

**DIRECCIÓN DE
INVESTIGACIÓN**

01-2018

INCENTIVOS INDIVIDUALES DOCENTES Y LOGROS ACADÉMICOS DE LOS ESTUDIANTES: EVIDENCIA DE ECUADOR

**JUAN PONCE
JORGE YÉPEZ**

Resumen

Este estudio evalúa el impacto del programa de Evaluación y Compensación Docente en el logro académico de los estudiantes. Los maestros que obtuvieron un puntaje por encima de cierto límite en la evaluación recibieron una compensación monetaria adicional durante cuatro años consecutivos. Aunque la evaluación se centró en el conocimiento de los docentes, el gobierno esperaba ver un impacto en el rendimiento de los estudiantes a través de la motivación de los docentes premiados. Por lo tanto, evaluamos el impacto del rendimiento de los estudiantes del programa al final de la implementación. Nuestra estrategia de identificación explota las discontinuidades de la regla de asignación de incentivos y se basa en un diseño de regresión discontinua (RD). Encontramos que el programa de compensación no tiene un impacto estadísticamente significativo en los resultados de los estudiantes. Esto significa que el incentivo financiero recibido por los maestros debido a su buena calificación no afectó el desempeño de sus estudiantes.

Introducción

Las autoridades educativas a nivel mundial han optado por vincular incentivos económicos con alguna medida de desempeño con el objetivo de mejorar los resultados tanto en maestros, como en estudiantes. Los programas de incentivos se pueden clasificar en: incentivos grupales o individuales. El último incentiva a los maestros de acuerdo con el desempeño de sus propios estudiantes, mientras que el primero premia a un grupo de maestros en base al desempeño promedio de todos los estudiantes del grupo. En general, los grupos están formados por todos los maestros que forman una escuela. Los resultados de estos programas, tanto individuales como grupales, son bastante variados.

Por ejemplo, con respecto a los incentivos grupales, Fryer (2013) evaluó un programa experimental sobre incentivos docente en más de 200 escuelas públicas de la Ciudad de Nueva York. Cada escuela participante podría ganar \$ 3,000 por cada maestro participante, siempre que la puntuación de la escuela cumpla con la meta establecida por el Departamento de Educación. El autor no encontró un impacto significativo en el rendimiento de los estudiantes¹. Glewwe et al. (2010) evaluaron otro programa grupal, pero en un país en desarrollo, Kenia. El programa se implementó al azar en 50 escuelas, mientras que otras 50 escuelas formaron parte del grupo de control. Los incentivos eran entre el 21% y el 43% del salario mensual de un profesor y estaban vinculados a los puntajes de las pruebas estatales de los estudiantes. Por un lado, los resultados encontraron una mejora de 0,14 desviaciones estándar (DE) en los exámenes. Por otro lado, no se encontró ningún impacto en los exámenes que no estaban vinculados a los incentivos, ni en los exámenes estatales una vez finalizado el programa. Los autores concluyen que los maestros respondieron al programa aumentando el número de sesiones de preparación para los exámenes durante la implementación del programa, por lo que los efectos positivos no perduraron.

En relación con los programas de incentivos individuales, los resultados también han sido mixtos. Por ejemplo, Springer et al (2010) evaluaron un programa en Nashville, Tennessee, donde maestros voluntarios fueron asignados al azar a grupos de tratamiento y control. Los maestros en el grupo de tratamiento fueron elegibles para diferentes bonos (\$ 5000, \$ 10000 y \$ 15000) según su valor agregado en el desempeño de los estudiantes (80, 90 y 95 percentiles, respectivamente). Los autores no encontraron impacto en el rendimiento de los estudiantes. Ellos atribuyen que los umbrales altos pueden haber desanimado a muchos maestros a participar. Otro experimento aleatorio que no encontró efectos significativos se llevó a cabo en las escuelas de Chicago. Escuelas enteras fueron asignadas al azar a la adopción temprana o posterior del programa. Bajo este programa, los maestros eran elegibles para recibir pagos de hasta \$ 6,400 (en promedio 10% de su salario anual) dependiendo de su contribución al valor agregado del desempeño de sus

¹ Además, se encontró un impacto negativo en las escuelas secundarias, especialmente en las escuelas más grandes, donde posiblemente el problema del polizone (*"free rider"*) podría haber sido importante debido a la gran cantidad de maestros participantes.

estudiantes y una observación de clase. Según Glazerman y Seifullah (2012), el programa no generó ningún impacto significativo en los resultados de los estudiantes².

Por otro lado, Fryer et al (2012) encontraron efectos positivos significativos en un programa de incentivos individuales en Chicago Heights, Illinois. El programa fue un experimento implementado en 9 escuelas. El sistema de incentivos usó un método de pago por percentil en el cual, al comienzo del estudio, cada estudiante es emparejado con sus vecinos más cercanos, en términos de desempeño en exámenes pre-programa, pero perteneciente a una escuela diferente. Luego, se clasificó a cada estudiante de acuerdo con su mejora entre el examen de referencia y el examen final del año. Cada maestro recibe una puntuación de percentil, que fue el percentil promedio de todos sus estudiantes. El programa recompensó \$ 80 por puntaje percentil, hasta una recompensa máxima posible de \$ 8,000. El valor esperado de la recompensa (\$ 4,000) fue equivalente al 8% del salario anual promedio de los maestros. Los maestros fueron asignados al azar en incentivos grupales o individuales. Además, los maestros tratados podrían clasificarse en grupos de "ganancia" o "pérdida". Los maestros que pertenecían al grupo de "ganancia" recibieron su premio al final del año, mientras que los maestros del grupo de "pérdida" recibieron \$ 4,000 al comienzo del programa, pero se comprometieron a devolver parte del incentivo si sus estudiantes no cumplían con los requisitos. Al mismo tiempo, si los estudiantes superan el rendimiento, los maestros recibieron un pago adicional al final del año. En resumen, los maestros de los grupos de "ganancia" y "pérdida" recibieron pagos iguales, pero diferentes en la aplicación. Los resultados, a pesar de no encontrar impacto en los grupos de "ganancia", encontraron un impacto positivo de 0.22 DE en los grupos de "pérdida". Además, los efectos de los incentivos grupales fueron mayores que los incentivos individuales.

Asimismo, Dee y Wyckoff (2015) evalúan un programa de incentivos implementado en Washington, DC. El programa se basó en el puntaje IMPACT, que estaba compuesto principalmente por el valor agregado del profesor en el desempeño de los estudiantes y una observación en el aula. Los maestros se dividieron según su puntaje: altamente eficiente (más de 350 puntos), eficiente (entre 250 y 349 puntos), moderadamente eficiente (entre 175 y 249 puntos) e ineficiente (menos de 175 puntos). Los maestros ineficientes fueron despedidos, mientras que los maestros ubicados en el grupo moderadamente eficiente, fueron amenazados con el despido si no calificaban para un mejor grupo en la evaluación del próximo año. Por otro lado, los maestros altamente eficientes recibieron \$ 25,000. Además, los maestros que obtuvieron calificaciones altamente eficientes durante dos años consecutivos recibieron un aumento salarial permanente de \$ 27,000. Los autores explotaron los umbrales de clasificación de los grupos mediante un diseño de regresión discontinua (RD). Encontraron que los maestros calificados como moderadamente eficientes en el segundo año, mejoraron sus calificaciones de tercer año en 0.26 DE. Asimismo, los docentes clasificados como altamente eficientes mejoraron en promedio 0.24 DE.

² Sin embargo, los autores informan que la implementación del programa no se llevó a cabo según lo previsto. Por ejemplo, los pagos de los maestros fueron más bajos que los establecidos originalmente. Tampoco hubo recompensas basadas en el valor agregado individual. Igualmente, el valor agregado fue calculado a nivel escuela y nivel de grado, ya que los datos individuales requeridos estaban incompletos.

Fuera de los Estados Unidos, Muralidharan y Sundararaman (2011) estimaron el efecto de los incentivos individuales y grupales en Andhra Pradesh, India. Los incentivos de los docentes se vincularon a las mejoras en los puntajes de los estudiantes. Los incentivos representaron, en promedio, el 3% del salario anual de los docentes. El experimento asignó aleatoriamente a las escuelas en tres grupos: sin incentivos, incentivos grupales e incentivos individuales. Los autores evaluaron el programa dos años después de su implementación. Ambos incentivos generaron mejoras significativas en los puntajes de los estudiantes. El programa tuvo un impacto de 0.27 DE en matemáticas y 0.17 DE en lenguaje. Por separado, los impactos son algo mayores en los incentivos individuales (0,28 DE) en comparación con los incentivos grupales (0,15 DE). Sin embargo, los autores encontraron evidencia de que los maestros asignaron más tiempo a la práctica de pruebas simuladas. Además, no se encontró evidencia sobre la tasa de asistencia de los maestros.

Finalmente, Lavy (2009) estimó el efecto de un programa de incentivo individual en las escuelas secundarias israelíes, donde los maestros recibieron incentivos basados en los puntajes y las tasas de aprobación de sus estudiantes. Los maestros se clasificaron según la diferencia entre el desempeño de sus estudiantes y el valor predicho de esta variable, en función de una regresión controlada por las características socioeconómicas de los estudiantes, su nivel (básico, intermedio, avanzado) y el tamaño de la clase. Además, debido a que el modelo de regresión incluía efectos fijos escolares, los maestros compitieron contra otros en la misma escuela. Los incentivos alcanzaron hasta \$ 7500. El autor encontró un efecto positivo del 5% en matemáticas y 14% en la tasa de aprobación.

El programa de incentivos para maestros ecuatorianos, tal como se describe en la siguiente sección, se basó directamente en los puntajes de los maestros de una evaluación que evalúa el conocimiento del maestro. Si bien hay pruebas limitadas de que los incentivos para maestros al no estar vinculados directamente con los logros de los estudiantes tengan impactos positivos, el gobierno esperaba que el programa afectara la motivación de los maestros y, a través de este canal, el incentivo mejoraría el rendimiento de los estudiantes entre los maestros premiados. Como detallamos en párrafos anteriores, otros países han optado por vincular de manera efectiva los incentivos docentes directamente a los resultados de los estudiantes. Adicionalmente, como el programa se canceló, algunos docentes no recibieron el número de pagos que en principio fueron ofrecidos. Además, las contribuciones de la economía del comportamiento se han explotado con éxito al implementar incentivos económicos, como en Dee y Wyckoff (2015) o en Fryer (2013). Esta falta de comprensión de la evidencia internacional pone claramente en peligro las intenciones del gobierno de generar un impacto en el rendimiento de los estudiantes.

Metodología

Base de datos y descripción del programa

El Sistema Nacional de Evaluación, establecido por el Ministerio de Educación del Ecuador en el año 2007, fue un proyecto que tenía como objetivo evaluar a los estudiantes, maestros³ y directores de escuelas. En 2009, se implementó la evaluación de maestros y directores, bajo el nombre de Ser Maestro 2009-2012, el cual incluye los siguientes ítems:

- (i) evaluación de conocimientos (50 puntos) (que se dividió en evaluación pedagógica (10 puntos), habilidades lingüísticas (10 puntos) y conocimiento (30 puntos))
- (ii) autoevaluación (5 puntos),
- (iii) coevaluación (5 puntos),
- (iv) evaluación de los estudiantes (8 puntos),
- (v) evaluación de los padres (12 puntos),
- (vi) evaluación del director (5 puntos) y
- (vii) observación en el aula (15 puntos).

Si bien la evaluación abarca una amplia variedad de áreas, la mayoría de los docentes (94%) obtuvo entre 45 y 50 puntos en los componentes que no estaban relacionados con la evaluación del conocimiento (ii) - (vii). En este sentido, solo se utilizaron los puntos de evaluación de conocimientos para discriminar los resultados de los docentes.

El programa otorgó a los maestros un incentivo de \$ 1200, por año durante 4 años, a los maestros que obtuvieron al menos 90 puntos, mientras que los maestros con al menos 80 puntos recibieron \$ 900 por año durante 4 años.

Del 2009 al 2011, el proceso de evaluación fue regulado por el Ministerio de Educación. Hasta entonces, el 50% de los docentes fueron evaluados. Para 2012, los maestros restantes fueron evaluados por el Instituto Nacional de Evaluación Educativa (INEVAL)⁴. Sin embargo, debido a que no hay registros de los pagos desde el 2012, excluimos de nuestro análisis a todos los maestros evaluados durante ese período.

La base de datos de Ser Maestro 2009-2012 tiene 70,464 observaciones⁵, que incluyen a todos los maestros que tomaron la evaluación, de todos los niveles escolares. De este universo de docentes, solo 3,073⁶ se observaron en los registros administrativos de la Comunidad Educativa

³ El programa evaluó maestros en todos los niveles de escolaridad.

⁴ Para ese año, la metodología se modificó ligeramente, por lo que los maestros que obtuvieron al menos 80 puntos también deben obtener al menos el 60% en cada uno de los elementos de evaluación de conocimientos para calificar para el incentivo. Además, el premio estaba destinado a ser pagado solo por un período único.

⁵ Limitamos nuestro análisis a los maestros de las escuelas públicas regulares.

⁶ Se emparejaron 5,929 maestros, pero Ministerio de Educación evaluó a solo 3,073 antes de 2012, el resto fue evaluado por INEVAL.

en Línea que vincula a los profesores con los estudiantes de secundaria⁷. Con este enlace, pudimos hacer coincidir a los maestros con los puntajes de sus estudiantes en el examen nacional estandarizado Ser Bachiller 2015-2016, que solo toman los estudiantes en 3ro de bachillerato. Con esta base de datos, es posible establecer si existe algún efecto entre los maestros que recibieron una bonificación durante los años 2009-2012 y los logros académicos de sus estudiantes en el año escolar 2015-2016. De este grupo de maestros, 2,214 obtuvieron calificaciones por debajo de 80, 829 entre 80 y 89, y 30 obtuvieron 90 o más. Sin embargo, solo 705 recibieron \$ 900, de los cuales 673 eran maestros durante el año de la evaluación, 31 eran directores⁸ y 2 eran directores y maestros. En principio, el programa tenía la intención de dar a los maestros un pago por año, hasta que completen 4 pagos. Sin embargo, el gobierno canceló el programa en 2012; por lo tanto, algunos maestros evaluados después de 2009, recibieron menos pagos o no recibieron ningún pago. Más precisamente, del grupo de maestros, solo 1 recibió un solo pago, pero tenía 3 pagos pendientes; 482 registraron 3 pagos y tenían 1 pendiente, 191 registraron 4 pagos y 1 maestro registró 5 pagos. Por otro lado, solo 29 maestros obtuvieron un puntaje superior a 90. De este grupo, 27 recibieron 3 pagos y registraron 1 pago pendiente, mientras que 2 recibieron 4 pagos.

Curiosamente, 1195 maestros con puntajes entre (80-90) no registran incentivos. Sin embargo, 976 se evaluaron en 2012, por lo que se desconoce si recibieron el incentivo. Por lo tanto, solo 219 maestros que fueron evaluados antes del 2012, con puntajes entre (80-90) no recibieron pagos. Esto podría explicarse porque los incentivos solo se otorgaron a los maestros con contratos con nombramiento y no con contrato. Por otro lado, 86 docentes registraron puntajes por encima de 90 puntos pero no recibieron incentivos. Sin embargo, 76 se evaluaron en 2012 y solo 10 se evaluaron antes. La siguiente tabla resume los pagos y el número de maestros.

Tabla 1: Pagos y puntaje

N	Pagos recibidos	Pagos pendientes	Incentivo por pago (\$)	Puntaje	DE	Min	Max
1	1	3	900	82.08			
482	3	1	900	83.51	2.64	79.52	89.46
191	4	0	900	82.72	2.23	79.53	89.11
1	5	0	900	83.64			
27	3	1	1200	90.74	0.98	89.53	92.98
2	4	0	1200	90.42	0.29	90.22	90.63
219	0	0	0	83.21	2.70	79.52	89.44
10	0	0	0	90.97	1.53	89.50	94.13
976*	-	-	-	83.63	2.59	79.51	89.46
76*	-	-	-	90.76	1.10	89.54	93.56

⁷ Si bien, los maestros de todos los niveles escolares fueron evaluados en la evaluación Ser Maestro 2009-2012, debido a las limitaciones de datos sobre el rendimiento de los estudiantes, restringimos el análisis de los estudiantes de último año de secundaria, que es el único nivel de educación evaluado.

De nuestra base de datos de 162,488 estudiantes que tomaron la evaluación Ser Bachiller 2015-2016, 98,247 (60%) se enlazaron a los maestros que tomaron la evaluación antes de 2012, 35,655 (22%) a maestros que recibieron \$ 900 y 2,371 (1.5%) a los maestros que recibieron \$ 1200.

⁸ Los maestros y maestros-directores que recibieron los incentivos tienen puntajes en el rango válido de 80 a 90, mientras que solo 7 directores tienen puntajes en la base de datos. De este grupo de directores, 4 obtuvieron puntuaciones por debajo del umbral de 80. Esto se puede explicar porque la evaluación del director fue diferente a la del docente.

* Se desconoce si los profesores evaluados en el 2010 recibieron incentivos.

En general, dentro del marco de la teoría económica del capital humano, se espera que los salarios estén relacionados con el nivel de productividad de un trabajador y que aumenten de acuerdo con las habilidades que el individuo está adquiriendo (Becker, 1994). Sin embargo, en el caso de la educación y la influencia de la calidad de los maestros en los resultados de los estudiantes, varios estudios han demostrado que los ingresos no están necesariamente relacionados con el rendimiento de los estudiantes. Principalmente porque la estructura salarial de los maestros no permite diferenciar adecuadamente sus habilidades (Hanushek y Rivkin, 2006). En este sentido, considerando la evidencia de que existe una relación positiva entre los incentivos monetarios para los maestros y los resultados de los estudiantes (Figlio y Lawrence, 2007), se podría esperar que una bonificación pueda tener un impacto positivo en el rendimiento de los estudiantes.

Sin embargo, como pudimos notar, el programa no se entregó tal como fue diseñado, y los maestros no recibieron la cantidad de pagos que el programa pretendía otorgar. Además, la bonificación se otorgó en base a una evaluación sobre el conocimiento de los maestros, sin considerar seriamente las características de otros maestros, como el apoyo emocional, la organización en el aula o el apoyo educativo⁹; y sin vincular a los maestros con los logros de sus estudiantes. Estas limitaciones podrían haber perjudicado claramente las intenciones de influir positivamente en los logros de los estudiantes a través de los incentivos de los maestros.

Estrategia de identificación

Utilizamos una estrategia de regresión discontinuidad (RD) para evaluar el impacto del programa de compensación docente. El hecho de que se otorgó el incentivo monetario (bonificación) a los maestros que obtuvieron un puntaje por encima de un cierto umbral se utiliza como fuente de variación exógena. Comparando los puntajes de los alumnos cuyos maestros obtuvieron un puntaje ligeramente por encima del punto de corte (quienes recibieron el Bono) con los puntajes de los alumnos cuyos maestros obtuvieron una calificación un poco por debajo del punto de corte (quienes no recibieron el Bono), podemos encontrar el efecto del programa en los logros académicos de los estudiantes.

La idea principal del método RD es que el programa se asigna de acuerdo con una variable observable continua, relativa a un punto de corte dado. En nuestro caso, esta variable es el puntaje de evaluación de los docentes. Además, la probabilidad de recibir el tratamiento es una función de esta variable discontinua en el punto de corte. El supuesto principal detrás de un diseño de discontinuidad es que las características no observables varían continuamente alrededor del punto de corte, y se utilizan características observables para determinar el

⁹ Caridad Araujo, et al. (2016) demostró que las características de estos maestros tuvieron un efecto significativo en los estudiantes ecuatorianos.

tratamiento¹⁰. Este supuesto puede no ser cierto si los individuos pueden cambiar su posición en relación con el punto de corte (Jacob y Lefgren, 2004). El supuesto de que no hubo manipulación de los puntajes alrededor del punto de corte debe ser probado. Estos resultados se presentan en la siguiente sección.

La idea detrás de un diseño de RD es que la comparación de individuos dentro de un intervalo muy pequeño alrededor del punto de corte es similar a un experimento aleatorio en el punto de corte. En este sentido, se espera que los individuos justo por encima de ($S > S_0$) y por debajo ($S < S_0$) el punto de corte (S_0) sean muy similares, y por lo tanto, tengan resultados comparables tanto en presencia como en ausencia de programa. Teniendo esto en cuenta, una comparación entre los que están por encima del punto de corte (quienes reciben tratamiento) y los que están por debajo (que no lo reciben), proporciona un estimador del efecto del tratamiento potencialmente imparcial.

En la literatura, hay dos tipos de diseño RD: "agudo" (sharp) y "difuso" (fuzzy). Bajo el diseño "agudo", los individuos se seleccionan para el tratamiento solo en función de una variable continua: aquellos que se encuentran debajo del punto de corte no reciben el tratamiento ($T_i = 0$ si $S_i < S_0$), mientras que los que están por encima del punto sí lo hacen. ($T_i = 1$ si $S_i \geq S_0$). En este caso, T es determinístico y solo depende del puntaje de evaluación del maestro $T_i = f(S_i)$. (Hahn, Todd y Van der Klaauw, 2001). Un problema potencial en nuestro caso es que el proceso de asignación del tratamiento no depende exclusivamente de la evaluación del maestro. Algunos maestros que recibieron una bonificación pueden estar por debajo del puntaje requerido, y otros con puntajes más altos pueden ser excluidos. Esto significa que la asignación del tratamiento depende de la evaluación del maestro, pero de una manera estocástica. En la literatura, este caso corresponde a un diseño "difuso".

Además de los puntajes de los maestros, hay otras variables que pueden influir en la asignación del tratamiento. Un punto importante a considerar es que, incluso en este caso, la función de probabilidad de participación en el programa (función de puntaje de propensión) $Pr(T = 1 | S)$ tiene una discontinuidad en el punto de corte, S_0 , como en el diseño "agudo". El diseño "difuso" puede ocurrir en los casos en que hay una asignación incorrecta en relación con el punto de corte. El término de error en un diseño "difuso" viene dado por: $T_i = E[T_i | S_i] + e_i$ (Hahn, Todd y Van der Klaauw, 2001). En este caso, se puede usar un método de variable instrumental (IV) para obtener estimadores no sesgados, donde el instrumento (Z) es la regla de asignación, y toma el valor de 1 para aquellos que están por encima del punto de corte y 0 para aquellos por debajo del punto de corte. Este instrumento satisface las condiciones básicas para ser un "buen" instrumento: Z está altamente correlacionada con la participación en el programa (ya que es la regla de asignación), pero no está correlacionada con el logro académico de los estudiantes, cerca del punto de corte. (Las muestras de RD¹¹). La ecuación de forma reducida en este caso es:

¹⁰ El supuesto principal al usar un diseño de RD, para obtener estimaciones imparciales del impacto del programa, es que no hay otras discontinuidades alrededor del punto de corte (80 y 90).

¹¹ La expansión del ancho de banda alrededor de la discontinuidad aumenta el número de observaciones; pero, al mismo tiempo, aumenta el riesgo de violar la tendencia común. Las muestras de RD que se utilizan son ± 2 , ± 4 , ± 6 . Adicionalmente, una ventaja del diseño de RD es que, cuando se utilizan variables

$$Y_i^{RD} = X_i^{RD} \beta_{rd} + f(S_i^{RD}) + \alpha_{rd} Z_i^{RD} + u_i \quad (1)$$

El modelo de MCO en dos pasos puede ser obtenido mediante:

$$Y_i^{RD} = X_i^{RD} \beta_{rd2} + f(S_i^{RD}) + \alpha_{rd2} \hat{T}_i^{RD} + e_i \quad (2)$$

donde $\hat{T}$ viene de la siguiente ecuación del primer-paso:

$$T_i^{RD} = X_i^{RD} \delta_1 + f(S_i^{RD}) + \delta_3 Z_i^{RD} + w_i \quad (3)$$

Donde el superíndice RD indica la muestra de regresión discontinua. Y_i es el puntaje promedio de los niños en varias disciplinas, de los estudiantes asignados al maestro i . X_i representa un vector de variables de control a nivel escolar y docente; S_i es el puntaje del maestro en su evaluación; T_i es el indicador de tratamiento. Alfa es el coeficiente de interés e indica el efecto causal promedio del tratamiento en el punto de discontinuidad (Imbens y Lemieux, 2007). El resultado bajo el método VI proporciona un efecto de tratamiento promedio local (LATE por sus siglas en inglés) y se refiere al impacto del programa en aquellos cuya participación puede ser manipulada con la regla de asignación (con el instrumento), es decir, aquellos que se encuentran alrededor del corte Punto (Imbens y Lemieux, 2007).

Resultados

Para comenzar, a continuación se presentan las estadísticas descriptivas que comparan la información de los maestros que están cerca del punto de corte (entre 78 y 82 puntos en la evaluación):

Tabla 2: Comparación de medias entre Tratamiento y Control

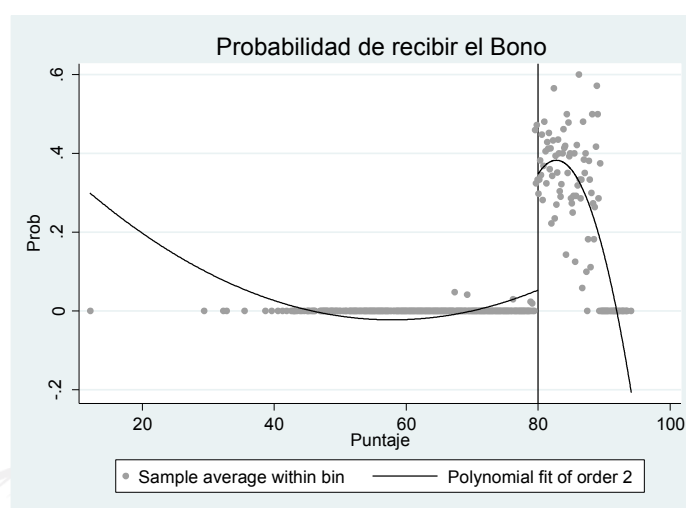
Variable	Control	DE	Tratamiento	DE	p-value
BGU	0.638	-0.481	0.641	-0.48	[0.900]
BI	0.068	-0.251	0.053	-0.225	[0.297]
Costa	0.361	-0.481	0.292	-0.455	[0.013]
Sierra	0.564	-0.496	0.643	-0.479	[0.006]
Amazonía	0.071	-0.257	0.065	-0.246	[0.668]
Escuelas Públicas	0.992	-0.088	0.996	-0.062	[0.356]
Escuelas Municipales	0.002	-0.04	0	0	[0.318]
Escuelas Bilingües	0.044	-0.205	0.025	-0.155	[0.067]
Puntaje de la evaluación docente	50.808	-6.638	50.32	-6.874	[0.349]

instrumentales, se puede justificar un estimador de Wald, incluso cuando se viola la condición de exclusión, el instrumento no se relaciona con el término de error (Hahn, Todd y Van Der Klaue, 1999 y 2001).

Profesores con maestría	0.159	-0.366	0.194	-0.395	[0.126]
Profesores con tercer nivel	0.406	-0.492	0.364	-0.482	[0.143]
Profesores con nivel técnico-tecnológico	0.016	-0.125	0.011	-0.106	[0.519]
N	635	527			

En general, los docentes que están por encima del punto de corte, en comparación con los que están por debajo, son muy similares. No se encontraron diferencias significativas en las variables presentadas en la tabla anterior, excepto en los indicadores regionales. Hay un mayor porcentaje de maestros que pertenecen a la sierra, que están por encima del punto de corte en comparación con los de abajo. Una característica importante en un análisis de regresión discontinua es verificar que, en efecto, hay una discontinuidad en el proceso de selección de bonificación, en el punto de corte. El siguiente gráfico nos permite observar que efectivamente existe una discontinuidad.

Figura 1. Probabilidad de recibir bono en el puntaje de evaluación. Punto de corte 80.



La probabilidad de recibir el bono tiene una discontinuidad estadísticamente significativa alrededor del punto de corte, 80.

Como se mencionó en la parte metodológica, estamos trabajando con un diseño "difuso". En este sentido, es importante verificar que nuestro instrumento (Z_i), que es la regla de asignación, es un buen predictor para la variable de tratamiento real (T_i). Para hacer esto, los resultados de la estimación de la primera etapa se presentan a continuación, donde T se regresa en Z . La Especificación 1 incluye, además del puntaje en forma polinómica cúbica, variables docentes como la edad y los indicadores de nivel educativo (maestría, tercer nivel y tecnólogo /técnico). La especificación 2 también incluye variables escolares (las mismas que se informan en la tabla 2). Finalmente, la especificación 3 incluye indicadores cantonales. En todos los casos, el coeficiente es significativo y con una capacidad predictiva de alrededor del 37%. El valor F para el instrumento es, en todos los casos, mayor que 10. Esto demuestra que nuestro instrumento es un buen predictor de la variable de tratamiento real. Los resultados para el punto de corte de 90 se

presentan en el apéndice. En este caso el instrumento no es lo suficientemente fuerte. Por lo tanto, los resultados para ese punto de corte deben tomarse con precaución.

Tabla 3: Primera-etapa

Variable	1	2	3
Z1	0.377	0.376	0.377
St.err	(0.027)	(0.027)	(0.027)
N	3814	3814	3814
r2	0.265	0.267	0.327
F value for the instrument	194.37	193.6	198.28

Errores estándar en paréntesis

La siguiente tabla presenta los resultados para los puntajes totales de las pruebas SER BACHILLER de los estudiantes, que incluyen: matemáticas, lenguaje, estudios sociales y ciencias naturales. Los resultados se informan para tres vecindarios diferentes alrededor del punto de corte: ± 2 , ± 4 , ± 6 . Además, los resultados se presentan utilizando mínimos cuadrados ordinarios (MCO) y mínimos cuadrados de dos etapas (2-SLS) (donde T está instrumentado por Z). Se utilizan las mismas tres especificaciones utilizadas en la primera etapa. En ninguno de los casos encontramos un efecto significativo de la bonificación.

Tabla 4. Estimadores MCO y VI sobre el puntaje promedio. Diferentes muestras RD.

Variable	RD. 78-82					
	MCO			VI		
	1	2	3	1	2	3
T	0.087	0.045	-0.067	-0.286	-0.377	-0.475
	(0.113)	(0.097)	(0.107)	(0.637)	(0.563)	(0.575)
N	686	686	686	686	686	686
Variable	RD. 76-84					
	MCO			VI		
	1	2	3	1	2	3
T	-0.016	-0.011	-0.082	-0.386	-0.284	-0.044
	(0.09)	(0.079)	(0.079)	(0.494)	(0.442)	(0.513)
N	1348	1348	1348	1348	1348	1348
Variable	RD. 74-86					
	MCO			VI		
	1	2	3	1	2	3
T	-0.022	-0.023	-0.081	-0.163	-0.189	0.11
	(0.077)	(0.068)	(0.067)	(0.725)	(0.625)	(0.692)
N	1961	1961	1961	1961	1961	1961

Errores estándar en paréntesis

La siguiente tabla reporta los mismos resultados para las matemáticas. En ninguno de los casos encontramos un efecto significativo.

Tabla 5. Estimadores MCO y VI para matemáticas. Diferentes muestras RD.

Variable	RD. 78-82					
	OLS			IV		
	1	2	3	1	2	3
T	-0.106 (0.12)	-0.149 (0.101)	-0.113 (0.114)	0.23 (0.659)	0.151 (0.547)	0.364 (0.603)
T	RD. 76-84					
	OLS			IV		
	-0.172 (0.09)	-0.147 (0.078)	-0.14 (0.08)	-0.255 (0.488)	-0.173 (0.433)	0.279 (0.525)
T	RD. 74-86					
	OLS			IV		
	-0.162 (0.078)	-0.15 (0.068)	-0.119 (0.069)	0.519 (0.75)	0.324 (0.617)	0.713 (0.737)

Errores estándar en paréntesis

La siguiente tabla presenta los resultados por idioma. No hay un efecto significativo de la bonificación de enseñanza en el rendimiento del estudiante en el idioma.

Tabla 6. Estimadores MCO y VI para lenguaje. Diferentes muestras RD

Variable	RD. 78-82					
	OLS			IV		
	1	2	3	1	2	3
T	-0.065 (0.106)	-0.116 (0.089)	-0.104 (0.106)	-0.381 (0.618)	-0.269 (0.54)	-0.101 (0.555)
T	RD. 76-84					
	OLS			IV		
	-0.113 (0.085)	-0.117 (0.077)	-0.115 (0.078)	-0.578 (0.473)	-0.323 (0.431)	0.195 (0.526)
T	RD. 74-86					
	OLS			IV		
	-0.1 (0.074)	-0.108 (0.067)	-0.09 (0.068)	-0.164 (0.709)	0.032 (0.599)	0.334 (0.708)

Errores estándar en paréntesis

No encontramos resultados significativos también para estudios sociales y ciencias naturales. Ver el Apéndice.

Tampoco hay efecto para el punto de corte de 90. La siguiente tabla presenta los resultados para el promedio general.

Tabla 7. Estimadores MCO y VI sobre el puntaje promedio. Diferentes muestras RD alrededor de 90.

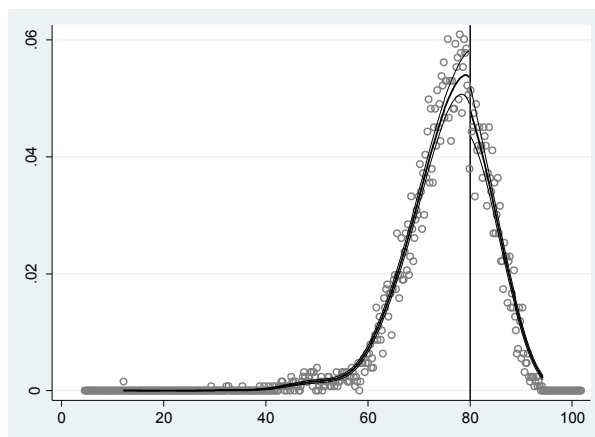
Variable	RD. 88-92					
	OLS			IV		
	1	2	3	1	2	3
T	0.303	-0.162	0.048	-2.705	-1.867	0.347
	(0.284)	(0.201)	(0.273)	(2.239)	(2.391)	(1.898)
N	98	98	98	98	98	98
Variable	RD. 86-94					
	OLS			IV		
	1	2	3	1	2	3
T	0.204	0.025	-0.115	-2.196	0.353	0.222
	(0.176)	(0.145)	(0.202)	(2.058)	(1.574)	(2.318)
N	213	213	213	213	213	213
Variable	RD. 84-86					
	OLS			IV		
	1	2	3	1	2	3
T	0.102	0.014	-0.065	0.137	0.541	-0.759
	(0.116)	(0.098)	(0.119)	(2.81)	(2.385)	(5.558)
N	383	383	383	383	383	383

Errores estándar en paréntesis

Como se mencionó en la descripción del programa, los maestros que obtuvieron una calificación superior a 90 recibieron una bonificación más alta (\$ 1,200). En este sentido, el análisis del efecto del programa, en este punto de corte, nos permite evaluar el efecto marginal de recibir \$ 1,200 versus \$ 900. Los resultados para el promedio general no encontraron un efecto significativo.

Una de las suposiciones del diseño de discontinuidad de regresión es que los beneficiarios o ejecutores del programa no pueden manipular la puntuación en el punto de corte para beneficiar a las personas que no deberían recibir el beneficio. Para probar que no hay manipulación, uno puede evaluar si no hay saltos en la distribución de la variable de asignación en el punto de corte. Para probar este supuesto, utilizamos la prueba de McCrary. El siguiente gráfico presenta los resultados para el primer punto de corte. Los resultados del punto de corte de 90 se presentan en el anexo.

Figura 2. McCrary test de no manipulación, punto de corte 80.



McCrary test: -0.10 con un error estándar de 0.07.

Según la prueba, no hubo manipulación en el punto de corte de 80. Se observan resultados similares en el punto de corte de 90. Consulte el apéndice.

Conclusiones

Este estudio evalúa un programa de incentivos monetarios (bonificación para maestros) vinculado a las calificaciones de las evaluaciones de los maestros. El programa otorgó un incentivo de \$ 900 a los maestros que obtuvieron una calificación de al menos 80, y \$ 1,200 para aquellos que obtuvieron una calificación de al menos 90. Los pagos se entregaron anualmente durante 4 años.

Estas discontinuidades en la regla de asignación (80 y 90 puntos) se usan como nuestra fuente de identificación. Comparamos las calificaciones de los alumnos de los maestros que obtuvieron una calificación un poco inferior a 80 (y que no recibieron la bonificación), con los alumnos de los profesores que obtuvieron una calificación ligeramente superior a 80 (y que sí recibieron la bonificación). Lo mismo se puede hacer para el otro punto de corte: 90 puntos. En este caso, existe el efecto marginal de recibir \$ 1,200 en comparación con recibir \$ 900. Por lo tanto, se evalúa el impacto del incentivo económico en el rendimiento académico de los estudiantes de secundaria en 2015, en varias disciplinas y en el promedio general.

Tal vez, porque el programa se canceló antes de que los maestros premiados recibieran los pagos prometidos, o porque la evaluación no tomó en cuenta las características de otros maestros además de su conocimiento, o porque no vinculó a los maestros con los logros de sus estudiantes, el estudio no encontró ningún impacto de las bonificaciones de los profesores. Esto significa que el incentivo financiero recibido por los maestros debido a su buen puntaje de evaluación no se reflejó en una mejora en los resultados de sus estudiantes.

Referencias

- Becker, G. (1994). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. The University of Chicago Press.
- Caridad Araujo, M., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher Quality and Learning Outcomes in Kindergarten. *IZA Discussion Papers 9796, Institute for the Study of Labor (IZA)*, 1-26.
- Dee, T., & Wyckoff, J. (2015). Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, 34(2), 267–297 .
- Figlio, D., & Lawrence, K. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, 901-914.
- Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, 32(2), 373–407.
- Fryer, R. G., Levitt, S. D., List, J. A., & Sad, S. (2012). Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment. *NBER Working Paper* , No 18237.
- Glazerman, S., & Seifullah, A. (2012). *An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years (Final Report)*. Washington, DC: Mathematica Policy Research, Inc.
- Glewwe, P. (2009). Teacher Incentives in the Developing World. En M. G. Springer, *Performance Incentives: Their Growing Impact on American K-12 Education*. Washington D.C: Brookings Institution Press.
- Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 205–227.
- Hahn, Jinyong; Todd Petra; Van der Klaauw Wilbert (1999). "Evaluating The Effect of an Antidiscrimination Law using a Regression-Discontinuity Design". *NBER Working Paper*. N.º 7 131: 1-40.
- Hahn, Jinyong; Todd Petra; Van der Klaauw Wilbert (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design". *Econometrica*. Vol. 69, N.º 1: 201-209.
- Hanushek, E., & Rivkin, S. (2006). Teacher Quality. In *Handbook of Economics of Education* (pp. 1051-1078).
- Hirano K; Imbens G y Ridder G. (2003). "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score". *Econometrica*. Vol. 71, N.º 4: 1 161-1 189.

Imbens, Guido y Thomas Lemieux. (2007). “Regression Discontinuity Designs: a Guide to Practice”. *National Bureau of Economic Research*. Technical Working paper 337: 1-34.

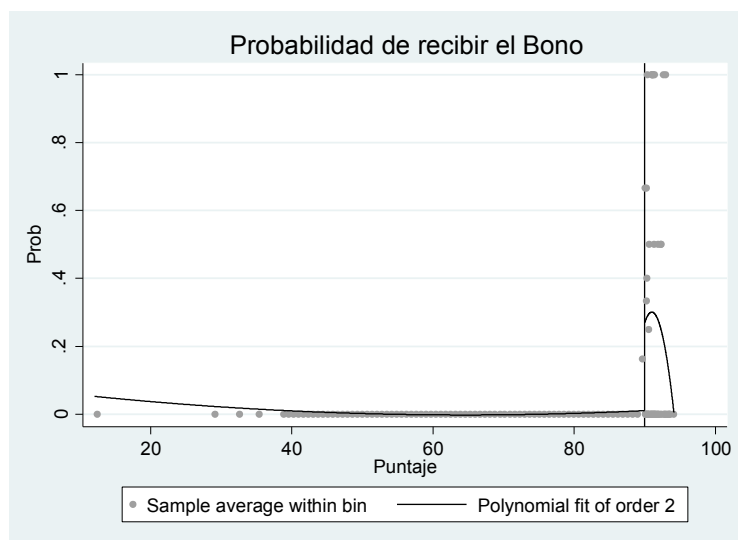
Lavy, V. (2009). Performance pay and teachers’ effort, productivity, and grading ethics. *American Economic Review*, 1979–2011.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1), 39–77.

Neal, D. (2011). The design of performance pay in education. En E. A. Hanushek, S. J. Machin, & L. Woessmann, *Handbook of the Economics of Education: Vol. 4* (págs. 495–550). Amsterdam:: Elsevier.

Apendice

A 1: Probabilidad de recibir bono en el puntaje de evaluación. Punto de corte 90.



A 2: Primera-etapa. Punto de corte 90.

Variable	1	2	3
Z2	0.158 (0.07)	0.159 (0.07)	0.156 (0.066)
N	3814	3814	3814
r2	0.112	0.113	0.163
F	5.08	5.08	5.52

Error estandar en paréntesis

A 3: Estimadores MCO y VI para estudios sociales. Diferentes muestras RD.

Variable	RD. 78-82					
	OLS			IV		
	Esp_1	Esp_2	Esp_3	Esp_1	Esp_2	Esp_3
T1	-0.069 (0.105)	-0.114 (0.083)	-0.113 (0.097)	-0.032 (0.596)	-0.242 (0.48)	-0.235 (0.531)
T1	RD. 76-84					
	OLS			IV		
	-0.156 (0.082)	-0.12 (0.07)	-0.104 (0.071)	-0.15 (0.452)	-0.137 (0.391)	0.339 (0.484)
T1	RD. 74-86					
	OLS			IV		
	-0.143	-0.124	-0.09	0.348	0.081	0.648

	(0.071)	(0.061)	(0.061)	(0.722)	(0.571)	(0.683)
--	---------	---------	---------	---------	---------	---------

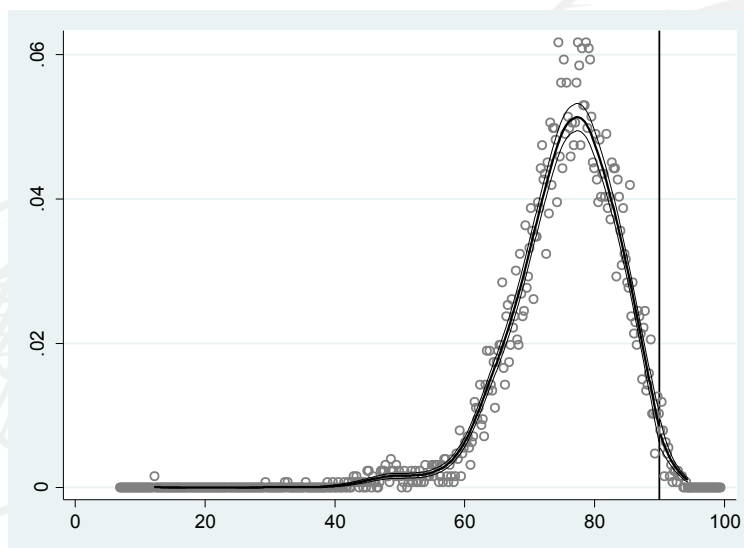
Error estandar en paréntesis

A 4: Estimadores MCO y VI para ciencias. Diferentes muestras RD.

Variable	RD. 78-82					
	OLS			IV		
	1	2	3	1	2	3
T	0.027 (0.112)	-0.011 (0.089)	-0.086 (0.102)	0.148 (0.637)	-0.046 (0.52)	-0.092 (0.567)
T	RD. 76-84					
	OLS			IV		
	-0.078 (0.088)	-0.044 (0.074)	-0.077 (0.077)	-0.403 (0.492)	-0.404 (0.423)	-0.044 (0.497)
T	RD. 74-86					
	OLS			IV		
	-0.061 (0.075)	-0.044 (0.064)	-0.035 (0.065)	0.083 (0.729)	-0.151 (0.589)	0.433 (0.689)

Error estandar en paréntesis

A 5: McCrary test de no-manipulación, punto de corte 90



McCrary test: -0.04 con un error estandar de 0.19.