

# LA EDUCACIÓN COMO INSTRUMENTO PARA CONTENER LA DESIGUALDAD SOCIAL

Antonio CABRALES\*

## Resumen

*En este artículo discutiré la utilización de intervenciones educativas como herramienta para mitigar o reducir el aumento de la desigualdad que se ha producido en muchas sociedades en las últimas décadas. Esta discusión está precedida por una introducción metodológica en la que se explica por qué son necesarios los ensayos controlados aleatorizados para tener evidencias firmes sobre las que basar las políticas públicas. Tras la introducción metodológica se discuten en detalle cinco grupos de ensayos controlados, y sus resultados.*

## Palabras clave

*Desigualdad social, intervenciones educativas, ensayos controlados aleatorizados, políticas públicas basadas en la evidencia.*

## Abstract

*In this article I will discuss the use of educational interventions as a tool to mitigate or reduce the increase in inequality that has occurred in many societies in recent decades. This discussion is preceded by a methodological introduction that explains why randomized controlled trials are necessary to have consistent evidence on which to base public policy. After the methodological introduction five groups of controlled trials and their results are discussed in detail.*

## Key words

*Social inequality, educational interventions, randomized controlled trials, evidence-based public policies.*

---

\* Department of Economics, University College London. Correo electrónico: a.cabrales@ucl.ac.uk

SUMARIO: I. Introducción; II. Política basada en la evidencia y ensayos aleatorios controlados; III. Intervenciones educativas; 1. El Quantum Opportunity Program; 2. Los programas Perry Preschool y Abecedarian; 3. El programa preescolar de Tennessee; 4. Escuelas «Sin Excusas»; 5. Los experimentos de Edlabs, incentivos financieros al estudio; IV. Conclusión; V. Bibliografía.

## I. INTRODUCCIÓN

A pesar de lo que a veces escuchamos, la desigualdad de la renta y la pobreza en el mundo en su conjunto ha disminuido en los últimos 30 años, como muestra por ejemplo Branko Milanovic (2013). Esto no quiere decir que la desigualdad no sea un grave problema, una buena parte del decrecimiento de la desigualdad se debe a que algunos países, notablemente China y la India, han aumentado mucho su renta y reducido el diferencial con respecto a Europa y Norteamérica. Sin embargo, en muchos países la proporción de la renta y la riqueza en manos del 1% más rico de la población ha aumentado de manera muy marcada en las últimas décadas, tal como documenta el volumen colectivo editado por Anthony Atkinson y Thomas Piketty (2010).

A la vista de este aumento de la desigualdad dentro de muchos países, resulta natural preguntarse por maneras de mitigarla o revertirla. Lógicamente la redistribución de la renta por vía fiscal debe ser un instrumento prioritario, y hay países como Holanda, en el que la desigualdad no ha aumentado y han podido mantenerse competitivos en el comercio internacional en buena medida a través de herramientas impositivas y de gasto público.

Pero esto no quiere decir que no se puedan utilizar otros instrumentos. En particular, la educación puede ser una herramienta apropiada. Hay que notar, por ejemplo, que la diferencia salarial entre graduados universitarios y personas con menor nivel de estudios ha aumentado de forma paralela a la desigualdad, como señalan Claudia Goldin y Lawrence Katz (2007) para Estados Unidos o François Bourguignon y Halsey Rogers (2008) resumiendo la literatura de un amplio grupo de países.

El propósito de este artículo es mostrar que la educación puede ser efectivamente un instrumento adecuado en la lucha contra la desigualdad social y la pobreza, pero que es una herramienta que debe utilizarse con cuidado, que no siempre funciona bien, y se corre el riesgo de desperdiciar recursos fiscales que pueden tener mejores usos. Para demostrarlo discutiré un conjunto de estudios sobre distintas intervenciones destinadas a mejorar los resultados escolares y sociales de jóvenes con un bajo nivel socioeconómico. Antes de revisar estos estudios haré una breve introducción metodológica que permite comprender mejor esos estudios.

## II. POLÍTICA BASADA EN LA EVIDENCIA Y ENSAYOS ALEATORIOS CONTROLADOS

Dado que vamos a hablar mucho en este artículo de políticas basadas en la evidencia, me parece interesante comenzar con una breve introducción metodoló-

gica y una motivación histórica. La medicina es la primera ciencia que utilizó de manera sistemática los ensayos controlados aleatorios para obtener evidencias sobre las que basar las intervenciones recomendadas y me parece que una buena parte de nuestras políticas sociales, sobre todo las educativas, se beneficiarían de una mayor atención a esta metodología.

El Washington Post, en su obituario sobre Paul Meier (escrito por David Brown, 2011), un famoso bioestadístico fallecido recientemente y considerado el padre de la moderna medicina basada en la evidencia nos decía que éste fue un «temprano y exitoso defensor de la introducción de ensayos clínicos aleatorizados. La idea de asignar a los sujetos en un experimento médico a un tratamiento u otro exclusivamente sobre la base del azar horrorizaba a muchos médicos». Pero como también dice el artículo, el uso de esta metodología ha salvado miles de vidas.

Si a muchos médicos les horrorizaba, es natural que sea aún peor para cualquier ciudadano que no está acostumbrados a tomar decisiones de vida o muerte en su experiencia cotidiana. De ahí que parezca oportuno entender por qué la metodología es tan útil. Técnicamente la razón es que el análisis estadístico suele resultar muy comprometido en presencia de variables omitidas; y los experimentos aleatorios evitan justamente ese problema. O más brevemente: correlación no implica causalidad, excepto si hay un experimento de por medio.

La mejor manera de explicarlo es con un ejemplo. Desde hace unos años los exámenes CDI (Conocimientos y Destrezas Indispensables) de la Comunidad de Madrid incluyen un cuestionario para los alumnos, y en este cuestionario una pregunta sobre el tiempo que los padres pasan haciendo deberes con los hijos. En el trabajo preparatorio de un artículo reciente (Brindusa Anghel y Antonio Cabrales 2014) hicimos unas regresiones relacionando los resultados de los alumnos con algunas variables asociadas a los mismos. En esas regresiones, el coeficiente de la variable «tiempo de los padres» sobre el resultado en los exámenes era negativo. Es posible que el esfuerzo de un padre con sus hijos no sea muy útil, pero ¿negativo? En este punto, el adagio «correlación no es causalidad» viene en nuestro rescate. El problema es que, aunque ese adagio nos permita mantener nuestra autoestima, no resulta muy útil para saber cuál es el efecto real de la variable «esfuerzo de los padres».

Un experimento aleatorio en el cual un grupo de padres escogidos al azar pasaran una hora al día con sus hijos y otro grupo no lo hiciera podría resolver el problema. La cuestión es: ¿por qué? Vamos a volver al ejemplo, y para simplificar imaginemos que hay sólo dos tipos de alumnos. Uno de ellos va a sacar un 7 de media sin ayuda de los padres. El otro va a sacar un 4 sin ayuda. En realidad da igual la causa, puede ser que tengan mejores genes o que hayan tenido un entorno infantil más cuidado, las dos cosas juntas o incluso una tercera variable omitida (una cuidadora parecida a Mary Poppins). Lo crucial es que el investigador no puede observar qué tipo de niño es cada uno. Ahora supongamos que una hora de tutoría paterna aumenta la nota media en dos puntos, sea cual sea el tipo del niño.

El paso crucial es darse cuenta de que, en ausencia de experimento controlado, el esfuerzo de los padres es *endógeno*. Y es posible que muchos padres, para simplificar digamos que todos, solamente hacen esfuerzo si prevén que el hijo no va a aprobar. Y los padres, al revés que el investigador, sí que saben si tienen un

hijo de 7 o uno de 4. Así que a falta de experimento la media de los chicos «de 4» acaba siendo un 6. Y la media de los «de 7» acaba siendo 7. Con lo cual el investigador verá un grupo de chicos a los que el padre ayuda y tienen un 6 de media; y otro grupo a los que el padre no ayuda y tienen un 7. Por consiguiente aquellos estudiantes a los que su padre ayuda tienen peor nota; que es lo que en el fondo señala el coeficiente de la regresión, una media condicionada a la ayuda de los padres más baja.

Esto explica por qué la correlación no es causalidad, pero todavía no he explicado por qué el experimento ayuda a solucionar el problema. La razón es que el agrupamiento aleatorio de los padres para la hora de ayuda nos evita el problema de que no conozcamos si los chicos de cada grupo son de un tipo u otro. Es fácil darse cuenta de que si el número de chicos «de 7» y «de 4» que caen en el grupo tratado es igual que los del grupo de control, la diferencia de medias entre los dos grupos será exactamente de 2. Y también que lo que determina la diferencia de medias entre los grupos es la proporción relativa de chicos «de 7» y «de 4» en cada grupo. Y justamente lo que garantiza la asignación aleatoria a cada grupo es que esa diferencia no será grande. Para hacerlo más visual imaginen que tenemos 40 chicos, 20 de cada tipo, y los ponemos aleatoriamente en un grupo de 20 que recibe ayuda y otro de 20 que no la recibe. Para que el grupo que recibe ayuda tenga una nota media más baja, tendría que haber 17 o más de los «de 4» en el grupo tratado. La probabilidad de que esto pase al azar es del 1,3 por mil si no nos engaña la tabla de la distribución binomial.

Espero que haya quedado claro que los experimentos aleatorios son una buena idea para conocer el efecto de políticas variadas. Pero además resulta que son un buen negocio. A todos se nos pueden ocurrir ideas aparentemente magníficas para mejorar la calidad de la enseñanza. Puede la llamada beca de 6000 que pagaba la Junta de Andalucía a los padres de familias desfavorecidas si sus hijos no abandonaban los estudios secundarios; la beca de mantenimiento del ministerio de educación para evitar el abandono escolar; el bachillerato de excelencia de la Comunidad de Madrid o los desdobles de clases que se han propuesto para fomentar la excelencia. Esto por mencionar solamente dos de los problemas más acuciantes de nuestra enseñanza; el fracaso escolar y la manifiesta falta de excelencia de nuestro sistema. Pero, que yo sepa, ninguno de estos programas se ha introducido con un protocolo experimental.

No todo está perdido. Algunos de estos programas pueden evaluarse de forma cuasi-experimental. Por ejemplo, la beca 6000 o la de mantenimiento tenían unos requisitos de renta máxima familiar para poder acogerse a los mismos; de manera similar el ingreso en el bachillerato de excelencia es competitivo. En la medida que estar justo por encima o por debajo del corte para disfrutar estos programas sea esencialmente aleatorio, una regresión de discontinuidad nos permite evaluarlos con ciertas garantías. El problema es que si estos programas tienen efectos muy distintos en el estudiante marginal que en otros puntos de la distribución, no habremos aprendido lo suficiente sobre sus efectos. Por tanto, en un contexto de severa limitación de los ingresos públicos sería bueno que aplicáramos los mismos criterios de rigor a la introducción de todo tipo políticas que a la introducción de medicamentos, siguiendo las ideas pioneras de Paul Meier.

### III. INTERVENCIONES EDUCATIVAS

#### 3.1 EL QUANTUM OPPORTUNITY PROGRAM

El primer experimento destinado a mitigar los efectos de las desigualdades sociales que quiero discutir es el Quantum Opportunity Program. Y la discusión está basada en un artículo de Núria Rodríguez-Planas (2012). Aunque es un programa de gran calidad, como veremos, los efectos en los adolescentes tratados tienen una corta duración y no se mantienen en el largo plazo, excepto para los más jóvenes. Otra evidencia que comentaré en siguientes epígrafes sugiere que las intervenciones en niños más jóvenes pueden ser duraderas y muy rentables desde el punto de vista social.

El Quantum Opportunity Program es una intervención de cinco años de duración sobre jóvenes estudiantes de secundaria (de unos 14 años al comenzar). El programa incluye un refuerzo educativo, un monitor para los chicos e incentivos monetarios para la participación en el programa. Este programa duraba cinco años y sus efectos se midieron en su último año, a los tres y a los cinco años de acabar. Para poder evaluarlo de manera rigurosa, se siguió al grupo intervenido y a otro grupo de control, y la participación en uno u otro grupo se seleccionó al azar.

La intervención se centraba en estudiantes con malas calificaciones en escuelas con problemas. La parte más innovadora del programa era que a cada estudiante tratado se le asignaba un monitor experimentado que debía desarrollar con él una relación de largo plazo durante los cinco años que duraba la intervención. Este monitor, cada uno de los cuales tenía solamente entre 15 y 25 chicos a su cargo, debía proporcionar el tipo de atención que un pariente interesado suele otorgar y de la que estos chicos normalmente carecen. Esta relación debía mantenerse incluso si el chico dejaba el programa, abandonaba la escuela, era encarcelado o se mudaba fuera de su barrio o ciudad.

Las actividades propiamente educativas tenían tres partes: refuerzo académico en matemáticas y lengua, así como ayuda en la preparación a la universidad y otros estudios postsecundarios; educación cívica y social (planificación familiar, higiene, salud, ayuda psicológica); y servicio comunitario (visitar residencias de ancianos o comedores sociales).

Finalmente, los estudiantes recibían 1,25 dólares por cada hora que participaban en el programa (incentivos por insumos) y entre 1000 y 3000 dólares por conseguir el título de bachillerato y/o matricularse en algún tipo de estudio postsecundario (incentivos por resultados).

Claramente una intervención de este tipo no es barata. El coste aproximado fue de unos 25.000 dólares por participante, y su generosa financiación fue compartida entre el departamento (ministerio) de trabajo americano y por la Fundación Ford.

Los datos que produjo la intervención son ricos en muchas dimensiones. Resultados de exámenes de lengua y matemáticas, datos sobre si los estudiantes recibieron el título de bachillerato, si se matricularon en algún tipo de estudios después del bachillerato, datos laborales (si trabajaron y cuánto ganaban), datos sobre conductas de riesgo y resistencia en el programa. Además, los datos sobre participa-

ción en estudios y empleo se recogieron al terminar el programa, tres y cinco años después de acabarlo. Un pequeño problema es que la única medida disponible de los estudiantes antes de su tratamiento es la nota de octavo curso (equivalente a nuestro segundo de ESO).

Los resultados son, como ya avancé, algo decepcionantes. El programa aumenta en un 7% la probabilidad de terminar el bachillerato a tiempo (comparando el grupo tratado con el de control) y en un 6% la de matricularse en la universidad o algún tipo de estudios postsecundarios (estas diferencias representan, respectivamente una mejora del 17,5% y del 23%). El problema es que las diferencias se anulan tres años más tarde, en ese momento en el grupo de control hay un 16% adicional de titulados de bachillerato, frente a un 10% de los tratados. Es decir, el 7% de ganancia inicial es esencialmente un adelanto de la titulación. En ese momento hay un 7% más de chicos trabajando en el grupo de control, lo cual parcialmente puede ser debido a que hay más participantes del grupo tratado estudiando. Finalmente, cuando pasan 5 años no se aprecian diferencias entre los dos grupos ni en porcentaje de empleados ni de titulados. Es decir, los efectos iniciales de la intervención desaparecen con el tiempo.

No todo es negativo. La riqueza del estudio permite analizar la diferencia de resultados entre diversos grupos. La más importante es la que existe entre los estudiantes de comenzaron noveno (nuestro tercero de ESO) con catorce años (es decir, en el tiempo normal) o los que empezaron con más de catorce (porque habían repetido algún curso, por ejemplo). Los que empezaron con catorce y participaron en el programa, tenían una probabilidad de haber acabado el bachillerato 7 puntos por encima de los que no participaron, a los cinco años de acabar el programa. También tenían una tasa de matriculación en la universidad 12 puntos por encima de los estudiantes comparables del grupo de control, y 10 puntos por encima para estudios postsecundarios en general. Para estos chicos, la intervención tiene un efecto duradero.

Como decía antes, una parte de las conclusiones es desconsoladora. Que no se pueda recuperar a algunos chicos incluso con una intervención que cuesta 25.000 dólares resulta dramático. La parte buena es que algunos chicos sí que son rescatables, incluso a los catorce años. Y los experimentos del programa Perry o del Abecedarian, que comentaré en el siguiente epígrafe, sugieren que si podemos intervenir a edades muy tempranas es posible ganar esta batalla. Hay gente a la que no le gusta el programa, pero el nombre «No Child Left Behind» («ningún niño dejado atrás») me parece que señala un buen objetivo.

### 3.2 LOS PROGRAMAS PERRY PRESCHOOL Y ABECEDARIAN

Otro grupo de experimentos de largo plazo muy interesantes tiene mejores resultados, relatados por ejemplo por James Heckman (2006). Estos experimentos muestran, en palabras de este autor que «actuaciones de alta calidad sobre niños en su primera infancia fomentan sus capacidades, lo que permite atacar la desigualdad en su origen. Las actuaciones precoces también aumentan la productividad de la economía».

Los experimentos mencionados en el párrafo anterior son el Programa Preescolar Perry (Perry Preschool Program) y el Programa Abecedario (Abecedarian Program). Estos programas también son ensayos controlados aleatorios. Y además se recogen datos durante periodos largos de tiempo (incluso tras la finalización del tratamiento). Así se pueden observar los efectos sobre los individuos tratados y no tratados en términos de logros académicos, rendimiento en el trabajo y conducta social, mucho tiempo después de la conclusión de las actuaciones. El Programa Perry consistió en una sesión diaria en el aula de 2,5 horas, además de una visita semanal de un profesor al hogar familiar de 90 minutos de duración. Los participantes fueron 58 niños negros pobres de Michigan entre 1962 y 1967, durante 30 semanas al año. Los grupos de tratamiento y de control fueron objeto de seguimiento hasta la edad de 40 años. El Programa Abecedario estudió a 111 niños, procedentes de familias con alta puntuación en un índice de riesgo social. La edad media de entrada fue de 4,4 meses. Este programa fue más intensivo. Su duración fue de un año completo y conllevaba atención durante todo el día. Los niños tratados fueron objeto de seguimiento hasta los 21 años.

Los datos procedentes de estos programas sugieren que los aumentos iniciales de cociente intelectual (CI) desaparecen con el transcurso del tiempo, pero el rendimiento académico es claramente superior. En el Programa Perry, por ejemplo, el porcentaje de estudiantes en educación especial se reduce a la mitad (del 34% al 15%) en el grupo de tratamiento con respecto al grupo de control, el porcentaje por encima del percentil 10% en resultados académicos pasa del 15% al 49%, más del triple, y el porcentaje que se gradúa a su debido tiempo en educación secundaria pasa del 45% al 66%. Esto se traduce en una proporción mucho mayor de individuos que ingresan por encima de 2.000 dólares al mes (del 7% al 28%), poseen una vivienda (del 13% al 36%) o nunca precisan asistencia social (del 14% al 29%). La caída de las tasas de criminalidad entre esos niños resulta igualmente drástica, ya que la proporción de arrestos cae prácticamente a la mitad. Heckman y sus coautores calculan a este programa una tasa de rentabilidad de alrededor del 10% que para un programa público es uno de los más elevados que conozco.

La evidencia experimental de estos programas confirma lo que sabemos por otras vías. La inteligencia humana tiene muchas dimensiones y el éxito social no depende en exclusiva del CI. El programa Perry y Abecedario no consiguen modificar a largo plazo el CI, pero sí los resultados académicos. Heckman y sus coautores han trabajado de manera intensiva sobre la formación de habilidades no cognitivas y sus resultados en el mercado de trabajo (véanse, por ejemplo, Flavio Cunha y James Heckman 2008). Estos estudios sugieren que vale la pena examinar con detalle los programas de educación preescolar porque dada la evidencia que discute Heckman y la evidencia complementaria Rodríguez-Planas me parece que una intervención temprana sobre los grupos de niños menos favorecidos tiene visos de ser una mejor inversión que una intervención tardía. Pero dado su coste, me parece que es importante que la inversión no sea indiscriminada, sino que se concentre de manera intensiva en los niños en situaciones de mayor riesgo y que la intervención se enriquezca con actuaciones fuera del aula.

### 3.3 EL PROGRAMA PREESCOLAR DE TENNESSEE

Antes de saltar al ruedo político defendiendo la educación preescolar como la bala mágica para acabar con la pobreza, veamos los resultados de otro experimento en esa dirección. No parece suficiente como para tirar la toalla, pero al menos requeriría un esfuerzo de investigación mayor y programas piloto para tener una idea más clara de cuáles son los factores que contribuyen al éxito de un programa de este tipo.

Aunque hay otros parecidos me voy a centrar en el programa preescolar de Tennessee porque es el de mayor calidad y que se acerca más a los estudios originales sobre el Programa Preescolar Perry y el Programa Abecedario. La discusión está basada en Mark Lipsey *et al.* (2013). A pesar de la similitud, una precaución importante cuando hagamos la comparación es que esos programas son mucho más caros que cualquier programa, incluido el de Tennessee, que vayan a ser llevados a cabo a gran escala. El programa Perry se calcula que costaría unos 20.000 dólares al año por cada niño hoy en día, y el programa Abecedario entre 16.000 y 40.000 dólares al año por niño. Es fácil de entender cuando se tiene en cuenta que el programa Abecedario acogía a los niños desde que tenían seis semanas de edad durante 50 semanas al año. Resulta poco plausible, a día de hoy, que un programa de ese coste y características vaya a ser replicado a gran escala incluso en países mucho más ricos que el nuestro.

Por lo tanto, la pregunta razonable es si un programa más modesto pero más realista, como el de Tennessee, tiene los efectos que se han observado en los «programas modelo».

El Programa Voluntario Preescolar de Tennessee (TN-VPK, por su sigla en inglés) es un programa estatal que se ofrece a los niños más necesitados del estado, lo cual incluye a los niños elegibles para comidas gratis en la escuela, así como a los que tienen algún tipo de discapacidad o no hablan (y están aprendiendo) inglés. El programa opera con el mismo calendario escolar que las escuelas públicas del estado, y en cada clase hay un profesor certificado y un ayudante, y no más de 20 alumnos. TN-VPK cumple 9 de las 10 características que un programa de calidad debería tener según el Instituto Nacional de Educación Temprana (NIEER). El Estado se gasta unos 90 millones de dólares anuales en las 935 aulas que acogen a unos 18.000 estudiantes. Por tanto el coste es aproximadamente unos 5.000 dólares por estudiante al año, sustancial y razonable para un programa a gran escala, pero lejos de los 16.000-40.000 del Abecedario.

El estudio se centra en unos 3000 estudiantes de dos cohortes, de los años 2009-10 y 2010-11. La asignación al programa se realizó de manera aleatoria, aprovechando que el programa resultaba atractivo y tenía más solicitudes que plazas. Esto, por cierto, es algo que nuestros gestores de la cosa pública deberían aprovechar más a menudo. En España estamos acostumbrados a que muchas cosas se decidan por lotería (el ingreso a una escuela con más solicitudes que plazas, o la asignación de pisos de protección oficial) y por tanto no habría una gran resistencia a hacer estudios piloto de casi cualquier programa atractivo diciendo que si hay más solicitudes que plazas el ingreso se hace por lotería. Y si no hay más solicitudes que plazas esto también nos dice claramente algo sobre el programa.



Una vez los estudiantes se asignaban de manera aleatoria al grupo tratado (participaban en el programa) o se quedaban en el de control se pedía el consentimiento a los padres para hacer una evaluación de los chicos en una batería de medidas, tanto cognitivas como, crucialmente, no cognitivas: como ya hemos dicho otras veces, los estudios de Heckman y otros autores demuestran que una parte crucial del efecto del programa tiene una base no cognitiva. Aquí surgió una dificultad, el consentimiento fue significativamente menor entre los estudiantes del grupo de control (básicamente porque fue más difícil contactarlos). Luego veremos que al menos en variables observables esto no parece que diera lugar a muestras muy desequilibradas, pero los autores toman precauciones econométricas para minimizar el problema.

La muestra final consiste en unos 1.075 estudiantes de 58 escuelas que representan a una variedad de entornos. A los estudiantes se les miden las capacidades cognitivas con dos pruebas psicométricas de habilidades verbales y dos de habilidades cuantitativas/solución de problemas, y cuando llegan a la escuela se les añaden dos pruebas más, una verbal y otra cuantitativa. Para medir los efectos no cognitivos se usan otros dos instrumentos, uno es la escala Cooper-Farran de medida de conducta (CFBRS). Este instrumento mide por un lado capacidades de trabajo (trabajar independientemente, escuchar al profesor, realizar tareas en el tiempo asignado), y por otro la conducta (comportamiento adecuado en trabajo de grupo, juego y actividades en el patio, respuesta a errores o problemas de los demás, expresión de sentimientos o ideas). Otro instrumento es el registro de conducta en el aula (ACBR) de Farran, Bilibrey y Lipsey que mide cuatro facetas: preparación para el nivel del curso escolar, cuánto les gusta el curso, problemas de conducta, y relación con los pares.

En cuanto a los resultados, dada la diferencia entre grupo tratado y de control respecto al consentimiento lo primero que comprueban los autores es que a pesar de todo en las variables observadas antes del tratamiento prácticamente no hay diferencias importantes.

El siguiente análisis se centra en los efectos en las medidas cognitivas para el año de preescolar. Los efectos son todos positivos, casi todos significativos y bastante grandes. El compuesto, por ejemplo, nos dice que en media tanto los estudiantes tratados como los del grupo de control mejoran, pero la mejora es un 44% mayor para los participantes en el programa. En las medidas no cognitivas también hay mejoras pero no son tan claras ni tan grandes. En el test Cooper-Farran la mejora es significativa en las dos medidas, en el test ACBR solamente en la primera, la preparación para el curso.

Pero los resultados descorazonadores llegan cuando se observa la evolución de los años siguientes. Tanto las habilidades cognitivas como las no cognitivas dejan de mejorar, y en algunos casos hasta retroceden en los primeros cursos de primaria, aunque es verdad que en muy pocos casos de manera significativa.

Ningún estudio aislado puede considerarse definitivo, y ciertamente debemos esperar e investigar más porque el problema lo merece. Pero la lección principal es que todos los resultados en ciencia social y políticas públicas son preliminares y debemos mantener la mente abierta a posibles cambios de dirección. Como en esa magnífica frase atribuida a Keynes: «Cuando las circunstancias cambian, yo cambio de opinión. ¿Usted qué hace?»

### 3.4 ESCUELAS «SIN EXCUSAS»

Volviendo a experimentos exitosos, Will Dobbie y Roland Fryer (2011) muestran que conseguir que una escuela funcione requiere: altas expectativas, más tiempo de clase, altas dosis de tutoría, información frecuente del profesor y el uso de datos para guiar la formación. Estos factores explican el cincuenta por ciento de la variación en los resultados de las escuelas investigadas. En otras palabras, trabajo duro y serio por parte de todos los implicados, ese es el modelo que denominan *sin excusas*.

Una de las dificultades más serias para desentrañar lo que hay detrás de la función de producción educativa viene de que este servicio es realmente complejo. Hay muchos factores implicados y hasta ahora ha sido difícil de lograr evidencia de muchos de ellos. A menudo se habla del número de estudiantes, de los ordenadores, o del gasto por estudiante pero el proceso educativo implica a muchos más factores y es difícil investigar su efecto. Por ejemplo, hay escuelas que ponen mucho énfasis en la instrucción artística porque aparte de su valor intrínseco piensan que esto mejora la creatividad y estimulan el desarrollo académico. En otros lugares pasa algo similar con el deporte, o el ajedrez.

Y hay estrategias sustitutivas entre sí cuyos impactos no conocemos bien: ¿qué pasa si aumentamos el tamaño de clase a cambio de aumentar las horas de instrucción? Esto se puede hacer sin tocar las horas de trabajo de los profesores y por tanto puede ser neutral en términos de coste. Otro ejemplo: se puede aumentar el tamaño de la clase, pero usar los recursos ahorrados para dar más tutorías individuales o en pequeños grupos. Contestar este tipo de preguntas es lo que puede hacerse con el tipo de datos que han recogido Dobbie y Fryer.

Los datos vienen de las «charter schools» de Nueva York. Estas son escuelas de titularidad privada y financiación pública. En eso se parecen a nuestras escuelas concertadas, pero la financiación por alumno es igual a la de la escuela pública, y en consecuencia no hay ningún copago de los padres. La ventaja de usar estas escuelas es que se introdujeron justamente para tener más libertad en las prácticas académicas (no así de admisión, ya que no pueden seleccionar a sus estudiantes, si hay exceso de demanda de plazas, se hace un sorteo para decidir quién entra). La idea era justamente que esta libertad generara variación y permitiera ver cuáles eran las mejores prácticas docentes. De las 48 escuelas elementales (de infantil a cuarto) y 37 de grado medio (de quinto a octavo) elegibles, participaron 22 elementales y 13 de grado medio.

Los datos recogidos de cada escuela fueron muchos y muy detallados. Varias entrevistas con el director aportaron información sobre: la formación continuada de los profesores, tiempo de clase, uso de datos para la instrucción, información e implicación de los padres, y cultura de la escuela. La información sobre el rigor curricular se obtuvo de los planes docentes para cada curso en matemáticas y lengua. También se obtuvo información sobre la cultura y las prácticas mediante visitas de día completo a las escuelas. Las visitas incluían observación y grabaciones de al menos una clase de matemáticas y una de lengua, así como entrevistas de profesores y alumnos escogidos al azar.

A la hora de mirar los resultados, lo primero es ver si ir a estas escuelas sirve para algo. Y como se pueden imaginar la respuesta es que sí. El efecto en matemá-

ticas es una mejora de alrededor de un sexto de una desviación estándar, y en lengua algo menos de la mitad. Puede parecer poco, pero es aproximadamente la mitad de la diferencia media entre un estudiante cuyo padre tiene formación universitaria y uno que solamente tiene la educación obligatoria.

La cuestión más novedosa de este artículo en particular es averiguar las razones por las que estas escuelas son más efectivas. Para esto se recopilaron los datos de los que hablamos antes. El procedimiento utilizado es hacer regresiones que expliquen el efecto estimado de cada escuela en las regresiones anteriores con algunas variables. La principal precaución con estos resultados es que no se pueden interpretar de manera causal, son correlaciones nada más y podría haber factores omitidos importantes. Pero son muy informativas, y en algún caso coinciden con lo que ya sabemos con metodologías estadísticamente más potentes.

Por ejemplo, ni el tamaño de clase, ni el gasto por alumno, ni el hecho de que el profesor tenga un máster parece ser importante. De hecho, como mucho parecen tener una correlación negativa si se ponen todas juntas. Algo que por cierto ya sabemos por otras vías.

En cambio, el efecto de las cinco políticas que expresé en la introducción: altas expectativas, más tiempo de clase, altas dosis de tutoría, información frecuente del profesor y el uso de datos para guiar la formación, explican el cincuenta por ciento de la variación en los resultados de las escuelas investigadas. Y no solamente esto, en una regresión del efecto de la escuela con un índice y cada una de las variables por separado, cuatro de las cinco estrategias (todas menos las altas expectativas) retienen valor explicativo.

La última parte del artículo se dedica a verificar, de una manera muy parcial, claro, algunas teorías sobre lo que hace una escuela efectiva.

Una de ellas es la teoría de la «escuela de servicios» en la cual el objetivo no son solamente los conocimientos sino el «bienestar global» del chico. Para comprobarla crean varios indicadores que son iguales a uno si la escuela tiene un trabajador social, servicios de salud, servicios «holísticos» o si el ideario escolar habla de proveer una «aproximación comprensiva al bienestar del estudiante». Con esto se genera un índice global que promedia estos 4. Otra teoría es la que se centra en la contratación y retención de los mejores profesores. Para probar esta teoría se crea un índice que promedia indicadores sobre si la escuela tiene un mayor número de requisitos para contratar que la media, si paga un salario superior a la media, si tiene una tasa de rotación mayor a la media y si ofrece bonos por rendimiento a los profesores. El último modelo comprobado es el modelo «sin excusas», una filosofía que se centra en más tiempo en la escuela, disciplina estricta y un mayor enfoque en habilidades lingüísticas y matemáticas básicas.

Cuando se realiza este análisis la «escuela de servicios» parece no ser muy relevante. Tanto la búsqueda de buenos profesores, como la filosofía «sin excusas» están relacionadas con el rendimiento. Luego los autores comprueban lo que sucede cuando se añade el índice de las cinco políticas básicas analizado antes. Para las dos primeras estrategias, escuela de servicios y contratación de buenos profesores, no hay cambios importantes. Pero para la filosofía «sin excusas» la introducción de esta variable es importante. Lo que quiere decir probablemente es que estas políticas están muy relacionadas con la filosofía.

Es posible que la conclusión resulte frustrante para algunos lectores. Un estudio científico tan exhaustivo de unos brillantes profesores de Harvard para concluir

que si profesores y estudiantes trabajan duro, mejorarán los resultados escolares. A veces parece tonto repetir lo obvio, pero quizá sea también necesario. Todos ponemos excusas para nuestro fracaso. Y muchas de estas quejas están fundamentadas, pero no siempre, y esa es la virtud de la ciencia y de la evidencia, que hacen más difícil esquivar el bulto.

### 3.5 LOS EXPERIMENTOS DE EDLABS, INCENTIVOS FINANCIEROS AL ESTUDIO

Esta sección es quizá la más ambigua. La buena noticia es que los incentivos a los estudiantes a veces funcionan. La mala es que solamente algunos incentivos son tan potentes como nos gustaría y que nos queda mucho por aprender. El resumen es que los premios por resultados no producen efectos apreciables. En cambio, las recompensas por «actividades», por ejemplo leer libros, sí que los dan. Los detalles son importantes, sin embargo porque nos pueden ilustrar sobre las razones de los resultados y, especialmente, sobre las circunstancias en que éstos se producen.

Fryer y su equipo del EdLabs realizaron experimentos en varias ciudades americanas. En todos ellos había un grupo de colegios (el grupo tratado) en el que los chicos podían recibir una recompensa por hacer o conseguir algo, y otro grupo de colegios (el grupo de control) que no recibía recompensa. La asignación de cada colegio a grupo tratado o grupo de control era aleatoria. Ambos grupos de colegios eran evaluados al final del año y la diferencia entre grupos sirve para medir el efecto del «tratamiento».

Los experimentos se realizaron en varios estados americanos. En Dallas a los estudiantes (segundo de primaria) del grupo tratado se les daba 2 dólares por leer un libro (y se verificaba si lo hacían con un test de comprensión por ordenador al acabar de leerlo), con un máximo de hasta 40 libros al año.

En Nueva York se trabajó con chicos de cuarto y séptimo (equivalente pues a nuestro primero de la ESO). Aquí se recompensaba (al grupo tratado) directamente por las notas en 10 pruebas estandarizadas de matemáticas y lengua. Los de cuarto recibían 2,5 dólares por cada punto (sobre un total de 10) en cada examen. Podían ganar, pues, hasta 250 dólares en el curso. Para los de séptimo se doblaban las ganancias.

En Washington los estudiantes recibían puntos por distintos factores de «producción» educativos. Todos ellos recibían puntos por asistencia y comportamiento y cada colegio escogía otros tres criterios con ayuda de los expertos de Edlabs. En particular, solían estar incluidos el llevar uniforme, hacer la tarea o los trabajos de clase. La recompensa media fue de 320 dólares y la máxima de 1300.

En Chicago los estudiantes eran de noveno (nuestro tercero de la ESO). Aquí se recompensaba (al grupo tratado) por las notas de clase en las cinco asignaturas troncales (matemáticas, lengua, ciencias naturales, ciencias sociales y educación física). Cada estudiante recibía 50 dólares por cada A (sobresaliente), 35 dólares por la B (notable), 20 dólares por la C (bien), y 0 por la D (aprobado). Si un estudiante suspendía un curso, no recibía nada por ninguna asignatura hasta que la

aprobase y entonces recuperaba el dinero de las otras. La máxima ganancia era de 2.000 dólares al año (la media fue de 695).

A todos los estudiantes de cada ciudad (tratados y de control) se les evaluó al final del año con un test estandarizado igual para todos, pero específico de cada ciudad. Por ejemplo, en Dallas con el test Iowa de destrezas básicas en matemáticas y lectura. En Nueva York con un test de matemáticas y lengua desarrollado por McGraw Hill. Washington y Chicago tienen también sus propios sistemas de evaluación locales.

Los resultados para Nueva York y Chicago muestran que los incentivos para sacar mejores notas, sea en los exámenes normales del colegio o en exámenes estandarizados no mejoran los resultados de los chicos. Para ninguno de las pruebas, de los cursos o de las metodologías estadísticas se obtiene un resultado estadísticamente distinto de cero. En el caso de Nueva York es claro que dar incentivos para hacerlo bien en el test no mejora los resultados del test. Pero, como dice Fryer, el resultado de Chicago es algo más difícil de interpretar. Dar incentivos para las notas no tiene impacto en la prueba estandarizada, pero en principio podría tenerlo en las notas. Esto es difícil de juzgar porque para las notas la comparación entre grupo tratado y de control puede estar contaminada, los profesores pueden haber cambiado (incluso sin querer) su forma de puntuar al saber el efecto que esto tendría en los estudiantes.

En cambio, el efecto en Dallas es significativo y bastante grande. Participar en el programa de incentivos a la lectura mejora el resultado en exámenes estándar de vocabulario o comprensión lectora aproximadamente un cuarto de desviación estándar, es decir, sube a un chico casi diez puntos porcentuales en la distribución. Para que se hagan una idea, es un tamaño que equivale a la diferencia entre un estudiante medio cuyos padres no han acabado la educación obligatoria respecto a los que tienen padres con estudios universitarios. Y esto a un coste máximo de 80 dólares por niño al año. No hay muchas intervenciones sociales que puedan ganar a esta en coste-efectividad. Y aunque no es definitivo, merece ser replicado y estudiado de manera más detallada.

Los resultados de Washington también dan resultados positivos, pero su significación estadística es marginal y hay otros problemas (la aleatorización no fue tan buena y el experimento es más complicado).

Por un lado, estas observaciones son interesantes porque quieren decir que se ha encontrado una intervención que funciona a un coste razonable. Pero es algo decepcionante que ni siquiera unos incentivos de un tamaño tan considerable a los alumnos por resultados consigan un efecto apreciable. Fryer sugiere que la ventaja de incentivar directamente el esfuerzo es que el estudiante ve una conexión más directa porque se reduce la incertidumbre. El problema es que no todos los esfuerzos son tan fáciles de medir como la lectura de un libro. Y esto hace inevitable que en algún caso haya que premiar resultados.

Por tanto yo no desesperaría en la búsqueda. Sabemos por los resultados de Joshua Angrist y Victor Lavy (2009) en Israel que unos incentivos algo mayores (unos 1.500 dólares en Israel, un país de renta algo menor) que los incentivos a los estudiantes por resultados pueden funcionar en algunas circunstancias. Ahora se trata de averiguar en cuáles funcionan. Hay otros indicios, aparte de la cuantía de cuándo lo hacen. Por ejemplo, es posible que los profesores o los padres de Israel estuvieran más implicados ya que allí se trataba del acceso a la universidad, por lo

que el «premio» para ellos era más grande. Además, los estudiantes de Dallas eran los más jóvenes de todos, y sabemos por los trabajos de Heckman que he referido antes que la intervención temprana es particularmente efectiva. Así que una recomendación para los responsables de política en particular, y los ciudadanos interesados en general, es que hay que seguir buscando hasta estar seguros de qué es lo que funciona mejor.

#### IV. CONCLUSIÓN

En este artículo he discutido la capacidad de las intervenciones educativas para mitigar o mejorar las desigualdades sociales y la pobreza. Dos conclusiones tentativas surgen de esta discusión. La primera es que no todas las intervenciones funcionan. Aquellas que tienen como objetivo prioritario a los más jóvenes, y que se centran en mejorar el trabajo y la disciplina tienen mejor pronóstico, pero no una garantía universal. Y, justamente por este motivo, es muy importante que todas las intervenciones públicas vengán acompañadas de estudios de impacto serios y rigurosos, a ser posible mediante ensayos controlados aleatorizados.

#### V. BIBLIOGRAFÍA

- ANGUEL, B. y CABRALES, A., «The Determinants of Success in Primary Education in Spain,» *Journal of Public Programs and Policy Evaluation*, N.º 2, 2014, pp. 22-53.
- ANGRIST, J. y LAVY, V., «The effects of high stakes high school achievement awards: Evidence from a randomized trial,» *The American Economic Review*, Vol. 99, N.º 4, 2009, pp. 1384-1414.
- ATKINSON, A. y PIKETTY, T. (ed.), *Top incomes: A global perspective*, Oxford University Press, 2010.
- BOURGUIGNON, F. y ROGERS, F., «Global Returns to Higher Education: Trends, Drivers, and Policy Responses», *ABCDE*, 2008, p. 25-44.
- BROWN, D., «Paul Meier, biostatistician and co-inventor of a famous graph, dies at 87,» *Washington Post*, 10 de agosto de 2011, disponible online en: [https://www.washingtonpost.com/local/obituaries/paul-meier-biostatistician-and-co-inventor-of-a-famous-graph-dies-at-87/2011/08/09/gIQAfHBH7L\\_story.html](https://www.washingtonpost.com/local/obituaries/paul-meier-biostatistician-and-co-inventor-of-a-famous-graph-dies-at-87/2011/08/09/gIQAfHBH7L_story.html)
- CUNHA, F. y HECKMAN, J. J., «Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation,» *Journal of Human Resources*, Vol. 43, N.º 4, 2008, pp. 738-782.
- DOBBIE, W. y FRYER, R. G. «Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone,» *American Economic Journal: Applied Economics*, Vol. 3, 2011, pp. 158-187.
- FRYER, R. G., «Financial Incentives and Student Achievement: Evidence from Randomized Trials,» *The Quarterly Journal of Economics*, Vol. 126, N.º 4, 2011, pp. 1755-1798.
- GOLDIN, C. y KATZ, L. F., «The race between education and technology: The evolution of US educational wage differentials, 1890 to 2005» *National Bureau of Economic Research*, 2007, N.º w12984.

- HECKMAN, J. J., «Skill formation and the economics of investing in disadvantaged children,» *Science*, Vol. 312, N.º 5782, 2006, pp. 1900-1902.
- LIPSEY, M. W., *et al.*, «Evaluation of the Tennessee voluntary prekindergarten program: Kindergarten and first grade follow-up results from the randomized control design,» Nashville (TN), Peabody Research Institute, 2013.
- MILANOVIC, B. «Global Income Inequality in Numbers: in History and Now,» *Global Policy*, Vol. 4, N.º 2, 2013, pp. 198-208.
- RODRÍGUEZ PLANAS, N., «Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the U. S.» *American Economic Journal: Applied Economics*, Vol. 4, 2014, pp. 121-139.

