las políticas públicas, como le ha sucedido a la propia evolución de la Economía Pública como una disciplina cada vez más empírica (Martínez-Vázquez, 2018), se ha visto favorecido por dos elementos claramente relacionados. Por una parte, la disponibilidad de micro datos administrativos de carácter longitudinal. Por otra, el desarrollo de nuevas técnicas, en especial, el uso de experimentos aleatorios (*Randomized Controlled Trials* o RCTs), recientemente refrendadas con la concesión de varios premios nobel de Economía (Banerjee, Duflo y Kremer en 2019 y Angrist, Card e Imbens en 2021).

Delimitado el objetivo del discurso y motivado su interés, en lo que sigue analizo, desde una perspectiva institucional, la situación de la evaluación de las políticas públicas en nuestro país en términos comparados. A continuación, hago un breve repaso a la revolución producida en las técnicas de evaluación antes de llevar a cabo una panorámica de los estudios realizados en España sobre evaluación de impacto en el ámbito de los resultados escolares. Cierro el trabajo con un epílogo en el que destaco algunas recomendaciones para este tipo de análisis.

### **La evaluación de políticas públicas en España desde una perspectiva institucional**

Desde una perspectiva institucional, como señalan Feinstein y Zapico (2010), las evaluaciones en España durante las décadas finales del pasado siglo y los primeros años del actual se han venido realizando de forma aislada y su retraso se debe a razones tanto institucionales, como de cultura político administrativa. Tenemos un sesgo hacia una evaluación de tipo más fiscalizador y procedimental frente a otra más enfocada en resultados e integrada en el proceso presupuestario. En definitiva y a diferencia de lo que sucede en otros países avanzados en este campo (EEUU y Reino Unido), falta en España una cultura de evaluación.

El primer intento serio de institucionalización de evaluación de políticas públicas en España fue la creación en 2006 de la Agencia Estatal de Evaluación de Calidad de los Servicios y de las Políticas Públicas (AEVAL). Después de una vida escasamente fructífera, esta Agencia acabó suprimiéndose en 2017 en un contexto de racionalización del sector público. En su labor apenas encontramos verdaderos análisis de eficiencia o evaluaciones de impacto de programas y políticas públicas, centrándose su actividad en

cuestiones metodológicas, programas de formación y elaboración de guías de evaluación. Junto a la falta de personal cualificado, su escasa, por no decir nula independencia (vinculada al ejecutivo y con una presidencia nombrada directamente por el gobierno) fue señalada como su principal problema de diseño (Albi y Onrubia, 2018).

Análisis recientes sitúan a nuestro país en un nivel medio bajo en evaluación de políticas públicas. Así sucede por ejemplo en el trabajo de Stockmann, Meyer y Taube (2020) que examinan la situación de 17 países a partir de encuestas de expertos nacionales en tres dimensiones (política, profesional y social)<sup>2</sup>. Afortunadamente, la situación parece estar cambiando y, como casi siempre sucede en todo movimiento modernizador en nuestro país, de la mano de Europa.

En primer lugar, destaca en este cambio el protagonismo de la Autoridad Independiente de Responsabilidad Fiscal (Airef), una institución de origen y vinculación claramente europeísta. Aunque la misión principal de este organismo es velar por la sostenibilidad de las finanzas públicas, en la práctica se ha convertido en el principal agente evaluador de las políticas públicas en nuestro país a lo largo de los últimos años. En 2021 se pone en marcha la División de Evaluación del Gasto Público de la Airef, consolidando una actividad que la institución venía realizando con sus revisiones de gasto (*Spending Reviews*) desde 2017. En una declaración del Eurogrupo de septiembre de 2016 se recomendaba a los países del euro utilizar los *spending reviews* o revisiones de gasto como herramienta para mejorar la calidad de las finanzas públicas y el propio Consejo formuló recomendaciones expresas a algunos países, entre ellos a España en 2017, para realizar una revisión integral del gasto. A diferencia de la AEVAL, la Airef goza de independencia (la elección de su presidencia requiere mayoría absoluta del Parlamento, su período de mandato no coincide con las legislaturas y se financia mediante una tasa) y desarrolla abundantes evaluaciones encargadas de forma separada del ciclo político por todos los niveles de las administraciones públicas en materias de gran relevancia (sanidad, educación, políticas activas de empleo, incentivos fiscales, ingreso mínimo vital, etc.).

En segundo lugar, El Plan de Recuperación, Transformación y Resiliencia (PRTR) de 2021 del Gobierno de España para acceder a los instrumentos de financiación comunitarios (Next Generation EU) a raíz de la pandemia, constituye un importante

---

<sup>2</sup> Para una revisión de la evaluación de las políticas públicas en España desde una perspectiva institucional pueden consultarse los trabajos de FEDEA (2021) y Casado y Del Pino (2021 y 2022).

impulso en la evaluación de las políticas públicas a juzgar por la forma en la que deben justificarse la percepción de los fondos asociados a ese plan. En efecto, el desembolso de los fondos se liga al desempeño, es decir, al cumplimiento satisfactorio de hitos y objetivos y no a la ejecución de los proyectos como venía ocurriendo con los anteriores fondos europeos.

La evaluación aparece explícitamente en algunos de los componentes del PRTR. En el número 11, dedicado a la modernización de las administraciones públicas, uno de sus hitos es una Ley de Evaluación de Políticas Públicas cuyo anteproyecto se acaba de aprobar en Consejo de Ministros. En el número 29, referido a la mejora de la eficacia del gasto público, se consolida el papel de la Airef como organismo evaluador planteándose un nuevo periodo de revisiones de gasto (2022-2026) y se crea una unidad en el Ministerio de Hacienda para el seguimiento de las propuestas de la Airef derivada de sus evaluaciones y su incorporación al ciclo presupuestario.

Termino este breve apunte con dos importantes limitaciones y un reto para la evaluación de políticas públicas en España.

Una primera limitación se encuentra en los datos. Contamos con una enorme riqueza de datos administrativos procedentes de diversas fuentes pero no están conectados ni disponibles ni preparados para su explotación. También es cierto que el diseño de los sistemas de información responde a la finalidad con la que se registran que, como ya dijimos, no es la de evaluar sino más bien controlar el gasto público. Por otra parte, la acusada descentralización de la administración española acentúa el problema de los datos. Existe coincidencia entre los investigadores (Almunia y Rey-Biel, 2021) y las instituciones dedicadas a la evaluación (Airef, 2020) en impulsar una estrategia nacional para la disponibilidad y uso de los datos que contienen los registros administrativos convenientemente anonimizados poniéndolos a disposición de las Administraciones y de la academia.

Una segunda limitación es la adecuada formación del personal dedicado a las tareas evaluadoras, en especial cuando nos referimos a los enfoques más avanzados sobre evaluación, en los que el acento se pone en el establecimiento de causalidad entre los programas públicos y sus resultados e impactos. En esos casos se exige una especialización de los recursos humanos que no se encuentra en la administración pública y que está al alcance solo de unos pocos académicos que la aplican en sus investigaciones.

Un reto evidente para la evaluación de las políticas públicas en España lo constituye la descentralización de nuestro sector público, que exige una elevada dosis de coordinación, de manera que puedan ponerse de manifiesto y utilizarse las ganancias en eficiencia identificando las mejores prácticas por comparación de distintas unidades en la prestación de los mismos servicios (benchmarking), como pronostica la teoría del federalismo fiscal (Oates, 1999). Los organismos internacionales vienen haciendo notar desde hace tiempo estas carencias en coordinación como una clara deficiencia de nuestro Estado de las autonomías (Jourmard y Giorno, 2005).

### **Sobre técnicas de evaluación, algunos ejemplos en la literatura internacional en educación**

En este apartado me aproximo a las técnicas de evaluación de impacto que intentan medir el efecto de una política educativa sobre la variable objetivo, en nuestro caso, el rendimiento escolar, acompañando las distintas alternativas con algunos ejemplos de investigaciones relevantes en la literatura internacional.

El problema que surge al llevar a cabo una determinada política educativa, como podría ser la reducción del número de alumnos por clase o la ampliación de la enseñanza preescolar, es que hay otras variables que pueden también influir en ese objetivo, el rendimiento escolar, y deberíamos ser capaces de identificar el efecto exclusivo que sobre él tiene la política.

En evaluación de políticas públicas, el principal inconveniente es de identificación, es decir, se trata de seleccionar un grupo de control adecuado; aquel que es exactamente igual al grupo de individuos tratados por la política, salvo en la aplicación de la misma. De este modo, podremos identificar los efectos causales, siendo el método más preciso la aleatorización. Seleccionando aleatoriamente a los individuos que recibirán la política (grupo tratado) y a los que no (grupo de control), garantizamos que las diferencias entre ambos grupos se deban a la política que pretendemos evaluar (Scholotter et al., 2011). Este método, habitual en ciencias de la salud, se aplica con mucha menor frecuencia en las ciencias sociales por razones de distinta naturaleza. Unas son éticas, parece más adecuado conceder becas a aquellos alumnos que las necesitan que no mediante un sorteo. Otras son económicas, debidas al elevado coste en tiempo y dinero de realizar experimentos; sin olvidar la imposibilidad, en muchas ocasiones, de su aplicación ex ante dado el carácter general con el que se aplican las políticas públicas. En todo caso, parece

preferible ensayar políticas públicas sobre las que existe un elevado grado de incertidumbre mediante un programa piloto en el que evaluemos su impacto, que aplicar de forma generalizada esa política al conjunto de la población.

Uno de los primeros experimentos aleatorios en el ámbito educativo fue el Proyecto Perry, un programa que se puso en marcha en los años sesenta en Estados Unidos con el propósito de analizar la efectividad de los programas de educación preescolar. Este estudio contó con la participación de 123 alumnos en riesgo de fracaso escolar divididos de forma aleatoria en dos grupos; uno siguió un programa preescolar de alta calidad basado en el enfoque de aprendizaje activo, y otro, grupo de control, sin acceso al programa. La principal particularidad de este estudio es que se hizo un seguimiento longitudinal de los alumnos. Han sido muchos los trabajos que han explotado este experimento, aunque los análisis más relevantes, aquellos realizados por Heckman y sus colaboradores, concluyeron con que los beneficiarios del programa obtuvieron mejoras significativas en los ámbitos educativo y laboral (acceso al empleo, estabilidad laboral y salarios), así como mejores niveles de salud y menores tasas de criminalidad (Heckman et al., 2010a; 2010b).

Cuando la aleatorización ex ante no es posible, a veces tal aleatorización se produce de forma natural. Es lo que se conoce como experimentos naturales. Son situaciones de la vida real que, adecuadamente analizadas, replican las condiciones de un experimento controlado. En ese sentido, un buen economista empírico en evaluación es aquel que es capaz de identificar grupos de control y grupos tratados que estén aleatorizados previamente.

Un ejemplo de este tipo, aunque en el ámbito de la salud, lo encontramos en la identificación de las causas del cólera a mediados del siglo XIX por John Snow, padre de la epidemiología, en la ciudad de Londres. De manera muy resumida (Coleman, 2019) podemos decir que Snow consiguió, mediante un experimento natural, demostrar que el cólera se transmitía por el agua y no por el aire como popularmente se creía. Con ese fin elaboró un mapa de casos mortales de la ciudad y observó que un elevado número de ellos se concentraban alrededor de la fuente de agua de Broad Street. El problema es que esos ciudadanos bebían de la fuente pero también se agrupaban para recoger el agua. Observó que en muchas calles el número de muertes difería claramente entre los habitantes de uno y otro lado de la calle, lo que no tenía sentido si el contagio se producía a través del aire. También observó que había dos compañías en Londres distribuyendo agua proveniente

del río Támesis pero desde lugares distintos y que cada compañía construyó su propia infraestructura (tuberías, etc.) mucho antes del comienzo de la epidemia, decidiendo a qué calles proporcionaba agua de forma aparentemente aleatoria. Esto significa que había vecinos que vivían uno enfrente del otro pero que recibían agua de orígenes distintos. En definitiva, contaba con un grupo tratado, el grupo con agua de una empresa, y un grupo de control, el que recibía agua de la otra empresa. Ambos grupos expuestos al mismo aire pero con agua de distinta procedencia con una diferencia notable en mortalidad media y cuyo origen tendría que ser el efecto del tratamiento (agua contaminada). De este modo consiguió demostrar la causa del cólera sin sesgo con unas condiciones similares a las de un experimento pero sin aleatorización previa, es decir, mediante lo que hoy conocemos como un experimento natural.

En la mayoría de los casos no es posible ni la aleatorización *ex ante* ni el experimento natural por lo que ha de recurrirse a métodos denominados cuasi-experimentales que permiten identificar efectos causales incluso cuando existen diferencias no observables entre los grupos tratado y de control, siendo los más comunes las diferencias en diferencias, regresión discontinua y las variables instrumentales<sup>3</sup>. La elección entre estos métodos dependerá tanto de las características de la política a evaluar como del tipo y calidad de los datos disponibles.

El método de diferencias en diferencias es el apropiado cuando evaluamos una política pública aplicada a individuos en diferentes momentos del tiempo. Se trata de comparar la diferencia en la variable objetivo entre ambos grupos (tratado y control) antes y después de llevar a cabo la política para identificar el efecto causal restando las diferencias no atribuibles a la política de las diferencias entre uno y otro grupo. Su aplicación exige la utilización de datos longitudinales lo que explica que gran parte de los estudios empíricos que utilizan esta técnica se hayan llevado a cabo en países desarrollados, donde este tipo de datos están disponibles gracias a la realización periódica de evaluaciones nacionales a gran escala o donde existen registros administrativos de carácter censal<sup>4</sup>.

---

<sup>3</sup> Cuando existen factores observables que diferencian a los dos grupos (tratamiento y control) se utilizan los modelos de regresión y los de emparejamiento (*matching*). Excelentes descripciones de los métodos que habitualmente se utilizan para analizar efectos causales de políticas públicas con sus fundamentos, ventajas e inconvenientes se encuentran en Artés y Rodríguez (2022) y en el manual de Angrist y Pischke (2014).

<sup>4</sup> En Estados Unidos existe una Evaluación Nacional del Progreso Educativo (NAEP por sus siglas en inglés) desde 1969. En Europa, los países nórdicos cuentan con registros administrativos desde mediados



Un ejemplo puede encontrarse en el trabajo de Machin y McNally (2008) que aplican esta técnica para evaluar el impacto de una hora semanal adicional de lectura introducida en las escuelas de enseñanza primaria de Inglaterra a finales de los años noventa, llegando a la conclusión de que tuvo un impacto positivo en los resultados escolares.

El método de regresión discontinua se aplica en aquellas situaciones en las que la asignación de los individuos al grupo de tratados por recibir la política depende del hecho de superar un umbral determinado de una variable continua como serían la nota académica, la renta o la edad. Se supone que aquellos individuos ligeramente por encima de ese corte son muy parecidos a los que quedan escasamente por debajo (grupo de control) por lo que, en el límite, estar por encima o por debajo, llega a ser aleatorio y su comparación permitiría identificar efectos causales. Dada la exigencia de disponer de una abundante información alrededor del punto de corte, no es extraño que la mayor parte de trabajos que emplean este enfoque utilicen datos de tipo administrativo.

Los ejemplos en políticas educativas son muy diversos pero donde más se ha utilizado tal vez sea en analizar el número de alumnos por clase aprovechando normas que obligan a desdoblarse grupos a partir de un tamaño máximo. En un trabajo reciente, Angrist et al. (2019), aplicaron este método al caso de escuelas públicas de Israel y no identificaron ningún efecto significativo sobre los resultados escolares de esta política de reducción del tamaño de la clase.

La técnica de variables instrumentales no es tan exigente en cuanto a los datos, aunque en ella resulta clave identificar un instrumento adecuado para eliminar el problema de endogeneidad entre el tratamiento y la variable objetivo. Se trata de encontrar una variable adicional (instrumento) que afecte a la probabilidad del tratamiento pero que, al mismo tiempo, no afecte a la variable objetivo. La función del instrumento es proporcionar un elemento de variación exógena de la variable endógena que permita identificar causalmente cómo afectan los cambios de esta en la variable dependiente.

El ejemplo típico en educación es el trabajo de Angrist y Kruger (1991) en el que, al analizar el efecto de los años de escolarización en los salarios, proponen como instrumento la fecha de nacimiento. Evidentemente, el instrumento afecta a los años de

---

de los años sesenta; en Reino Unido existe una base de datos censal (NPD) desde mediados de los noventa y en Italia (INVALSI) desde 2005.

escolarización ya que no es lo mismo un niño nacido en enero que en diciembre y, sin embargo, esa fecha no afecta a la variable objetivo. En su análisis para EE.UU. los autores obtienen un efecto positivo y significativo entre años adicionales de escolarización y el salario futuro de los individuos.

### **Evaluación de impacto y resultados escolares en España**

Los especialistas en Economía de la Educación utilizan la función de producción educativa como instrumento para organizar las distintas categorías de factores que pueden afectar al rendimiento escolar. Propuesta por Levin (1974), esta función incluye desde los inputs escolares, como el profesorado o los recursos físicos con los que cuentan los centros, hasta las características socioeconómicas del alumno, pasando incluso por elementos de carácter institucional como el carácter público o privado de los centros o su grado de autonomía.

El interés por la función de producción educativa tiene su origen en las polémicas conclusiones obtenidas en el conocido Informe Coleman (Coleman et al., 1966) cuyo objetivo fue comprobar la importancia de los recursos escolares, principalmente en las clases más desfavorecidas. La investigación utilizó datos de más de 700.000 estudiantes de Estados Unidos y su conclusión fue que esos recursos explicaban una parte muy pequeña de las diferencias registradas en los niveles de desempeño de los alumnos en comparación con la importancia de los factores relacionados con el entorno socioeconómico. Posteriormente, Hanushek (1986), en otro de los trabajos fundamentales dedicados a la función de producción educativa, llegó a una conclusión similar, destacando la escasa importancia de ese tipo de recursos.

El análisis empírico de la función de producción educativa se ha visto impulsado tanto por el desarrollo de las técnicas de evaluación, comentadas en el apartado anterior, como por la disponibilidad de datos derivados, en parte, de la generalización de evaluaciones estandarizadas internacionales de los sistemas educativos impulsadas fundamentalmente por dos instituciones, la IEA (*International Association for the Evaluation of Educational Achievement*) y la OCDE (*Organisation for Economic Cooperation and Development*)<sup>5</sup>.

---

<sup>5</sup> La IEA lleva coordinando estudios internacionales a gran escala desde los años sesenta y es la responsable de las evaluaciones TIMSS (*Trends in Mathematics and Science Study*) y PIRLS (*Progress in International Reading Literacy Study*) centradas en la educación primaria. La OCDE comenzó con estas evaluaciones a finales de los años noventa, promoviendo el estudio PISA (*Programme for International Student*

Cuando descendemos a la evaluación de políticas educativas en España, nuestra asignatura pendiente es la utilización de experimentos aleatorios controlados. A continuación, presento algunas de las pocas excepciones existentes; en concreto, el estudio piloto desarrollado por el Banco de España para evaluar programas de educación financiera, el proyecto ACT de la Unión Europea y las iniciativas de la Fundación la Caixa de evaluación de programas de innovación educativa.

El programa “Finanzas para todos” del Banco de España y la Comisión Nacional del Mercado de Valores, aplicado con generalidad en nuestro país, permitió diseñar un experimento aleatorio de un curso de 10 horas sobre finanzas personales impartido en la educación secundaria obligatoria (tercer curso de la ESO) a 3.000 alumnos de 78 centros que solicitaron participar en el programa (Bover et al., 2018) y que fueron divididos en dos grupos: tratados y control. El contenido del curso abarcó la elaboración de presupuestos, las relaciones bancarias y los vehículos de ahorro, así como la conciencia sobre los resultados futuros.

La comparación entre los resultados de los estudiantes pertenecientes a ambos grupos mostró que la participación en el curso aumentó las puntuaciones en las pruebas de conocimiento financiero un sexto de la desviación estándar. Los jóvenes tratados tenían más probabilidades de involucrarse en asuntos financieros en el hogar y mostraron un mayor grado de paciencia en las opciones hipotéticas de ahorro. Los mejores resultados son estadísticamente significativos en los estratos con estudiantes con antecedentes relativamente más desfavorecidos. Sin embargo, la participación en el programa no mejoraba otras cuestiones como la elaboración de un presupuesto y la capacidad de ajustarse al mismo.

El objetivo del proyecto ACT (*Active Citizenship Project to enhance pupils' social and civic competences*), financiado por el Programa Erasmus + de la Unión Europea, es profundizar en la competencia ciudadana de los estudiantes de Educación Secundaria mediante el desarrollo de los valores cívicos. Se trata de un programa experimental a gran escala en el que intervinieron Francia, Reino Unido, Grecia y España en los que estaban implicados sus correspondientes Ministerios de Educación y equipos de evaluación independientes procedentes de prestigiosos centros de investigación de cada uno de esos

---

Assessment) en educación secundaria y PIAAC (*Programme for the International Assessment of Adult Competencies*) referida a competencias en adultos.

países<sup>6</sup> (Briole et al., 2022). El experimento social consta de dos intervenciones. Una, de formación a los profesores en prácticas pedagógicas interactivas centradas en los alumnos y relacionadas con un marco teórico común en competencia cívica. Otra, la puesta en práctica por parte de esos profesores de un proyecto de ciudadanía en sus centros en el que los alumnos tenían un papel importante tanto en su diseño como en su ejecución.

El proyecto se basó en una metodología experimental. En primer lugar, mediante una intervención piloto en tres centros educativos de cada país durante el curso 2017-2018, se puso a prueba tanto la formación como su implementación en el aula. Posteriormente, en el curso 2018-19, el experimento se generalizó a 312 centros de Educación Secundaria de los cuatro países participantes. Entre los centros participantes aleatoriamente se seleccionaron los que serían tratados (sus profesores recibirían la formación e implementarían el proyecto en el aula) y los controles. La evaluación del programa se llevó a cabo tanto cualitativa como cuantitativamente mediante la distribución de cuestionarios antes y después de las intervenciones, entrevistas en profundidad tanto a profesores como a alumnos y observaciones en el aula a lo largo del curso escolar mencionado. El resultado global, desde un punto de vista cuantitativo, es que no se observó efectos en Grecia ni en España, mientras que los resultados del Reino Unido son difíciles de interpretar por el elevado abandono que se registró. Por el contrario, en Francia se observó un efecto positivo y significativo sobre las competencias cívicas de los alumnos. Además, en los tres países con datos fiables, el programa resultó ser más eficaz para los alumnos que inicialmente mostraban una mayor competencia ciudadana y participación democrática, aumentando aún más las diferencias de participación democrática y actitudes cívicas entre los representantes estudiantiles y el resto de alumnos.

El proyecto “Evaluación de programas educativos” de EduCaixa, iniciado en 2018, constituye otro ejemplo relevante de experimentos aleatorios en educación escolar. Su objetivo es fomentar la evaluación de intervenciones e innovaciones educativas mediante el contacto entre centros educativos con propuestas educativas innovadoras y equipos de investigación<sup>7</sup>. Las convocatorias constan de cuatro fases. En la primera fase, los centros presentan sus solicitudes de participación exponiendo el programa educativo que desean

---

<sup>6</sup> Ecole d'Économie de Paris, London School of Economics and Political Sciences, University of Athens y Fundación Europea, Sociedad y Educación, respectivamente.

<sup>7</sup> Para más información véase <https://educaixa.org/es/evaluacion-programas-educativos>

evaluar. En la segunda fase, se seleccionan tanto las intervenciones educativas como los equipos de investigación encargados de realizar las evaluaciones. En la tercera fase, a los programas seleccionados por un comité de expertos de los presentados se le asigna un grupo de investigación independiente. Los centros educativos, junto con los grupos de investigación asignados, elaboran las propuestas de implementación y evaluación de las intervenciones educativas. En la fase final, las propuestas seleccionadas reciben la financiación correspondiente que permite su implementación en los centros docentes y la correspondiente evaluación de las mismas. Algunos ejemplos de intervenciones educativas que han participado o están participando actualmente en las convocatorias son el ajedrez educativo, la digitalización de bibliotecas escolares para promover la lectura, la metodología del Aprendizaje Basado en Proyectos (ABP) o el impacto de la música en el aprendizaje y el desarrollo emocional.

Si nos centramos en trabajos que utilizan experimentos naturales o técnicas cuasi-experimentales en la evaluación del desempeño escolar en España, los ejemplos son más abundantes, aunque se trata de estudios relativamente recientes, desarrollados en los últimos siete u ocho años. A continuación, ofrecemos una panorámica en la que se incluyen los más relevantes agrupados en temas relacionados con la función de producción educativa.

Un tópico que ha atraído la atención de los investigadores es la política de repetición de curso sobre el rendimiento escolar debido a que nuestro país alcanza una de las mayores tasas de repetición de Europa (López-Rupérez et al., 2021). La obtención de resultados fiables sobre el impacto de esta medida exige tener en cuenta el problema de causalidad inversa que habitualmente existe en los datos (si los repetidores presentan peores resultados puede ser debido a la repetición o a que ya tenían un peor rendimiento antes de repetir). García-Pérez et al. (2014), mediante el uso del trimestre del nacimiento del alumno como variable instrumental, encuentran un efecto negativo sobre el rendimiento de la estrategia de repetición, siendo más perjudicial en primaria que en secundaria.

Si parece haber coincidencia entre los expertos en la poca efectividad de la política de repetición de curso para abordar un tema tan complejo como es el fracaso escolar, también hay consenso, como sucede en el ámbito internacional, sobre la eficacia de las actuaciones educativas en edades tempranas. Un par de trabajos, al menos, avalan esta política con evidencia cuasi-experimental en España. Felfe et al. (2015) analizaron el impacto que tuvo la ampliación gratuita de la educación infantil a los tres años con la implantación de

la Ley de Ordenación General del Sistema Educativo a principios de los años noventa, explotando el experimento natural que se produjo como consecuencia del distinto año de aplicación de la reforma entre las comunidades autónomas. Utilizando datos de diferentes oleadas de PISA y un enfoque de diferencias en diferencias, se comprobó que la reforma afectó positivamente a la promoción de curso durante la educación primaria y supuso una mejora en los resultados educativos en lectura y matemáticas que se mantuvo en el largo plazo. Además, este efecto fue mucho más acusado en los alumnos con peor nivel socioeconómico. Santín y Sicilia (2015) por su parte, utilizando datos de la Evaluación General de Diagnóstico de Primaria (2009), aprovechan el experimento natural derivado de que esa base de datos incluye un grupo de escuelas en las que la asignación de grupos se realiza aleatoriamente y, por tanto, existe un grupo que, en media, tiene más años de educación infantil que otro, siendo posible utilizar esa diferencia como tratamiento del experimento. Los resultados muestran que los años de educación infantil tienen un impacto positivo y significativo en el resultado de matemáticas y lectura.

El efecto de los inmigrantes en los resultados escolares ha sido abordado en Pedraja et al. (2016). Mediante un enfoque de diferencias en diferencias, analizan el impacto del incremento exógeno de alumnos inmigrantes en España producido entre 2003 y 2009 con la información ofrecida por PISA en ambos años. Debido a que el porcentaje de alumnos inmigrantes difiere de unas escuelas a otras, se introduce un tratamiento por dosis para comprobar el efecto medio de tener estudiantes inmigrantes en las escuelas y cómo afecta su acumulación a los resultados escolares. La conclusión es que la llegada de alumnos inmigrantes no reduce, en media, las tasas de promoción de curso en las escuelas y resulta incluso beneficioso para los nativos. Sin embargo, la acumulación de inmigrantes sí tiene un impacto negativo sobre los porcentajes de repetidores tanto de inmigrantes como de nativos cuando alcanza cifras superiores al 15% del total del alumnado de los centros.

Lamentablemente, no disponemos en nuestro país de estudios sobre el input escolar con mayor efecto en los resultados, los profesores o, más concretamente, la calidad del profesorado. En este sentido, Hanushek (2011) demuestra que reemplazar entre un 5 y un 8 por ciento de los profesores de peor calidad por profesores de calidad media llevaría a situar a Estados Unidos en los primeros puestos en las pruebas internacionales de matemáticas y ciencias (un valor de 100 billones de dólares). Sobre una cuestión relacionada con el profesorado, el efecto de los diferentes estilos o prácticas docentes en los resultados, sí existen algunos trabajos pero los resultados no son concluyentes.

Hidalgo y López-Mayán (2018) aplican la técnica de diferencias en diferencias a los datos proporcionados por la Evaluación General de Diagnóstico de Primaria (2009) aprovechando las diferencias en prácticas docentes utilizadas por los profesores dentro de la misma escuela. Basándose en la información proporcionada por los alumnos, detectan un mayor efecto sobre el rendimiento de las prácticas docentes modernas frente a las más tradicionales. En cambio, Cordero y Gil-Izquierdo (2018), con datos de PISA y aplicando variables instrumentales, llegan a la conclusión contraria, la superioridad de los métodos más tradicionales.

Los programas de incentivos dirigidos tanto a alumnos como a profesores cuentan con abundantes ejemplos en la literatura internacional. En el caso de España, se han implementado algunas políticas llevadas a cabo por Andalucía y Cataluña que no han sido evaluadas rigurosamente. La excepción la tenemos con el Programa 18-25 implementado por la Comunidad autónoma de Extremadura cuyo objetivo era elevar el nivel educativo de los desempleados extremeños entre 18 y 25 años ofreciendo el incentivo económico de 1.000 € a aquellos que obtuvieran el título de la enseñanza secundaria obligatoria (ESO), incrementando con ello sus posibilidades de incorporación al mercado de trabajo. Pedraja et al. (2022), mediante el uso de una regresión en discontinuidad, demuestran que la posibilidad de obtención del título de la ESO no fue estadísticamente diferente entre el grupo de tratados y el grupo de control, lo que implica que esa política de incentivos no resultó efectiva.

Una cuestión polémica que viene suscitando la atención de los grupos de interés del mundo educativo es la efectividad de las tareas escolares. En un curioso estudio aplicado a Andalucía, Jerrim et al. (2019), aplican la técnica de diferencias en diferencias a datos administrativos longitudinales procedentes de la Evaluación de Diagnóstico de esa comunidad a una muestra formada por alumnos gemelos. Lo característico de la muestra es que les permite controlar por todo lo que tienen en común (características de la escuela, padres, antecedentes socioeconómicos, etc.). El principal resultado es que el tiempo dedicado a los deberes no parece contribuir significativamente a la mejora del rendimiento académico.

Para finalizar, una cuestión que ha suscitado interés, acrecentado por la actual pandemia, ha sido el impacto de la incorporación de las TIC en las escuelas mediante diferentes programas impulsados tanto por el Ministerio de Educación como por las Comunidades Autónomas. Lamentablemente, la mayoría de estas iniciativas no obedecieron a un plan

consensuado para todo el territorio nacional ni establecieron un criterio de evaluación riguroso, por lo que resulta muy difícil comprobar si resultaron eficaces. La excepción ha sido el programa Escuela 2.0, cuyo objetivo era mejorar la dotación de ordenadores de los centros educativos, para el que existen dos trabajos que han analizado su impacto mediante la aplicación de técnicas de inferencia causal a los datos disponibles en PISA (Villaplana, 2014; Cabras y Tena, 2016). Ambos concluyen que el efecto fue moderadamente positivo y concentrado únicamente en los casos que esa política viniera acompañada con cambios en la metodología docente del profesorado.

Relacionado con la pandemia, en un reciente estudio (Arenas y Cortazar, 2022) se presentan las primeras estimaciones de pérdida de aprendizaje por el cierre de las escuelas con datos de evaluaciones externas del Instituto Vasco de Evaluación e Investigación Educativa (ISEI-IVEI)<sup>8</sup>. El ejercicio empírico sigue una técnica de diferencias en diferencias midiendo las competencias antes y después de la pandemia. Las estimaciones muestran una importante pérdida de aprendizaje estadísticamente significativa en matemáticas y euskera, aunque no en castellano. Además, la pérdida de aprendizaje es más heterogénea en función de las características de los centros que de los alumnos, siendo los centros públicos los que han sufrido especialmente los efectos de la pandemia.

## **Epílogo**

Cierro este discurso con tres recomendaciones.

La primera se sitúa en el ámbito institucional de las políticas públicas y tiene que ver con la difusión y la utilización de los resultados de las evaluaciones. De nada sirve la realización de evaluaciones rigurosas si no se difunden con claridad a todos los niveles y si sus resultados no afectan al diseño de las futuras políticas públicas. Son necesarios compromisos políticos con relación a la elaboración de registros de resultados y evidencias de las evaluaciones. En ese sentido, todas las evaluaciones deberían incluir un plan de difusión de resultados y de utilización de los mismos. También es imprescindible que las propuestas de políticas públicas que se hagan incorporen en su diseño las evidencias que existen en ese campo. La panorámica presentada en el caso de evaluaciones de impacto en relación a los resultados escolares, ofrece interesantes

---

<sup>8</sup> La pérdida de aprendizaje es un tema muy relevante por las significativas consecuencias económicas y sociales que puede provocar. Hanushek y Woessman (2020) estiman que una pérdida de aprendizaje del 25% de un año académico sin pandemia reduce el PIB per cápita a largo plazo en un 1,9%.



evidencias sobre cuestiones clave como son, entre otras, la repetición de curso, la actuación en edades tempranas, la política de incentivos o el uso de las TIC.

La segunda es una cautela en relación con las consecuencias no intencionadas de las políticas públicas. Un repaso de esos efectos en sanidad y educación se encuentra en el trabajo de Graves y Zucchelli (2022). En políticas educativas se insiste en los incentivos que se producen en las respuestas a determinados objetivos (resultados en los exámenes) con la reducción del tiempo dedicado a desarrollar contenidos o capacidades no evaluadas tanto o más importantes. También se llama la atención sobre los problemas de implementar políticas a gran escala a partir de los resultados obtenidos al aplicar una política en un ámbito reducido, como por ejemplo la reducción del número de alumnos por aula cuyos efectos beneficiosos acaban más que contrarrestándose con una pérdida de calidad del profesorado que ha de contratarse masivamente para aplicar esa política. Hemos de ser cuidadosos con las cláusulas *ceteris paribus* tan habituales en Economía como inapropiadas al evaluar políticas de gran alcance que se aplican con generalidad a toda la población. Como reflexión final, se aboga por una evaluación periódica de las políticas teniendo en cuenta que buena parte de esos efectos se producen pasados bastantes años desde que las políticas públicas se aplican.

La tercera, relacionada con la segunda recomendación, llama la atención sobre el output educativo que normalmente se mide en las pruebas nacionales e internacionales, el output cognitivo que es relevante, pero resulta incompleto ya que deja fuera otra parte esencial de la función de la escuela, la dimensión no cognitiva; es decir, valores afectivos, comportamiento personal, desarrollo social, artístico, etc. Heckman (2008), premio nobel de Economía, en su trabajo con psicólogos y neurólogos, llama la atención sobre las habilidades no cognitivas demostrando, con evidencia experimental, como estas son tan importantes e incluso superiores a las habilidades cognitivas en sus efectos sobre los salarios y la productividad de la economía. Cunha y Heckman (2010), concluyen su trabajo confirmando además el impacto positivo del ámbito no cognitivo sobre el cognitivo (la evidencia en la dirección contraria es muy débil). En definitiva, utilizar medidas de resultados académicos para predecir la productividad y el crecimiento económico lleva a conclusiones exageradas (sesgadas) si omitimos las variables no cognitivas en educación.

## Bibliografía

- AIReF (2020). *Opinión para una estrategia de acceso a datos administrativos*. Opinión. <https://www.airef.es/es/opiniones/estrategia-acceso-datos-administrativos/>
- Albi, E. y Onrubia, J. (2018). *Supervisión, control y evaluación para una gestión pública transparente*. Madrid: Instituto de Estudios Fiscales.
- Almunia, M. y Rey Biel, P. (2021). La gestión de los datos administrativos en España: diagnóstico y retos de futuro. *Cuadernos Económicos del ICE*, 102, 103-131.
- Angrist, J. D. Y Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4), 979-1014.
- Angrist, J. D. y Pischke, J. S. (2014). *Mastering metrics: The path from cause to effect*. Princeton University Press.
- Angrist, J. D., Lavy, V., Leder-Luis, J. y Shany, A. (2019). Maimonides' rule redux. *American Economic Review: Insights*, 1(3), 309-24.
- Arenas, A. y Gortazar, L. (2022). Learning loss One Year After School Closures: Evidence from the Basque Country. *Working Paper #1*, Esade Center for Economic Policy.
- Artés, J. y Rodríguez-Sánchez, B. (2022). Métodos de evaluación de políticas públicas, *Papeles de Economía Española*, 172 (en prensa).
- Bover, O., Hospido, L. y Villanueva, E. (2018). The impact of high school financial education on financial knowledge and choices: Evidence from a randomized trial in Spain. *IZA Discussion Papers*, N° 11265, Institute of Labor Economics (IZA).
- Briole, S., Gurgand, M., Maurin, É., McNally, S., Ruiz-Valenzuela, J., y Santín, D. (2022). The Making of Civic Virtues: a school-based experiment in three countries. *IZA Discussion Papers*, N° 15141, Institute of Labor Economics (IZA).
- Cabras, S. y Tena, J. D. (2016). A Bayesian non-parametric modeling to estimate student response to ICT investment. *Journal of Applied Statistics*, 43(14), 2627-2642.
- Casado, J. M., y del Pino, E. (2021). Evolución, situación actual y retos de la evaluación de políticas públicas en las Administraciones españolas (2000-2021). *Cuadernos Económicos del ICE*, 102, 13-38.
- Casado, J. M., y del Pino, E. (2022). Similitudes y diferencias en la evaluación de políticas públicas en ocho países: España en perspectiva comparada. *Papeles de Economía Española*, 172 (en prensa).
- Coleman, J., Campbell, E., Hobson, C., Mc Partland, I., Modd, A., Wenfeld, F. y York, R. (1966). *Equality of educational opportunity*, Department of Health, Education and Welfare, Office of Education, Government Printing Office, Washington.
- Coleman, T. S. (2019). Causality in the Time of Cholera: John Snow as a Prototype for Causal Inference. *Working Paper*, Version 1.2. Chicago University Press. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3262234](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3262234)

- Cordero, J. M. y Gil-Izquierdo, M. (2018). The effect of teaching strategies on student achievement: An analysis using TALIS-PISA-link. *Journal of Policy Modeling*, 40(6), 1313-1331.
- Cunha, F., Heckman, J. y Schennach, S. (2010): Estimating the Technology of Cognitive and Noncognitive Skill Formation, *Econometrica*, 78 (3), 883-931.
- FEDEA, Fundación de Estudios de Economía Aplicada. (2021). La evaluación de políticas públicas en España: antecedentes, situación actual y propuestas para una reforma. *FEDEA Policy Papers 2021/09*.
- Feinstein, O. y Zapico-Goñi, E. (2010): Evaluation of Government Performance and Public Policies in Spain. *ECD Working Paper Series 22*. World Bank. Independent Evaluation Group.
- Felfe, C., Nollenberger, N. y Rodríguez-Planas, N. (2015). Can't buy mommy's love? universal childcare and children's long-term cognitive development. *Journal of Population Economics*, 28(2), 393-422.
- García-Pérez, J. I., Hidalgo-Hidalgo, M. y Robles-Zurita, J. A. (2014). Does grade retention affect students' achievement? Some evidence from Spain. *Applied Economics*, 46, 1373-1392.
- Graves, J. y Zucchelli, E. (2022). Consecuencias no intencionadas de las medidas de política pública: ejemplos de sanidad y educación, *Papeles de Economía Española*, 172 (en prensa).
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141-1177.
- Hanushek, E.A., 2011. The economic value of higher teacher quality. *Economics of Education Review*, 30(3), 466-479.
- Hanushek, E. A. y Woessmann, L. (2011). The economics of international differences in educational achievement, en Hanushek, E., Machin, S. y Woessmann, L. (eds.) *Handbook of the Economics of Education*, Vol. 3, 89-200. Amsterdam: North Holland.
- Hanushek, E.A y Woessmann, L. (2020). The economic impacts of learning losses. *Education Working Papers*, No. 225. OCDE
- Heckman, J. (2008): Schools, skills and synapses, *Economic Inquiry*, 46, 289-324.
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. y Yavitz, A. (2010a). The rate of return to the High Scope Perry Preschool Program. *Journal of Public Economics*, 94(1-2), 114-128.
- Heckman, J., Moon, S. H., Pinto, R., Savelyev, P. y Yavitz, A. (2010b). Analyzing social experiments as implemented: A reexamination of the evidence from the High Scope Perry Preschool Program. *Quantitative Economics*, 1(1), 1-46.
- Hidalgo, A. y Lopez-Mayan, C. (2018). Teaching styles and achievement: Student and teacher perspectives. *Economics of Education Review*, 67, 184-206.
- Jerrim, J., Lopez-Agudo, L. A. y Marcenaro-Gutierrez, O. D. (2019). The relationship between homework and the academic progress of children in Spain during compulsory elementary education: A twin fixed-effects approach. *British Educational Research Journal*, 45(5), 1021-1049.
- Jourmard, I. y Giorno, C. (2005): "Getting the most out of public sector decentralization in Spain", *Working Paper 23*, OCDE.

- Levin, H. M. (1974): Measuring Efficiency in educational production, *Public Finance Quarterly*, 2, 3-24.
- López-Rupérez, F., García-García, I. y Expósito, E. (2021). La repetición de curso y la graduación en Educación Secundaria Obligatoria en España: análisis empíricos y recomendaciones políticas. *Revista de Educación*, 394, 325-353.
- Machin, S., y McNally, S. (2008). The literacy hour. *Journal of Public Economics*, 92(5-6), 1441-1462.
- Martínez-Vázquez, J. (2018). Perspectives on the last quarter century of research in Public Economics, *Revista de Economía Aplicada*, 76 (XXVI), 9-33.
- Oates, W. E. (1999). An essay on fiscal federalism. *Journal of Economic Literature*, 37(3), 1120-1149.
- Pedraja, F., Santín, D. y Simancas, R. (2016). The impact of immigrant concentration in schools on grade retention in Spain: a difference-in-differences approach. *Applied Economics*, 48(21), 1978-1990.
- Pedraja, F., Santín, D. y Simancas, R. (2022). Show me the money! The impact of a conditional cash transfer on educational achievement. *Empirical Economics*, en prensa. <https://doi.org/10.1007/s00181-022-02211-x>
- Santín, D. y Sicilia, G. (2015). El impacto de la educación infantil en los resultados de primaria: evidencia para España a partir de un experimento natural. *Reflexiones sobre el Sistema Educativo Español*, Fundación Europea Sociedad y Educación.
- Schlottter, M., Schwerdt, G. y Woessmann, L. (2011). Econometric methods for causal evaluation of education policies and practices: a non-technical guide. *Education Economics*, 19(2), 109-137.
- Stockmann, R., Meyer, W. y Taube, L. (eds.) (2020). *The Institutionalization of Evaluation in Europe*. Palgrave MacMillan.
- Villaplana, C. (2014). Evaluación del programa Escuela 2.0 para la asignatura de Matemáticas a partir de PISA 2012. *Investigaciones de Economía de la Educación*, 9, 631-652.
- Wolfe, B. y Zuvekas, S. (1997). Non-market effects of education. *International Journal of Education Research*, 27(6), 491-502.