

Centro de Investigación y Docencia Económicas A.C.



**Evaluación de impacto a políticas educativas.
Experiencias en el caso de México**

T E S I S

QUE PARA OBTENER EL GRADO DE:

DOCTOR EN POLÍTICAS PÚBLICAS

P R E S E N T A

JAVIER ADRIÁN GUÍZAR MONTERO

Asesor de tesis:

Dr. Rubén Irvin Rojas Valdés

Ciudad de México

Junio 2022

A Paulina, por su amor

Los conceptos desarrollados, análisis realizados y conclusiones de la presente tesis son responsabilidad exclusiva del autor, y no reflejan la postura del Centro de Investigación y Docencia Económicas. Lo mismo ocurre con cualquier error presente en este trabajo.

ÍNDICE

Listado de figuras	viii
Listado de gráficas.....	ix
Listado de tablas	x
Resumen	1
Agradecimientos	2
Introducción general	6
Capítulo 1. Las evaluaciones de impacto como herramienta para el estudio de políticas públicas en México	12
1.1. Introducción.....	12
1.2. Marco teórico.....	15
1.2.1. Los diseños experimentales en evaluación de impacto	18
1.2.2. Los diseños cuasiexperimentales o no experimentales.....	23
1.2.2.1. Regresión discontinua	23
1.2.2.2. Diferencias en diferencias	24
1.2.2.3. Método de aparejamiento (matching).....	25
1.2.2.4. Variables instrumentales	28
1.2.2.5. Diseño <i>pipeline</i> o de línea de espera	29
1.3. Metodología para la revisión de literatura.....	29
1.4. Evaluaciones de impacto para el caso mexicano.....	32
1.4.1. Evaluaciones de impacto de <i>PROGRESA/Oportunidades/Prospera</i>	33

1.4.2. Evaluaciones de impacto a programas federales	39
1.4.3. Otras evaluaciones de impacto	50
1.5. Limitaciones	54
1.6. Conclusión: ¿Qué pueden aportar las evaluaciones de impacto en México?	55
1.7. Referencias	58
Capítulo 2. Escuelas de doble turno: una evaluación con métodos no experimentales	81
2.1. Introducción.....	81
2.2. El doble turno	82
2.3. Descripción de la intervención	91
2.4. Metodología.....	93
2.4.1. Datos.....	96
2.4.2. Análisis descriptivo	100
2.4.3. Metodología de aparejamiento por marcador de propensión	104
2.4.3.2. Procedimiento para el aparejamiento	107
2.5. Resultados del diseño cuasiexperimental	117
2.5.1. Limitaciones	120
2.6. Conclusión.....	123
2.7. Referencias	125
Anexo 2A.....	133
Anexo 2B.....	137
Anexo 2C.....	142

Capítulo 3. Una intervención para mitigar la pérdida de habilidades de lectura en verano	149
3.1. Introducción.....	149
3.1.1. Definición del problema	150
3.1.2. Una intervención en verano para mejorar el desempeño en lectura	152
3.2. Revisión de literatura.....	154
3.2.1. La pérdida de habilidades lectoras.....	154
3.2.1.1. La pérdida de habilidades lectoras en México.....	156
3.2.2. Experimentos para mitigar la pérdida de habilidades lectoras	158
3.3. Metodología.....	164
3.3.1. Diseño del estudio	164
3.3.2. Aleatorización y recolección de los datos	164
3.3.2.1. Participantes	165
3.3.2.2. Instrumentos	167
3.3.2.3. Metodología de evaluación y línea basal.....	168
3.3.3. Balance del tratamiento	172
3.4. Resultados.....	174
3.4.1. Efectos capturados por condición de tratamiento.....	175
3.4.2. Discusión de resultados	177
3.5. Conclusión y recomendaciones de política pública.....	181
3.6. Referencias	185
Anexo 3A.....	196
Anexo 3B.....	198
Anexo 3C	199

Capítulo 4. Conclusiones generales 206

Glosario de siglas 210

LISTADO DE FIGURAS

Capítulo 2

Figura 2.1 Línea de tiempo de la intervención 92

Capítulo 3

Figura 3.1 Línea de tiempo del estudio 169

LISTADO DE GRÁFICAS

Capítulo 2

Gráfica 2.1a-i Distribución de las diferencias entre turnos; resultados seleccionados. Escuelas primarias	108
Gráfica 2.2a-i Distribución de las diferencias entre escuelas pareadas (NN3). Variables de interés	115
Gráfica 2.4a-i Diferencias en las tendencias entre grupos. Cambios en resultados seleccionados (NN3). Escuelas primarias.....	119

Capítulo 3

Gráfica 3.1 Distribución de los encuestados en los distintos niveles socioeconómicos de la AMAI	171
Gráfica 3.2 Distribución de la diferencia en puntajes entre alumnos en los grupos de tratamiento y control.....	177
Gráfica 3.3 Tamaño de muestra requerido para obtener efectos significativos.....	178

LISTADO DE TABLAS

Capítulo 1

Tabla 1.1. Resultados de documentos en buscadores.....	31
Tabla 1.2. Evaluaciones de impacto del programa <i>PROGRESA/Oportunidades/PROSPERA</i>	36
Tabla 1.3. Evaluaciones de impacto a programas federales en educación en México	40
Tabla 1.4. Evaluaciones de impacto de otros programas federales en México	45
Tabla 1.5. Otras intervenciones en México que utilizan evaluación de impacto	51

Capítulo 2

Tabla 2.1. Efectos de la implementación de escuelas de doble turno alrededor del mundo	84
Tabla 2.2. Valores promedio y desviaciones estándar, panel de escuelas primarias.....	95
Tabla 2.3.1. Regresión MCO de la diferencia entre turnos en escuelas primarias en el grupo de tratamiento (primer grado, ciclo escolar 2007-2008 a 2011-2012).....	102
Tabla 2.3.2. Regresión MCO de la diferencia entre turnos en escuelas primarias en el grupo de tratamiento (tercer grado, ciclo escolar 2010-2011).....	103
Tabla 2.4. Escuelas en los grupos de tratamiento y control, por turno	107
Tabla 2.5. Modelo logístico de participación en el tratamiento	110
Tabla 2.6. Escuelas en los grupos pareados de tratamiento y control, por turno	114
Tabla 2.7. Impactos estimados utilizando diferencias en diferencias (NN3); resultados seleccionados en escuelas primarias pareadas.....	121

Tabla 2.9. Impactos estimados utilizando diferencias en diferencias (NN3); otros resultados seleccionados en escuelas primarias pareadas.....	122
-----------------------------------------------------------------------------------------------------------------------------------------------	-----

Capítulo 3

Tabla 3.1. Prueba de balance de las características entre grupos	172
Tabla 3.2. Características y puntajes promedio antes del tratamiento.....	173
Tabla 3.3. Auto-reporte de lectura de la población en el grupo de tratamiento	175
Tabla 3.4. Diferencias entre los puntajes de la prueba ENLACE antes y después del verano	176

RESUMEN

Evaluación de impacto a políticas educativas.

Experiencias en el caso de México

Esta tesis se compone de tres documentos de investigación que desarrollan, en la teoría y en la práctica, distintas aplicaciones de las herramientas que constituyen la Metodología de Evaluación de Impacto; aplicándolas al contexto mexicano con un énfasis en políticas educativas. Con este trabajo se busca abonar a la comprensión de la importancia que ha cobrado esta metodología en nuestro país, debido a la certeza con que permite determinar los impactos de un programa o política sobre uno o varios resultados de interés.

Las políticas evaluadas en esta tesis resultan novedosas en México; y tienen importantes implicaciones de política educativa; pudiendo extrapolarse a otros contextos, principalmente de América Latina. El objetivo final es promover el uso de evidencia robusta para estimar los efectos que distintas políticas públicas tienen sobre el desarrollo y bienestar de la población.

AGRADECIMIENTOS

Quiero agradecer en primer lugar a mi madre, sin cuyo inagotable apoyo no hubiera podido completar esta tarea, y a quien debo mis logros y mi ser. A mi papá y a mis segundos padres: mis abuelitos. A mis tíos Aída, Débora, Graciela, Adrián, Julián, Marino y René; y a sus familias. A mi primo David, quien ha sido mi hermano toda la vida y a mis demás primos, esperando que este logro los motive aún más.

A Roberto Gerhard, Roslyn Jiménez, Raúl Arana y todo el equipo de la DGIET. A Eloy, colega y amigazo. A la Dra. Hilda Lokpez, con mi más profundo respeto. A todos mis familiares y amigos, quienes han hecho mi vida productiva y placentera tanto en el CIDE como fuera de él. En el ámbito académico, quiero agradecer en primer lugar el apoyo de mi director de tesis, Irvin Rojas, cuya asesoría ha permitido alcanzar el nivel de calidad del trabajo aquí presentado. El resultado final se vio beneficiado sobremanera a partir de sus observaciones. A mis lectores: Víctor Carreón -por su apoyo y motivación para iniciar y concluir con esta meta- y a Gustavo del Ángel por su apoyo, acompañamiento e interés en el seguimiento de mi proceso doctoral a lo largo de todas sus etapas. A Eva Arceo y Carlos Vilalta; quienes enriquecieron grandemente esta tesis y mi investigación; fueron maestros, consejeros y amigos. En el caso de la profesora Arceo, en particular, por su apoyo en el desarrollo de la metodología de aparejamiento. En el caso del profesor Vilalta, en particular, por brindarme la oportunidad de enseñar en el CIDE como laboratorista por dos años consecutivos. A Sergio Cárdenas, de quien aprendí a lo largo del doctorado, y cuya enseñanza y consejos ayudaron en las etapas de investigación y desarrollo profesional contenidas en esta tesis. Espero que nuestro tiempo juntos haya contribuido a que ambos seamos mejores personas.

Sé que no es usual agradecer a una institución y no a una persona, pero estudiar en el CIDE ha significado para mí vivir bajo normas que parecerían reservadas a un mundo utópico, como el descrito por Tomás Moro en su más famoso libro. Mi sentir ha sido el de caminar de la mano de gigantes. Estudiar en él ha abierto mis ojos a la fuerte necesidad de que la verdad y el mérito prevalezcan. Mirar la realidad desde diferentes escenarios, me ha mostrado que el esfuerzo tiene distintos alcances en las distintas circunstancias de vida; lo que me ha ayudado a ser más empático.

Si tuviera algo que destacar -de lo mucho en lo que esta universidad ha influido positivamente en mi vida- es la convicción y el deseo de que sea ésta la calidad de educación superior a la que aspiren todos aquellos a quienes amo. No concibo un mejor camino para México, ni para el mundo, que aquél en el que el acceso a esta educación (en todos los ámbitos del saber) no dependa más que del mérito y el esfuerzo académico, sea asequible al mayor número posible de alumnos, y genere en ellos los valores que, quizá arrogante e infundadamente, pregoné que ha inculcado en mí. Pido a Dios la oportunidad de contribuir a que más jóvenes y adultos mexicanos tengan una oportunidad de aprendizaje tan rica como la mía; allanando el camino para que sus logros superen a los de aquellos que les precedimos. Haré lo bueno que en mis manos esté para lograr tal propósito; que no será sino dar algo de lo mucho que he recibido.

A Conacyt, por la beca brindada que permitió financiar mis estudios, ayudó a mi manutención, y sin la cual no hubiese elegido este camino. A mi antiguo jefe, Enrique Cárdenas Sánchez, quien también me motivó a emprender esta aventura, y también me infundió el valor para iniciarla. Al coordinador del Doctorado en Políticas Públicas, Mauricio Dussauge, por su incondicional apoyo; y a Alberto Casas. Al antiguo coordinador del doctorado en Políticas Públicas, Eduardo Villarreal, así como a su antecesor, Víctor Figueras, por su sobresaliente apoyo al alumnado, calidez y trato humano. A los especialistas en ética, justicia y transparencia

María Amparo Casar, Ana Laura Magaloni, Layda Negrete, David Arellano, Guillermo Cejudo, José Ramón Gil, Sergio López Ayllón, Alejandro Madrazo, Mauricio Merino, Edgar Ramírez, Ricardo Raphael, José Roldán Xopa, Juan Salgado y Gabriel Purón; por forjar en mí una fuerte convicción de que la equidad, el mérito y la verdad son el mejor camino, y deben ser criterios rectores tanto en el servicio público y la academia como en la vida privada; aún cuando en la práctica no siempre sean los más fáciles de elegir o hacer prevalecer.

A los profesores David Crow y Alfonso Miranda, por su apoyo incondicional en la enseñanza de métodos aplicados en Stata[®]. A los profesores Blanca Heredia, Susan Parker, Alejandra Ríos, Javier Aparicio, Rodolfo Cermeño (q.e.p.d.), Robert Duval, David González, Fausto Hernández, Ignacio Lozano, David Juárez, Raúl Pacheco-Vega, Alejandro Villagómez y a todos aquellos a cuyas cátedras tuve la fortuna de asistir; aunque sé que me he beneficiado indirectamente aún de aquellos profesores con quienes no tomé clases, así como de las enseñanzas de múltiples investigadores externos al CIDE, a quienes hago extensivo mi agradecimiento.

A los compañeros de las primeras siete generaciones del doctorado con quienes he compartido experiencias, conocimientos y visiones del mundo; y quienes en más de una ocasión me han tendido la mano. A los alumnos de las generaciones 2012 y 2013 de la Licenciatura en Derecho; a quienes fue todo un gusto enseñar y de quienes no he dejado de aprender. Un agradecimiento especial para Anaid Reyes, por la entrega en formato electrónico de las bases de datos del proyecto *Veranos de lectura* del CIDE y la explicación del proceso de su implementación, que se describe en el capítulo 3 de esta tesis. A David Sánchez y todo el equipo del proyecto de educación inicial. A Luz y Alma Barajas, quienes desde la Coordinación me brindaron un enorme apoyo en las gestiones, trámites y requisitos administrativos; al grado de sentirme francamente apapachado.

Para Ana Laura Mar, Mary, Nuri, Don Jaime y todo el personal de biblioteca del CIDE que me apoyó para encontrar hasta la última referencia bibliográfica contenida en esta tesis. A todas las personas que laboran en el CIDE y facilitaron mi vida de estudiante (servicios administrativos, cafetería, comedor, servicios generales, etc.). Gracias a todos.

INTRODUCCIÓN GENERAL

"Para realizar el viaje más difícil solo debemos dar un paso a la vez, pero debemos seguir andando"

Proverbio chino

El hilo conductor de esta tesis es el estudio de implementaciones de la metodología de evaluación de impacto (MEI) en México. El Capítulo 1 presenta una revisión exhaustiva de documentos de investigación publicados en revistas científicas, instituciones gubernamentales o instituciones de educación superior que han utilizado la MEI para evaluar políticas públicas de México, desde la década de los 90 hasta 2013. Se sistematizó dicha revisión, buscando que la publicación fuera acreditada para garantizar un mínimo de calidad y rigurosidad. El capítulo comienza con un recuento de los distintos diseños de evaluación que componen dicha metodología, seguido de una estrategia para recopilar los documentos de interés al presente trabajo; terminando con una cifra de 59. Cada una de estas investigaciones es esquematizada en términos del programa o política que estudia; las variables de interés seleccionadas y los impactos estimados. Con ello, ofrece una perspectiva sobre las tendencias metodológicas que ha seguido esta veta de investigación en México; actores que han influido en su difusión en dicho periodo -programas e intervenciones recurrentes- y termina con una reflexión de los hallazgos; a manera de conclusión.

En el Capítulo 2, el lector encontrará una aplicación de la MEI, explotando una política implementada en un estado de la República Mexicana con el objetivo de generar una asignación más justa entre turnos escolares. La iniciativa de las autoridades permitió llevar a cabo una lotería para decidir qué alumnos son asignados al turno matutino y quiénes son asignados al

turno vespertino. La situación resultante contrasta con asignaciones tradicionales basadas, por ejemplo, en desempeño académico; antecedentes familiares o discrecionalidad de los directores de cada plantel. Estas otras asignaciones han propiciado desigualdades en equidad entre turnos escolares; como verá el lector evidenciado en la literatura revisada y las estadísticas descriptivas presentadas.

En este capítulo se emplea una combinación de diferencias en diferencias y aparejamiento, encontrando una reducción en las brechas de puntajes entre turnos para los alumnos de más bajo desempeño en español. Los resultados muestran, sin embargo, un incremento en las brechas de extraedad y repitencia a lo largo del periodo estudiado, arrojando luz sobre otros efectos que esta política podría tener y llaman a mejores intervenciones acompañadas de un diseño de investigación más limpio y robusto; como lo es el experimental.

En el capítulo 3 se parte de un diseño experimental con el objetivo de paliar los efectos de la llamada *pérdida de habilidades lectoras en verano*. A través de una aleatorización (de alumnos de quinto grado en el nivel de primaria), un grupo de tratamiento que recibió tres libros durante el verano -y sesiones de acompañamiento por parte de sus padres- fue contrastado con sus compañeros; a fin de observar si hubo algún impacto en su desempeño lector.

Con un marco teórico que señala los efectos derivados de mejoras en la habilidad lectora, así como experiencias previas en experimentos con este fin, se explica en el texto la importancia de estudiar políticas como esta en nuestro país. Los resultados -si bien constreñidos por los recursos y la duración de la exposición a los libros durante el periodo observado- permiten vislumbrar efectos positivos que podrían ser estadísticamente significativos si la intervención se potenciase.

Las evaluaciones de impacto como herramienta para el estudio de políticas públicas en México (Capítulo 1)

La implementación de las políticas públicas -tanto en nuestro país como en el resto del mundo- está sujeta a competencia, debate y situaciones específicas al contexto. Por ello, contar con evidencia robusta respecto a los potenciales efectos resulta un valioso insumo para favorecer o rechazar una política. En este primer capítulo se busca aportar al conocimiento de la forma en que la práctica de la evaluación de impacto se ha desarrollado en nuestro país, particularmente en educación. Siendo esta metodología específica para aislar los efectos debidos a un programa o política particular, resulta un importante elemento de cualquier programa de evaluación de políticas públicas. El capítulo comienza describiendo brevemente los dos diseños en que se divide la MEI: los experimentales y los cuasiexperimentales; para posteriormente llevar a cabo una recopilación de las investigaciones que se han llevado a cabo en el contexto mexicano, terminando con una cifra de 59, hasta el año 2013. Tras describir la metodología para llevar a cabo la búsqueda, se clasificaron estas investigaciones a partir del programa o política estudiado. Los resultados muestran que el detonante de la MEI en México fue el programa *PROGRESA/Oportunidades/Prospera*; aunque existen antecedentes de evaluaciones previas al mismo. También se observa que, con excepción de los estudios al programa antes citado, la agenda de investigación en México estuvo dominada por diseños cuasiexperimentales, o de primera generación; siendo hasta años recientes cuando los diseños experimentales han ganado relevancia. El capítulo concluye con una breve reflexión sobre la importancia de una agenda de investigación mediante el uso de evaluación de impacto para nuestro país; así como la importancia de contar con comités de ética para regular su práctica.

Escuelas de doble turno: una evaluación con métodos no experimentales (Capítulo 2)

El doble turno es el arreglo más común para ofrecer diferentes calendarios de trabajo en un mismo plantel educativo. En México, la evidencia disponible señala la existencia de diferencias entre los recursos disponibles para los alumnos en el turno matutino y aquellos en el turno vespertino; en el nivel de primaria. Ello conduce a desventajas sistemáticas para estos últimos, y constituye una ventana de oportunidad para generar políticas públicas orientadas a fomentar la equidad entre turnos escolares en escuelas públicas de México.

El capítulo 2 evalúa una política, llevada a cabo por autoridades estatales de una entidad de la República Mexicana, consistente en una lotería para asignar de manera aleatoria a los alumnos a los turnos matutino y vespertino en las primarias públicas de dicho estado. La hipótesis que se explora es si esta asignación podría conducir a una reducción en la brecha de desempeño entre turnos escolares. Para llevar a cabo la evaluación se hizo uso de los datos administrativos de las escuelas; los resultados de los exámenes de la prueba ENLACE y del programa de desarrollo humano *Oportunidades*.

Se optó por utilizar un diseño cuasiexperimental de diferencias en diferencias y aparejamiento para buscar un grupo de comparación que permitiese comparar a las escuelas de la entidad donde se implementó el tratamiento con un grupo de escuelas estadísticamente similares que no recibieron la política.

Los resultados muestran una reducción en la brecha de reprobación entre el turno vespertino y el matutino después del primer año de la intervención; pero un incremento de dicha brecha en los años subsecuentes, acusando una dilución de los efectos del programa. Asimismo, se presenta un aumento en las brechas de extraedad y repitencia en los tres ciclos escolares que comprende la intervención, lo que podría indicar una recomposición de las cohortes de alumnos

de los turnos matutino y vespertino cuando transitaron al segundo y tercer grado en el nivel de primaria.

Finalmente, los resultados de la prueba ENLACE son congruentes con una reducción en la brecha entre turnos en el examen de Español -para los alumnos de más bajo desempeño. El capítulo termina con la recomendación de estudiar políticas innovadoras acompañadas de un diseño de investigación experimental; para llevar a cabo una evaluación más robusta sobre los efectos de intervenciones orientadas a fomentar la equidad y mejorar el desempeño de los alumnos en las escuelas públicas -particularmente en el turno vespertino- debido al potencial efecto positivo sobre la vida de millones de niños mexicanos.

Una intervención para mitigar la pérdida de habilidades de lectura en verano (Capítulo 3)

Los beneficios de inculcar el gusto por la lectura y el hábito de leer se cosechan durante toda la vida, y tienen repercusiones para el desarrollo de habilidades, como matemáticas y de comprensión (Noguero, 2002; Freire, 2006). Abordar el tema de la brecha en habilidades lectoras de niños y jóvenes, en particular, es necesario como un factor para explicar el desempeño escolar, el aprendizaje, e incluso las oportunidades de acceso a niveles educativos superiores.

En el capítulo 3 se estudia la implementación de un diseño experimental cuyo objetivo es mitigar los efectos de la llamada *pérdida de habilidades lectoras en verano*. A través de una aleatorización (de alumnos de quinto grado en el nivel de primaria), un grupo de tratamiento que recibió tres libros durante el verano (y sesiones de acompañamiento por parte de sus padres) fue contrastado con sus compañeros; a fin de observar si hubo algún efecto sobre su desempeño lector.

Utilizando diferencias en diferencias para corregir por un desbalance en las características de la línea base entre alumnos tratados y no tratados, no fue posible encontrar cambios entre el desempeño de los alumnos antes y después del verano, ni en el grupo que recibió los libros ni en el de sus contrapartes. La implementación del programa no tuvo un efecto estadísticamente significativo sobre la población tratada de alumnos, frente al desempeño del grupo de control. En el capítulo se exploran dos posibilidades para este resultado: que el tamaño de muestra no haya sido suficiente dada la baja magnitud del estimador, y que el programa no haya tenido la duración suficiente para que los efectos, de existir, se viesen materializados.

Dichos resultados, pese a no tener significancia estadística, abonan al conocimiento sobre el potencial de estas intervenciones en nuestro contexto, y abren la puerta para explorar políticas con una evidencia de mayores efectos en el plano internacional. Por ejemplo, las intervenciones recurrentes durante dos o más veranos; cuyos efectos acumulativos pueden generar ganancias en el mediano o largo plazo, y pueden buscar un equilibrio entre intensidad y duración; para maximizar el provecho que de ellas puedan obtener los menores.

CAPÍTULO 1

Las evaluaciones de impacto como herramienta para el estudio de políticas públicas en México

“...hay dos clases de magia. Hay una magia que es obra del diablo y que se propone destruir al hombre mediante artificios que no es lícito mencionar. Pero hay otra magia que es obra divina, ciencia de Dios que se manifiesta a través de la ciencia del hombre, y que sirve para transformar la naturaleza, y uno de cuyos fines es el de prolongar la misma vida del hombre”

Humberto Eco, El Nombre de la Rosa.

1.1. Introducción

En México, el estudio de la administración pública ha buscado mejorar la eficiencia y eficacia de las intervenciones gubernamentales (véase por ejemplo Cejudo, 2003 y Sánchez González, 2005). Este esfuerzo se inscribe en una corriente teórica y práctica, de alcance global, cuya finalidad es incrementar la base de conocimiento robusto respecto al impacto de los programas públicos¹.

Hoy en día existen, además, espacios de acción para otros sujetos -provenientes tanto del ámbito internacional como de la sociedad civil organizada (Bardach, 1998; Barzelay, 1992)- que tienen como objetivo incidir sobre la política pública: diseñar, implementar y evaluar planes y programas que tengan un impacto en el bienestar y desarrollo de la población; a través de intervenciones de alcance universal o dirigidas a grupos específicos. Las dificultades de hacerlo

¹ Para una descripción de esta trayectoria véase la narración correspondiente en Ravallion (2015).

no son menores. Una primera dificultad por superar, para implementar nuevas intervenciones, es la resistencia en el ámbito político. Como señala el libro de Merino *et. al.* (2010):

la calidad de las definiciones, la disponibilidad de los medios, la calidad de los liderazgos, la certeza ofrecida por el marco institucional o la complejidad de los intereses en juego, entre otros factores, pueden incidir de modo irreparable en los resultados obtenidos (pp. 16-17).

Esto es, la implementación de una política pública está sujeta siempre a conflictos de interés al tratar de modificar el *statu quo*, por lo que es necesario contar con las herramientas adecuadas (legales, políticas, de negociación, etc.) para vencer dichas inercias.

Vinculada con el problema anterior, se encuentra la necesidad de competir por fuentes de financiamiento. Normalmente, la implementación de políticas se encuentra imbuida en una constante competencia; resultado natural, en primera instancia, de la demanda de recursos (siempre escasos) que enfrentan las diferentes propuestas.

Los recursos pueden provenir, en el caso de las iniciativas impulsadas por los gobiernos, de dinero público (y en última instancia de los contribuyentes) y adicionalmente de patronatos, beneficencias y organizaciones sin fines de lucro. Asimismo, otras organizaciones también financian este tipo de proyectos; por ejemplo, las no gubernamentales o de la sociedad civil. Lo que hermana a todos estos dispensadores es el deseo de ver que los fines que sirvieron de justificación para tales políticas sean alcanzados; y que el uso de los recursos (materiales y humanos) sea eficiente y esté bien canalizado.

Un tercer problema se refiere a la rendición de cuentas a la que están obligadas las intervenciones de política pública. Como se señala en Bovens (2007), la rendición de cuentas efectiva incluye consecuencias que pueden ser formales e informales; y en última instancia “puede ayudar a asegurarse que la legitimidad de la gobernanza se mantenga intacta o se

incremente” (*ibíd.*, p. 464; traducción propia). Estos requisitos hacen de la evaluación una herramienta esencial en el proceso de implementación de políticas (Bracho; en Merino *et. al.*, 2010; *op. cit.*). Como se señala en Desplatz y Ferracci (2016) “si los métodos de evaluación pueden en ocasiones parecer técnicos, en realidad son guiados por preguntas simples: ¿ha cumplido su objetivo la política? ¿ha generado efectos no anticipados por el legislador?” (p. 7, traducción propia).

El capítulo que aquí se presenta está enfocado a revisar la evidencia disponible sobre la efectividad de las políticas en distintos ámbitos de interés. En ésta, la robustez de la teoría de causalidad, que subyace a la estimación de los efectos, aporta evidencia tanto para poner en marcha alguna política pública específica (cuantificando los efectos potenciales), como para rendir cuentas de los resultados alcanzados (cuando la política ya se ha implementado y se quieren medir sus efectos sobre la población tratada). La evaluación de impacto es la única metodología de evaluación con un procedimiento específico para establecer una hipótesis alternativa sobre el estado de cosas, si el programa no se hubiera implementado; haciendo que las consecuencias del programa sean atribuibles a éste y no a otras causas (Gertler, Martínez, Premand, Rawlings y Vermeersch, 2011, p. 33).

Este argumento de causalidad es lo que distingue a la MEI de otro tipo de evaluaciones; ya que busca aislar los efectos debidos al programa de los otros factores que pudieron incidir sobre el resultado observado de la política o programa particular. Como se señala en Gertler *et. al.* (2011, *op. cit.*) “A diferencia de las evaluaciones generales -que pueden responder a diferentes tipos de preguntas- las evaluaciones de impacto se preocupan por saber *cuál es el impacto (o efecto causal) de un programa sobre un resultado de interés.*” (p. 7; cursivas en el original).

Para abonar al conocimiento de las evaluaciones de impacto en México, el capítulo está organizado de la siguiente manera: la sección 1 presenta un marco teórico que describe el

objetivo que persigue la MEI; así como los diferentes diseños que la componen. La sección 2 describe la forma en que se recolectó información de los estudios para México que han recurrido a la MEI, mientras que la sección 3 expone la influencia de esta metodología en el campo de las políticas públicas, a partir de la bibliografía recolectada. Por último, se presentan las limitaciones y conclusiones derivadas de este estudio.

1.2. Marco teórico

La principal característica de las evaluaciones de impacto es que buscan aislar los efectos atribuibles a un programa o política (Dhaliwal, Duflo, Glennerster y Kremer, 2012). Debido a la forma en que el efecto del programa es separado de los otros factores que pudieron influir en los resultados, las evaluaciones de impacto constituyen un insumo para llevar a cabo análisis de costo-beneficio (Dhaliwal *et. al.*, 2012), siendo capaz el evaluador de distinguir -entre una gama de políticas- aquella que presenta el mayor efecto (nuevamente, sobre un resultado específico) al menor costo (Ravallion, 1999; Khandker, Koolwal y Samad, 2010).

Otra forma de expresar lo anterior es decir que una evaluación de impacto tiene como objetivo hallar el **contrafactual**; esto es, el resultado para los participantes en un programa si dicho programa no se hubiera implementado. El contrafactual es imposible de observar en la realidad, por lo que se debe estimar cuál sería el estado de las cosas en ausencia del programa.

Escrito en forma algebraica, siguiendo el ejemplo presente en Duflo, Glennerster y Kremer (2007), sea I el impacto causal de un programa o política particular. Sea Y_1 la medición de la variable de interés para un individuo que ha participado en el programa, y Y_0 la medición de la misma variable para dicho individuo; pero en ausencia del programa. Sea D una variable dicotómica que indica la participación (o no) en el programa o política. Tenemos que:

$$I = E(Y_1 - Y_0 | D = 1) \quad (I)$$

$$= E(Y_1|D = 1) - E(Y_0|D = 1) \quad (\text{II})$$

Esto es, el impacto (I) que tiene un programa es el resultado de aislar el efecto promedio de aquellas personas que participaron en él (también llamados *grupo de tratamiento*), con una medición de sus características una vez que han recibido el programa y, al mismo tiempo, con una medición de las características que poseerían si no lo hubieran recibido. Éste es el contrafactual al que trata de aproximar la evaluación de impacto, o $E(Y_0|D = 1)$ en la ecuación de arriba.

En la práctica resulta imposible medir dos estados distintos para la misma persona al mismo tiempo. En su lugar, utilizamos aproximaciones al mismo mediante un grupo de comparación (también llamado grupo de control). En la medida que las características de ese grupo sean más parecidas a las del grupo de tratamiento, las diferencias que resulten tras la aplicación del programa pueden ser atribuidas a este.

Cuando comparamos grupos tratados y no tratados, estamos capturando el efecto verdadero de una intervención más el sesgo de selección. El sesgo de selección es el valor de las diferencias existentes entre los grupos de tratamiento y control, si a ninguno se le hubiera aplicado el programa.

Utilizando la misma notación algebraica de arriba, notemos que lo anterior se puede escribir como:

$$I = E(Y_1|D = 1) - E(Y_0|D = 1) \quad (\text{III})$$

Pero en la realidad, aproximamos $E(Y_0|D = 1)$ a través de un grupo de comparación que no ha participado en el programa, obteniendo el valor promedio de nuestro indicador a través de $E(Y_0|D = 0)$. El impacto que estimamos entonces, denotado por I' , es:

$$I' = E(Y_1|D = 1) - E(Y_0|D = 0) \quad (\text{IV})$$

Que es igual a:

$$I' = E(Y_1|D = 1) - E(Y_0|D = 1) + E(Y_0|D = 1) - E(Y_0|D = 0) \quad (V)$$

$$= I + E(Y_0|D = 1) - E(Y_0|D = 0) \quad (VI)$$

Así, el impacto estimado I' es igual al impacto verdadero, I , más el sesgo de selección; denotado en la ecuación de arriba por $E(Y_0|D = 1) - E(Y_0|D = 0)$.

El problema de hallar el contrafactual, para estimar el impacto, puede verse también como el problema de eliminar el sesgo de selección para aislar el efecto de un programa o política particular. Un ejemplo generalmente usado para ejemplificar el sesgo de selección lo constituyen los programas subvencionados de empleo (Angrist y Pischke, 2008). Debido a la focalización de estos programas hacia grupos vulnerables o población de riesgo, al comparar los salarios entre esta población y la población general vemos que los salarios de los primeros son menores, en promedio, aún después de recibir la capacitación; lo que se podría interpretar como falta de efectos positivos de los programas de capacitación para el empleo. Sin embargo, comparar a los recipientes de capacitación con el trabajador promedio es equivocado cuando se trata de evaluar los efectos de recibir la capacitación; precisamente porque existe un sesgo de selección al elegir a la población general para compararla con grupos vulnerables o población de riesgo. Si, en su lugar, utilizamos un grupo de comparación equiparable al que recibió el tratamiento -particularmente respecto a su condición de vulnerabilidad- entonces los efectos de la capacitación son generalmente positivos (Angrist y Pischke, *op. cit.*, 2008, pp. 12-13).

Desde una perspectiva analítica, existen dos estrategias para resolver este problema: los diseños de control experimental y los diseños cuasiexperimentales o no experimentales.

1.2.1. Los diseños experimentales en evaluación de impacto

Los diseños experimentales consisten en la *aleatorización* de una intervención entre la población participante; a fin de resolver el problema generado por el sesgo de selección en el cálculo del impacto. Estos métodos -conocidos genéricamente por el nombre de métodos experimentales- son la base de metodologías de evaluación de programas sociales de amplia difusión y uso (Banerjee y Duflo, 2008).

La investigación con diseños experimentales “comienza por un interés acerca de la identificación confiable de los efectos de un programa en presencia de múltiples y complejas formas de causalidad. Los experimentos hacen posible variar un factor a la vez, proveyendo estimadores “internamente” válidos del efecto causal” (*Ibidem*, p. 4; traducción propia).

Una primera etapa consiste en la identificación de la población sobre la que se quiere aplicar el programa o política; es decir, la población elegible para participar. Por ejemplo, si el objetivo es ofrecer becas a los alumnos de mejor desempeño, es necesario definir el umbral de calificaciones a partir del cual recibirían los beneficios.

Una vez identificada la población objetivo, el siguiente paso es seleccionar aleatoriamente una muestra sobre la cual llevar a cabo el experimento. Es importante verificar que esta muestra sea estadísticamente indistinguible, en las características observables y no observables, de la población de la que se extrajo².

La siguiente etapa es característica del diseño experimental, y consiste en seleccionar aleatoriamente -de entre la muestra seleccionada- un grupo de individuos que recibirán el programa; y otro que no lo recibirá. Es decir, seleccionar aleatoriamente a grupos de la población objetivo sobre los que se quiere aplicar un programa social; exponer al grupo de tratamiento al

² Para un análisis sobre el cálculo de tamaño de muestras, véase el Apéndice C en Siegel y Swanson (2004)

programa, y dejar fuera de éste al grupo de control. Aunque en la práctica es imposible que la submuestra que recibe el tratamiento sea idéntica a la que no lo recibe, es importante verificar que toda diferencia entre uno y otro grupo sea debida al azar.

Dado que antes de la aplicación del experimento no había diferencias estadísticamente significativas entre ambos grupos -y que la aplicación del programa fue aleatoria- se concluye que toda diferencia estadística observable en el promedio de una variable de interés en estos grupos es debida a la aplicación del programa o política (el así llamado efecto promedio del tratamiento, o *average treatment effect* como se le conoce en inglés).

El efecto promedio del tratamiento se refiere al promedio de los impactos individuales que la intervención tuvo sobre la población objetivo. En cambio, el efecto del tratamiento sobre los tratados (o *average treatment on the treated*) es el efecto sobre la población potencialmente expuesta al tratamiento, comparada con el contrafactual de que no hubiesen sido expuestos.

En Duflo, Glennerster y Kremer (2007) se distinguen cinco formas para generar una variación aleatoria, aunque cada intervención puede utilizar más de una:

Al inicio de un programa. Es el tipo de aleatorización con mayor recurrencia en la literatura. Se lleva a cabo simultáneamente con el nuevo programa. Selecciona una muestra representativa de la población y se le divide entre los grupos de tratamiento y control. En Schultz (2000), por ejemplo, se describe una intervención como ésta para el caso del programa PROGRESA. En el Capítulo 3, el lector puede encontrar la evaluación a un programa de lectura en verano que recurre a esta metodología.

Por lotería. Cuando la demanda por un programa o política particular excede los recursos de los que dispone el agente implementador, una manera justa de decidir quiénes reciben dichos recursos es a través de loterías. Esto permite además generar un diseño experimental de forma automática, ya que la probabilidad de pertenecer al grupo de los que reciben o no reciben el

programa es la misma; y esto garantiza que las características de los que están en uno u otro grupo sean estadísticamente indistinguibles (para muestras suficientemente grandes). La investigación de Beuermann *et. al.* (2015) es un ejemplo de este tipo de intervenciones.

Por desfaseamientos. Otra forma de introducir aleatoriedad entre quienes reciben o no un programa ocurre cuando, en principio, hay una población en la que todos son elegibles; pero no es posible implementar el programa para todos al mismo tiempo. En este caso, se puede introducir el programa paulatinamente, eligiendo al azar quiénes lo recibirán primero y, de esta manera, generar un primer grupo de tratamiento y de control. Aunque eventualmente todos recibirán el programa, las brechas en el tiempo entre quienes lo reciben primero y quienes lo reciben después permiten estimar si el programa generó impactos estadísticamente significativos entre uno y otro grupo. El requisito es que las diferencias en el tiempo entre los primeros y los segundos duren lo suficiente como para que los efectos del programa, de existir, se vean materializados. En Behrman, Parker y Todd (2011) se estiman los efectos de la ampliación del programa *PROGRESA/Oportunidades* en lo que constituye una muestra estratificada.

Por grados o grupos de comunidades. Si no es posible decidir al azar quiénes reciben un programa y quiénes no, otra forma es asignar la intervención a diferentes subgrupos dentro del grupo de los tratados. Esto arrojaría estimaciones creíbles de impacto, siempre y cuando los efectos del programa o política particular no contaminasen a los subgrupos que no hayan sido elegidos. Por ejemplo, si no fuera posible introducir un programa aleatoriamente a un grupo de escuelas, podría al menos buscarse que el programa se implemente en salones de clase diferentes al interior de cada una. En la medida que los salones que no reciben el programa no se vean beneficiados por éste, constituyen un buen grupo de control para compararse con grupos similares dentro de la misma escuela; arrojando un resultado confiable al aglutinar un número

suficientemente grande de observaciones. Kim (2006) es un ejemplo de este tipo de intervenciones.

Por incentivos diferenciados. En este caso el implementador expone -de forma aleatoria- a individuos elegibles para el programa a incentivos distintos al resto; generando así una mayor propensión a la participación que permite estudiar los efectos resultantes. Esta técnica es particularmente útil para estudiar programas en los que no pueden ponerse barreras para que la gente participe, pero sí alterarse los estímulos de algunos para incrementar su deseo de participar (por ejemplo, ofreciendo pequeños incentivos económicos que no alteren sus características, pero sí su deseo de entrar al programa); o facilitándoles más información que al resto sobre cómo inscribirse al programa, así como los posibles beneficios de su adscripción. En Banerjee *et. al.* (2010) puede encontrarse un ejemplo de este método de aleatorización.

Es importante mencionar que el evaluador enfrenta barreras y limitaciones en el ejercicio de su profesión; principalmente relacionadas con los criterios éticos con los que debe conducirse (INEE, 2015). Pero también de naturaleza legal, institucional y normativa; entre otras. Por ejemplo, en el caso del evaluador de desempeño, el Instituto Nacional para la Evaluación de la Educación (INEE) estableció un Código de Ética al que debían adscribirse sus evaluadores, y distinguía cinco principios rectores: 1) La mejora continua como propósito de la evaluación. 2) Equidad; revirtiendo la tendencia de reproducir la desigualdad. 3) Justicia; respetando en todo momento los derechos de las personas. 4) Diversidad; reconociendo, valorando y atendiendo las diferencias contextuales y culturales. 5) Participación; logrando consensos sociales que permitan ver la evaluación como una oportunidad para crecer y mejorar (INEE, *op. cit.*, pp. 4-5).

Adicionalmente, este instituto estima pertinente la promoción de la integridad, los valores éticos, la transparencia y la rendición de cuentas. Esto es particularmente relevante en el caso de la

implementación de los diseños experimentales en ciencias sociales, cuyo propósito siempre es servir de evidencia respecto a aquello que potencialmente tiene el mayor efecto sobre el bienestar presente y futuro de las personas.

Por ejemplo, cuando se trata de repartir una fracción del presupuesto educativo de una entidad federativa y se tiene un abanico de opciones viables, aprobadas por la ciudadanía, todas con efecto positivo sobre aquellos que recibirán el recurso -pero estas opciones se encuentran en conflicto y está en las facultades de la autoridad decidirse por alguna- lo ideal sería contar con evidencia robusta sobre sus distintos efectos; antes de tomar la decisión. Los comedores escolares, la distribución de dispositivos electrónicos con contenido pedagógico, la entrega de libros de texto, o la instalación de bibliotecas escolares son todos ejemplos de políticas públicas cuya gratuidad sería deseable para las escuelas; y puede haber más. El evaluador puede encontrar efectos positivos sobre el aprendizaje de cualquiera de estas políticas, y recomendar la implementación de todas ellas si existe el recurso suficiente para hacerlo. Pero si dicho recurso no fuera suficiente; o fuera suficiente -pero se sumaran otras propuestas igual o más deseables- el evaluador debe proveer evidencia robusta de los efectos de las distintas políticas para el tomador de decisión; tanto de las variables que le interesan a éste como de potenciales efectos no previstos que el tomador de decisión deba tener en cuenta. La evidencia robusta enriquece de esta manera la perspectiva valorativa del tomador de decisión, y se incorpora a su visión a la hora de tomar la o las decisiones que le competan.

El evaluador, por su parte, debe partir de un sustento teórico que justifique el experimento, e idealmente contar con la aprobación de un comité de ética para ejecutarlo. Comprometerse con la integridad en su actuar. Con la transparencia en la manera en que implementaron, procesaron y obtuvieron los resultados a partir de la puesta en práctica de las distintas intervenciones. Debe también comprometerse con el código de ética que corresponda al entorno institucional; tanto

del ejercicio de sus funciones como de la población involucrada en la intervención. Debe entregar sus resultados únicamente a las personas en cuya competencia esté recibir dicha información; abstenerse de influir en ámbitos que no le correspondan, y rendir cuentas del presupuesto recibido y los resultados obtenidos. También debe informar cualquier resultado no previsto del que haya tenido conocimiento.

Además de los diseños experimentales, existen otros diseños de investigación para estimar el valor del contrafactual, agrupados bajo el término de diseños cuasiexperimentales o no experimentales.

1.2.2. Los diseños cuasiexperimentales o no experimentales

En circunstancias donde no es posible llevar a cabo un diseño experimental antes de la intervención, existen técnicas estadísticas para estimar un contrafactual que sirva para comparar al grupo beneficiario del programa. Estas técnicas producen estimadores de impacto que son consistentes e insesgados bajo ciertos supuestos; y son robustos si dichos supuestos se cumplen. Las técnicas más comunes son brevemente descritas a continuación.

1.2.2.1. Regresión discontinua

El método de regresión discontinua se utiliza cuando las reglas de operación del programa imponen una restricción para participar en el mismo (por ejemplo, una edad mínima, un nivel de ingreso máximo, etc.) que segmenta a la población en elegibles y no elegibles a través de una regla. La discontinuidad así generada puede ser aprovechada para comparar a los individuos, en ambas poblaciones, que están en una vecindad cercana a dicha discontinuidad. Si tomamos a los que caen justo en el límite para ser elegibles y los comparamos con aquellos que cayeron justo en el límite para no serlo, puede considerarse casi como aleatoria la decisión de que participen en el programa, y tratarlos como si proviniesen de un diseño experimental. Buddelmeyer y Skoufias (2003) es un buen ejemplo de la aplicación de esta metodología, donde se contrastan

los resultados con los que son producto de una asignación aleatoria. El problema con este método radica en que los impactos estimados solo son válidos para las personas que se encuentran en dicha vecindad, pudiendo ser distintos al tratarse del resto de los individuos que participan en el programa.

Otra limitante radica en que existe una disyuntiva (*trade-off*) entre el tamaño de la muestra y la varianza del estimador. Entre más cerca esté el umbral considerado de la línea de discontinuidad, la estimación será más confiable³; sin embargo, esto implica reducir el número de observaciones que contiene dicho umbral, incrementando el error estándar del estimador. Por ello, el cálculo de impactos basados en este método requiere en general de grandes tamaños de muestra.

1.2.2.2. Diferencias en diferencias

El método de diferencias en diferencias se utiliza en casos donde se tiene información de los individuos que reciben la intervención antes y después de que esta ocurre, pero no se cuenta con un grupo de control producto de una aleatorización. En su lugar, este método utiliza información -recopilada antes y después de la implementación de un programa- para comparar los cambios sufridos en el grupo de tratamiento.

En lugar del grupo de control, se hace uso de un grupo de comparación que sea lo más parecido que se pueda al de tratamiento⁴, pero que no haya sido afectado por el programa. Lo que se hace es comparar los cambios experimentados por el grupo tratado (antes y después de la intervención) con los cambios que sufrió en ese mismo periodo el grupo de comparación. Se espera que la *diferencia de las diferencias* observadas en el tiempo por ambos grupos refleje el

³ En el límite, si se tomaran los valores que se encuentran justo en el punto de discontinuidad, el estimador tendría la misma confiabilidad que si se hubiera utilizado un diseño aleatorio para su estimación.

⁴ Se busca que este grupo sea parecido no solo en los indicadores relevantes antes de la intervención, sino que también haya sido expuesto a circunstancias similares a las que experimentó el grupo de tratamiento (excluyendo la intervención) durante la exposición al programa.

impacto de dicho programa. Puesto en forma algebraica, sea Y_1 un individuo en el grupo de tratamiento (igual que en el caso de un diseño experimental) y Y_0 un individuo elegido para formar parte del grupo de comparación. Sea t' el momento del levantamiento de información de la línea base, y t el levantamiento de información que se ha llevado a cabo tras la implementación del programa. El estimador de diferencias en diferencias del impacto del programa (DD) viene dado por:

$$DD = E(Y_{1t} - Y_{1t'} | D = 1) - E(Y_{0t} - Y_{0t'} | D = 0) \quad (\text{VII})$$

El supuesto necesario para que este estimador no esté sesgado es -como se dijo antes- que los cambios experimentados por los tratados y los no tratados (en ausencia del programa) hayan sido los mismos en el periodo estudiado. Este supuesto se conoce en la literatura como *supuesto de tendencias comunes* (J-PAL, 2010). Una forma de expresar esto es:

$$E(Y_{0t} - Y_{0t'} | D = 1) = E(Y_{0t} - Y_{0t'} | D = 0) \quad (\text{VIII})$$

Este supuesto, pese a ser más restrictivo que en el caso de los diseños experimentales, resulta más laxo que, por ejemplo, el que se requiere para comparar al grupo de tratamiento con otro grupo cualesquiera cuando no existe una línea basal⁵. Esto se debe a que el estimador de diferencias en diferencias elimina el sesgo generado por variables no observables; siempre y cuando las variaciones en el tiempo de estas variables no sean distintas entre ambos grupos. Duflo (2000) es un ejemplo clásico de la implementación de este método.

1.2.2.3. Método de apareamiento (Matching)

Consiste en utilizar técnicas estadísticas para construir un grupo de comparación similar al grupo que recibe el programa, a partir de sus características observables (Rosenbaum y Rubin, 1983).

⁵ Ya que, en ese caso, es necesario suponer que el grupo de tratamiento se hubiera comportado igual que el grupo de comparación en ausencia del programa, es decir, que $E(Y_0 | D = 1) = E(Y_0 | D = 0)$ o, dicho de otro modo, que no existe un sesgo de selección entre ambos grupos.

Existen diversas técnicas para aparear dichos grupos. La forma más directa es con un apareamiento exacto; es decir asignar, a cada individuo, otro que tenga exactamente las mismas características en todos los aspectos observables de la medición.

Debido a que la dificultad para encontrar individuos idénticos se incrementa conforme aumenta el número de características que se usan para hacer el apareamiento, otras técnicas han sido desarrolladas (Becker e Ichino, 2002; Angrist y Pischke, 2008).

Existen diversas opciones respecto a los algoritmos para llevar a cabo el apareamiento. El método de apareamiento más recurrentemente utilizado por la literatura es el de marcador de propensión (*propensity score*). El primer paso para aplicar esta metodología es estimar las probabilidades de participación para el grupo de tratamiento; basándose en sus características. Posteriormente, a cada uno de estos participantes se le asigna una probabilidad única de pertenecer al programa. La tarea entonces es encontrar un individuo (o combinación lineal de individuos) que no haya recibido el tratamiento, pero cuyo marcador de propensión sea parecido al de un individuo en el grupo tratado. Esto es lo que se conoce como apareamiento por marcador de propensión (*propensity score matching*).

En Rosenbaum y Rubin (1983, *op. cit.*) se prueba que, bajo algunos supuestos⁶, aparear a los individuos a través de su marcador de propensión es estadísticamente equivalente a hacerlo a través de todas sus características relevantes (en este caso, las características relevantes para participar en el programa o política de interés). De esta manera, el impacto del programa se calcula obteniendo el promedio de las diferencias observadas entre aquellos que participaron en el programa y sus “pares” en términos del marcador de propensión. Para ello necesitamos

⁶ Se requiere cumplir con los supuestos de *independencia condicional* (que todos los criterios para participar en el programa sean conocidos y observados) y *soporte común* (que existan individuos en el grupo de comparación con un rango de características similar a las del grupo de tratamiento). Estos requisitos son necesarios, aunque no exhaustivos.

información de ambos grupos después de la implementación del programa (de donde obtenemos el efecto).

El supuesto necesario para que el impacto así estimado sea un buen indicador del impacto del programa es que, bajo el marcador de propensión $P(X)$, el valor promedio del indicador entre el grupo de tratamiento y el de comparación sea el mismo en ausencia del programa. Puesto en términos algebraicos:

$$E(Y_0|P(X), D = 1) = E(Y_0|P(X), D = 0) \quad (\text{IX})$$

Podemos pensar en este supuesto como la no existencia de sesgo de selección *condicional* al marcador de propensión⁷. Díaz y Handa (2006) utiliza esta metodología, evaluando su precisión de cara a un diseño experimental, para el caso del programa PROGRESA.

Un supuesto menos restrictivo -cuando se dispone de información tanto antes como después de la implementación- es el de apareamiento de diferencias en diferencias (*difference-in-difference matching*):

$$E(Y_{0t} - Y_{0t'}|P(X), D = 1) = E(Y_{0t} - Y_{0t'}|P(X), D = 0) \quad (\text{X})$$

Aquí el supuesto necesario es que, condicional a su marcador de propensión, las diferencias en el tiempo entre los grupos de tratamiento y de comparación en ausencia del programa sean las mismas. Se espera que el método de diferencias en diferencias ayude a mitigar la incidencia de heterogeneidad no observada entre el grupo tratado y el que sirve de comparación. En Behrman *et. al.* (2012) se implementa una aplicación de este método, siendo un buen referente para su uso.

El método de apareamiento a través de cálipers (*caliper matching*) recurre a intervalos definidos por encima y debajo del valor de cada característica; ampliando el espectro de individuos

⁷ Reordenando, se tiene $E(Y_0|P(X), D = 1) - E(Y_0|P(X), D = 0) = 0$.

elegibles en el grupo de control. En Sianesi (2004) se aplica este método en el contexto de programas para el empleo.

El aparejamiento por clústers (*cluster matching*), a su vez, utiliza vecindades cercanas a las de las características de los individuos en el grupo de tratamiento para elegir al grupo de control; mientras que el aparejamiento por referencias (*benchmark matching*), lo hace a través de una medición multivariada (Shadish, Cook y Campbell; 2002 p. 119).

En el Capítulo 2, el lector encontrará una revisión más detallada del método de aparejamiento, pues se recurre a esta metodología para estimar el efecto de una asignación de alumnos a los turnos matutino y vespertino.

1.2.2.4. Variables instrumentales

Cuando existen variables no observables -correlacionadas con el hecho de participar (o no) en el programa cuyo impacto queremos estudiar- los resultados que encontremos a través de cualquiera de los métodos antes descritos tendrán presente un sesgo, incluso cuando la asignación del tratamiento sea aleatoria. Una forma de eliminar los efectos de este sesgo es recurrir a un instrumento; es decir, a una variable (Z) que esté altamente correlacionada con el hecho de participar en el programa (T); pero que no esté correlacionada con las variables latentes. Formalmente, se busca una Z tal que:

$$Cov(Z, T) \neq 0 \quad (XI)$$

$$Cov(Z, e) = 0 \quad (XII)$$

Donde e denota los no observables⁸. En general, cuando la decisión de qué individuos participan en el programa es aleatoria, se utiliza esta regla de asignación como variable instrumental; y,

⁸ e sería el término de error en la regresión $I_i = \alpha + \beta X_i + bT_i + e_i$, donde I son los impactos del programa, X las variables observables y T la participación en el programa.

cuando existe alguna otra regla de asignación, dicha regla se convierte usualmente en la Z deseada.

Una vez que el instrumento ha sido hallado, se procede a realizar la estimación de los impactos utilizando mínimos cuadrados en dos etapas; introduciendo en la primera etapa, como variable explicativa, el instrumento elegido; y T como la variable dependiente. En la segunda etapa, se utilizan los valores predichos para T como las variables explicativas, y la variable que mide el impacto como la variable a explicar; siendo el coeficiente asociado a dicha variable el que mide el impacto del programa. Como ejemplo, el trabajo de Angrist y Lavy (1997) estima, mediante este método, los impactos de una regla de asignación del número de estudiantes por salón de clase.

1.2.2.5. Diseño *pipeline* o de línea de espera

Este diseño aplica cuando existen individuos que podrían pertenecer al grupo de tratamiento (son elegibles), pero que por factores exógenos al programa se encuentran en una lista de espera. El método asume que no existe otra diferencia entre los tratados y los individuos en espera más allá del tiempo en el que mostraron interés por participar en el programa. Así, si esta diferencia en tiempo es “pequeña”, los grupos son comparables entre sí y pueden ser tratados como parte de un diseño aleatorio. El trabajo de Ángeles *et. al.* (2011), aplicado al caso del Programa de Estancias Infantiles, en México, utiliza esta metodología.

Además de algunos ejemplos antes citados, en el contexto de nuestro país existe una variada gama de investigaciones que han aplicado la MEI. El objetivo del siguiente capítulo será el de presentar la metodología seguida para recopilar dichas investigaciones; hasta el año 2013.

1.3. Metodología para la revisión de literatura

El criterio rector para seleccionar los estudios que componen la revisión fue la utilización de la MEI aquí descrita. Se incluyeron evaluaciones enfocadas en México (para maximizar su validez

externa hacia nuestro contexto⁹). Se dejó fuera aquellos estudios que -aún utilizando evaluación de impacto- no estuvieran centrados en un programa, política o intervención en nuestro país. Cabe aclarar que la metodología priorizó los documentos más citados por la literatura y metodológicamente más robustos.

La estrategia utilizada fue la técnica de saturación (Bryman 2001; p. 18). Comenzó con una revisión de las bases de datos donde estas investigaciones se presentan; normalmente en revistas y publicaciones de economía y educación. Se utilizaron los buscadores *Google Scholar*, *EconLit*®, *Educational Resources Information Center (ERIC)*, *JStor*, *Science Direct* y el *National Bureau of Economic Research (NBER)*; utilizando las palabras clave México e, intercambiablemente, evaluación de impacto (*impact evaluation*) o impacto. Primero se introdujeron como palabras clave los tres términos (evaluación + impacto + México).

Después se limitó la búsqueda únicamente a impacto y México y finalmente se introdujo el término conjunto evaluación de impacto (*impact evaluation*) y México; con el objetivo de minimizar el número de resultados arrojados. La revisión se llevó a cabo entre septiembre y noviembre de 2013. Comenzando por el número mínimo de resultados obtenidos por buscador (ordenados por relevancia), se revisaron cada una de las entradas correspondientes a un máximo de 10 páginas de resultados. La Tabla 1.1 presenta el número de documentos encontrados en esta primera etapa.

De cada entrada, correspondiente a un documento de investigación o libro, se revisó si correspondía a una intervención en México, y si la metodología sugerida en la introducción correspondía a evaluación de impacto, fuese de forma explícita o como consecuencia de las características de diseño del que se partía. Esto generó una primera lista de documentos que,

⁹ Véase la crítica hecha en Pritchett, Samji, y Hammer (2012) respecto a la validez externa de las evaluaciones de impacto.

tras omitir duplicidades (aún aquellas publicadas en revistas diferentes, o por distintos autores, pero con las mismas variables de interés)¹⁰, ascendió a 34.

Tabla 1.1. Resultados de documentos en buscadores

Buscador	Número máximo de resultados	Filtrados por impacto	Filtrados por “evaluación de impacto” (<i>impact evaluation</i>)	Documentos finales incluidos en la primera revisión
Google Scholar	1,390,000	1,320,000	8460	100
ERIC	121	--	7	7
EconLit	2	235	1	100
NBER	3,240	491 ^a	9	9
JSTOR	45,620	--	326	100
Science Direct	56,024	8,804 ^b	506	100
ProQuest	112,581	298^c	821	200

Notas: a) Filtrado por impacto y evaluación; y México, b) Se añadió la palabra impacto como palabra clave, c) Se solicitó que impacto estuviera ya sea en el resumen o como palabra clave. Fuente: cálculos propios.

A continuación, la bibliografía de cada uno de los documentos encontrados fue revisada, a fin de descubrir documentos previamente publicados. Esto permitió sumar 8 documentos más. Estos documentos se depuraron revisando por completo la sección de metodología, dejando 42 documentos que efectivamente llevaban a cabo evaluación de impacto.

Adicionalmente, se revisaron las publicaciones de la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES), así como tesis y tesinas a través de *ProQuest* (cuya búsqueda se reporta también en la Tabla 1.1); por presentar investigaciones noveles que podrían cubrir los requisitos de esta revisión. Por último, se añadieron los documentos reconocidos como evaluaciones de impacto de programas sociales por el Consejo Nacional para

¹⁰ El caso de investigaciones de distintos autores, pero centradas en las mismas variables fue particularmente recurrente en el caso de PROGRESA, ya que este programa ha servido para la exploración de distintos modelos y aproximaciones teóricas (por ejemplo, para hacer pronósticos). En esta revisión solo se incluyeron los documentos que, o bien se enfocan en estimar diferentes variables (todas de impacto), o se enfocan en las mismas variables a través de métodos diferentes, pero siempre pertenecientes a la MEI.

la Evaluación de la Política de Desarrollo Social (Coneval)¹¹. Tras una depuración de estos documentos (que ascendió a la cifra de 16), se contó con un total de 57 documentos de investigación; a los que se sumaron 2 documentos producto de una revisión reciente por parte del Banco Mundial¹². El propósito de la siguiente sección es presentar las intervenciones estudiadas por estos documentos -así como las variables que evalúan y los impactos que estiman- para clasificar y ponderar aportaciones de la MEI al estudio de políticas públicas en México.

1.4. Evaluaciones de impacto para el caso mexicano

Las fuentes principales de aplicación de la MEI en México, -salvo algunas excepciones¹³- fueron las evaluaciones al Programa de Educación, Salud y Alimentación (PROGRESA), posteriormente *Oportunidades*^{14,15} y luego *Prospera*¹⁶. La investigación, no obstante, siguió un patrón diferente; ya que en Pritchett, Samji y Hammer (2012, *op. cit.*) se distinguen tres etapas que ha seguido la MEI en su desarrollo: la etapa de diseños cuasiexperimentales como fuente principal de evidencia sobre el impacto de programas; la etapa de proliferación de estudios a

¹¹ Disponibles en la página http://www.coneval.gob.mx/Evaluacion/Paginas/Evaluaciones_Programas/Evaluacion_Impacto/Evaluacion_Impacto.aspx. Consultada el 17 de septiembre de 2013.

¹² Disponible en <http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluationHome.aspx?sD=E>

¹³ En particular, los estudios sobre el impacto de los programas de educación compensatoria y de capacitación para el empleo que pueden encontrarse en Revenga, Riboud y Tan (1994) y Acevedo (1999).

¹⁴ Un logro adicional de PROGRESA fue el de sobrevivir a los cambios de administración del año 2000 en México; que fueron producto de una transición democrática del partido al frente del Poder Ejecutivo por más de 70 años (Partido Revolucionario Institucional) hacia la oposición (Partido Acción Nacional). Para una descripción de las diferencias y similitudes entre PROGRESA y Oportunidades, véase Hevia de la Jara (2009).

¹⁵ PROGRESA fue el detonante de una veta de literatura durante la década del 2000; veta que trascendió nuestro país y creó una discusión en torno a la efectividad de los así llamados programas de transferencias condicionadas (*conditional cash transfer programs*) que se han implementado a lo largo de todo el mundo.

¹⁶ De acuerdo con la página oficial del gobierno mexicano, *Prospera Programa de Inclusión Social* constituye un fortalecimiento del Programa *Oportunidades*, “que articula y coordina la oferta institucional de programas y acciones de política social, incluyendo aquellas relacionadas con el fomento productivo, generación de ingresos, bienestar económico, inclusión financiera y laboral, educación, alimentación y salud; dirigida a la población que se encuentre en situación de pobreza, bajo esquemas de corresponsabilidad que les permitan a las familias mejorar sus condiciones de vida y aseguren el disfrute de sus derechos sociales y el acceso al desarrollo social con igualdad de oportunidades.” (Disponible en <http://www.gob.mx/prospera>, consultado el 13 de febrero de 2018).

partir de diseños experimentales, y una tercera etapa de investigación que plantea la necesidad de generar aprendizajes locales y políticas que tengan una constante evolución.

1.4.1. Evaluaciones de impacto de *PROGRESA/Oportunidades/Prospera*

PROGRESA se diseñó como un experimento aleatorio, asignando los hogares de 320 comunidades de México al grupo de tratamiento y 186 comunidades al grupo de control. El programa levantó encuestas cada seis meses para dar seguimiento a los beneficiarios.

La primera evaluación del programa se asignó al *International Food Policy Research Institute* (IFPRI). En Skoufias (2001) se detallan los resultados de dicha evaluación, que encontró efectos positivos sobre asistencia escolar (evitando la deserción de los niños), salud, nutrición y rendimiento físico de niños y adultos. A partir de allí, se han llevado a cabo múltiples evaluaciones académicas para estimar los efectos de PROGRESA sobre diferentes medidas de bienestar.

La Tabla 1.2 resume las evaluaciones de PROGRESA que corresponden a la MEI. En general, los resultados dan cuenta de la efectividad del programa. Estos pueden agruparse -de forma analítica- en tres grupos, acordes a su nombre: educación, salud y alimentación. No obstante, una característica distintiva del programa reside en su carácter multidimensional; que busca combatir los principales problemas que aquejan a las personas en situación vulnerable; proponiéndose generar un círculo virtuoso que refuerce su bienestar y les permita generar capacidades para salir de la pobreza.

En el caso de educación, el programa involucraba un monitoreo de la asistencia escolar; condicionando la permanencia del programa a una asistencia de 85% por parte de los alumnos beneficiarios, y el hecho de que no reprobasen (PROGRESA, 1997). Indirectamente, también tenía como objetivo que éstos no desertaran por factores asociados a carencias económicas; a partir de la ayuda que otorga.

10 de los estudios coinciden en que el programa incrementa la asistencia escolar y el número de grados cursados; y disminuye la reprobación. La magnitud de impacto de PROGRESA se estima en alrededor de un año más de escolaridad esperada para los niños que beneficia (Behrman, Parker y Todd, 2011); con una mejora en indicadores intermedios de asistencia escolar y desempeño académico. Si bien los resultados en términos de desempeño de los alumnos son limitados, las evaluaciones muestran que la mayor permanencia de los menores en la escuela tiene un “efecto de derrama” sobre otros indicadores de bienestar. Por ejemplo, alejándolos de actividades nocivas en perjuicio de ellos, sus familias o la sociedad (Cabrera-Hernández, 2015). Para el caso de salud, los resultados de cuatro estudios son consistentes en que las familias que participan en el programa presentan menos enfermedades (Gertler, 2004), niños con menor desnutrición (Rivera *et. al.*, 2004), crecimiento más sano y mayor talla para la edad, principalmente en los primeros años de vida (Behrman y Hoddinott, 2000; Dominguez *et. al.*, 2008). Los estudios señalan que los mayores beneficios de PROGRESA los reciben los niños de menor edad; y se potencializan cuando la madre es beneficiaria desde su embarazo.

En el caso de alimentación, cuatro estudios coinciden en que la nutrición de las familias mejora; no solo con más variedad de alimentos (Davis *et. al.*, 2002), sino también con una dieta más variada y rica en nutrientes (Hoddinott y Skoufias, 2004). La capacitación que PROGRESA y sus variantes ha otorgado busca combatir los problemas que afectan a la población más vulnerable, como son: la desnutrición, la obesidad y la diabetes.

La mayoría de los estudios en este rubro utilizaron las bases de datos de la Secretaría de Desarrollo Social, producto de los levantamientos de las encuestas de seguimiento a hogares del programa. Una excepción es Rivera *et. al.* (2004); que al requerir información sobre micronutrientes llevó a cabo un levantamiento propio sobre una submuestra de los hogares tratamiento y control del programa.

Respecto a los autores de las evaluaciones presentadas en la Tabla 1.2, la mayoría pertenece al ámbito de la academia; seguidos de organizaciones no gubernamentales o intergubernamentales como IFPRI¹⁷, el Banco Interamericano de Desarrollo, la Organización de las Naciones Unidas para la Alimentación y la Agricultura, o el Banco Mundial. El tercer grupo más numeroso de autores corresponde a los pertenecientes al sector gubernamental; principalmente a través del Instituto Nacional de Salud Pública, cuyos investigadores tienen un perfil más académico.

Los resultados encontrados en estos estudios han servido de evidencia contribuyendo a que *PROGRESA/Oportunidades/Prospera* permanezca y haya incrementado su cobertura por más de quince años. Adicionalmente, la evidencia de los efectos del programa ha tenido incidencia sobre la operación y los recursos de éste, contribuyendo a justificar la ampliación de los beneficios y beneficiarios (Secretaría de Desarrollo Social, 2008). El éxito de estas investigaciones para evidenciar los efectos de *PROGRESA/Oportunidades/Prospera* motivó a que otros programas del gobierno federal en México comenzaran a enriquecer sus estudios con evaluaciones de impacto para medir sus efectos. El rubro que abarca el volumen más importante de estudios bajo la MEI son, en general, las evaluaciones a programas del gobierno federal.

¹⁷ Que concentra la mayor proporción de evaluaciones entre las ONG's.

Tabla 1.2. Evaluaciones de impacto del Programa *PROGRESA/Oportunidades/Prospera*

Autor y año	Método	Impactos encontrados del programa	Magnitud
Behrman y Hoddinott (2000)	Diseño experimental aleatorio, controlado por efectos fijos.	Talla en niños menores de 3 años (+), retraso en el crecimiento (-).	Un centímetro más de estatura por año; para niños entre 12 y 36 meses.
Handa, Huerta, Pérez y Straffon (2000)	Diferencias en diferencias.	Pobreza (-).	4% menos personas por debajo de la línea de pobreza (fijada en el percentil 25 de consumo <i>ex ante</i>).
Schultz (2000)	Diferencias en diferencias.	Asistencia escolar de niños (+) y niñas (+).	0.72 años más de educación para las niñas, y 0.64 años para los niños.
Skoufias, Parker, Behrman y Pessino (2001)	Diferencias en diferencias.	Trabajo infantil de niños (-), deserción escolar de niños y niñas (-).	Para grupos de edad específicos, se encuentran reducciones, en tres rondas consecutivas, de hasta -22% en la probabilidad de los niños de trabajar, y de hasta -25% en el trabajo doméstico de las niñas. Así como un incremento en la asistencia escolar de las niñas de hasta 11% y de los niños de hasta 6%.
Ruiz-Arranz, Davis, Stampini, Winters y Handa (2002)	Aleatorización combinada con variables instrumentales	Consumo de alimentos (+) e ingesta de calorías (+) ¹⁸ .	30 centavos más de gasto en alimentos por cada peso recibido de PROGRESA. 2.8 calorías diarias más por cada peso transferido por el programa.
Buddelmeyer y Skoufias (2003)	Regresión discontinua.	Asistencia escolar de niños y niñas (+).	Incremento de 4.8% en la asistencia escolar de niños, y 9.9% en la de niñas.
Dubois, de Janvry y Sadoulet (2003)	Diseño experimental aleatorio con estimación por un modelo de transición.	Deserción (-), desempeño (+).	2 puntos porcentuales más de probabilidad de completar el grado escolar que cursan (primaria). 6.1 puntos porcentuales más en desempeño (primaria).

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

¹⁸ Este documento también evalúa el impacto de PROCAMPO. Los hallazgos encontrados se detallan en la Tabla 1.3.

Tabla 1.2. Evaluaciones de impacto del Programa *PROGRESA/Oportunidades/Prospera* (Cont.)

Autor y año	Método	Impactos encontrados del programa	Magnitud
Gertler (2004)	Diseño aleatorio.	Enfermedades en niños de 0 a 5 años (-), enfermedades en adultos (-), días incapacitados (adultos -).	12% menos incidencia de enfermedades en niños de 0 a 5 años; 19% menos dificultades por enfermedad en adultos; 17% menos días incapacitados (mayores de 50 años).
Hoddinott y Skoufias (2004)	Diferencias en diferencias con un modelo de efectos fijos	Consumo calórico, proveniente principalmente de productos animales y vegetales (+).	El hogar promedio ingiere 6.4% más calorías, y de mejor calidad.
Rivera, Sotres-Alvarez, Habicht, Shamah y Villalpando (2004)	Diseño experimental aleatorio.	Anemia en niños de hogares rurales (-), talla de niños (+).	0.37 g/dL más de hemoglobina y 1.1 cms. más de talla en menores de año y medio o menos, tras un año de intervención.
Behrman, Sengupta y Todd (2005)	Diseño experimental con impacto estimado por matrices de transición.	Progresión escolar (+) y reprobación (-).	Progresión: las matrices de transición muestran impactos que fluctúan entre el 7 y el 11% anual entre los seis y los diez años. Repetición: las matrices de transición muestran impactos que fluctúan entre el -7 y el -8% anual entre los seis y los diez años.
Díaz y Handa (2006)	Aparejamiento por marcador de propensión.	Gasto en alimentos (+), asistencia escolar (+).	El diseño experimental arroja un gasto en alimentos por persona mayor en 35 pesos; y un 6.6% más de inscripción escolar entre los 13 y los 16 años. Las muestras pareadas son estadísticamente indistinguibles en el rubro de asistencia escolar, pero presentan un sesgo a la baja respecto al gasto en alimentos, y una sobreestimación del impacto de PROGRESA sobre trabajo infantil.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.2. Evaluaciones de impacto del Programa *PROGRESA/Oportunidades/Prospera* (Cont.)

Autor y año	Método	Impactos encontrados del programa	Magnitud
Parker, Todd, Wolpin (2006)	Aparejamiento, variables instrumentales.	Escolaridad niños y niñas (+).	6% más de asistencia escolar en el corto plazo. Un pronóstico de medio año más de escolaridad promedio en el largo plazo.
De Brauw y Hoddinott (2008)	Aparejamiento.	Asistencia escolar (+ con condicionamiento efectivo).	9% más de asistencia escolar.
Leroy, García-Guerra, Domínguez, Rivera y Neufeld (2008)	Aparejamiento.	Talla de niños menores de 6 meses en localidades urbanas (+).	1.5 cms de talla y 0.76 kgs de peso.
Bobonis y Finan (2009)	VARIABLES instrumentales.	Inscripción (+) y progresión (+) escolar.	0.49 puntos porcentuales de incremento en la inscripción escolar para vecinos de alumnos en PROGRESA (efecto de pares).
Lalive y Cattaneo (2009)	Diseño experimental y variables instrumentales.	Asistencia escolar (+)	Incremento en asistencia escolar de 5.8 puntos porcentuales para alumnos elegibles. 2.1 puntos porcentuales para alumnos no elegibles, en localidades con PROGRESA.
Attanasio, Meghir y Santiago (2011)	Diseño experimental aleatorio; modelo de equilibrio general.	Inscripción en niños (+) trabajo infantil (-).	El impacto promedio sobre inscripción se estima en alrededor de 5 puntos porcentuales. La disminución de trabajo infantil en 3.3%.
Behrman, Parker y Todd (2011)	Diferencias en diferencias.	Impactos crecientes sobre escolaridad (niños y niñas +) y trabajo infantil (niños -).	<i>Escolaridad.</i> 0.8 años más de escolaridad en jovencitas entre 15 y 18 años y entre 0.5 y 1.3 años más de escolaridad en varones entre 15 y 21 años. <i>Trabajo infantil.</i> 9% más probabilidad de trabajar de las mujeres entre 19 y 21 años. 19% menor probabilidad de trabajar en los varones entre 15 y 16 años.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.2. Evaluaciones de impacto del Programa *PROGRESA/Oportunidades/Prospera* (Cont.)

Autor y año	Método	Impactos encontrados del programa	Magnitud
Behrman, Gallardo-García, Parker, Todd y Vélez-Grajales (2012)	Diferencias en diferencias y apareamiento por marcador de propensión.	Inscripción a la escuela (+), asistencia escolar (+), tiempo dedicado a hacer tareas (+) y empleo infantil (- para niños). Para áreas urbanas.	4% más de escolaridad (para edades entre 12 y 14 años); 2% más de inscripción a la escuela para niñas y niños entre 8 y 11 años. Entre 14 y 21% más de horas de estudio a la semana para niños entre 12 y 14 años; y 12-14 puntos porcentuales menos de trabajo en niños entre 12 y 14 años.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

1.4.2. Evaluaciones de impacto a otros programas federales

La Tabla 1.3 presenta las investigaciones sobre programas federales relacionados con políticas de educación¹⁹; mientras que la Tabla 1.4 presenta el resto de los programas evaluados al menos en una ocasión a partir de la MEI. El motivo detrás de esta división es que, dejando de lado *PROGRESA/Oportunidades/Prospera*, la evaluación de impacto de programas educativos abarca más de la mitad del total de evaluaciones halladas a programas.

Siete de las 11 publicaciones estudiadas por la MEI, en el caso de educación, van dirigidas a estudiar intervenciones que benefician a los alumnos de educación básica y media superior (que son el grueso del sistema educativo nacional). Sus resultados enfatizan el modesto, pero estadísticamente significativo, impacto que tienen programas focalizados en poblaciones vulnerables; como es el caso de Acciones Compensatorias del Consejo Nacional de Fomento Educativo (CONAFE) y el Programa para Abatir el Rezago Educativo (PARE).

¹⁹ No necesariamente políticas educativas.

Tabla 1.3. Evaluaciones de impacto a programas federales en educación en México

Autor y año	Programa evaluado	Método	Impacto	Magnitud
Revenga, Riboud y Tan (1994)	Programa de Becas de Capacitación para Desempleados (PROBECAT).	Aparejamiento por marcador de propensión.	Tiempo desempleado (- para individuos con experiencia laboral), ingresos mensuales para varones (+), horas trabajadas por semana (+). Desempeño académico (+).	0.8 meses menos de desempleo (hombres mayores de 25 años). 1.1 meses menos de desempleo (Mujeres con experiencia laboral). Ingresos mensuales 17.7% mayores para varones. Más horas trabajadas por semana: 8.4 (hombres) y 1.2 (mujeres).
López-Acevedo (1999)	Programa para Abatir el Rezago Educativo (PARE).	Diseño experimental aleatorio.		Incremento en desempeño entre 19 y 38% para estudiantes en el medio rural. Este incremento puede alcanzar hasta 90% para estudiantes indígenas.
Calderón-Madrid y Trejo (2001)	Programa de Becas de Capacitación para Desempleados (PROBECAT).	Aparejamiento por marcador de propensión.	Reducción en el tiempo para encontrar empleo (- para la modalidad mixta en hombres; diferencias por regiones del país en la modalidad escolarizada, así como para mujeres).	Los resultados varían entre 0.2 y 4 puntos porcentuales dependiendo de las instituciones, el género de los participantes y la zona geográfica donde se ubican.
Canton y Blom (2004)	Becas de la Sociedad de Fomento a la Educación Superior (SOFES).	Regresión discontinua.	Inscripción a la preparatoria (+), a la universidad (+), desempeño académico (+), probabilidad de trabajar mientras se estudia en licenciatura (-) y en maestría (+).	24% más de probabilidad de entrar a la universidad (efecto que podría estar sobreestimado si la beca está correlacionada con aptitudes académicas). Mayor diversidad socioeconómica en la población de estudiantes: familias de estudiantes son en promedio, 48% más ricas que familias de estudiantes con beca <i>SOFES</i> .

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.3. Evaluaciones de impacto a programas federales en educación en México (Cont.)

Autor y año	Programa evaluado	Método	Impacto	Magnitud
			Elección de la profesión (+).	<p><i>Con base de datos SOFES:</i> 3% en desempeño para universitarios (sin controles; límite inferior). <i>Con encuesta:</i> 2% más de desempeño para universitarios. Caída en reprobación de -0.3%. -10% en la probabilidad de trabajar de los estudiantes de licenciatura; entre 17 y 30% más de trabajar para los estudiantes de maestría. Los estudiantes tienden a elegir menos carreras de ingeniería y más economía y gestión; a una razón de 1% más por año Estudiar en una preparatoria privada incrementa la inscripción a carreras de ingeniería y ciencias naturales en 5%.</p>
Hunter, Guernsey de Zapien, Papenfuss, Fernández, Meister y Giuliano (2004)	Programa Promotora.	Diseño experimental aleatorio.	Participación en programa de prevención y educación para la salud (+).	35% más visitas sucesivas a los exámenes preventivos anuales de salud.
Shapiro y Moreno-Treviño (2004)	Acciones Compensatorias para Abatir el Rezago en Educación Inicial y Básica (CONAFE).	Apareamiento o por marcador de propensión.	Desempeño en matemáticas en primaria (+) y en español y matemáticas en secundaria (+) Tasa de reprobación y deserción en primaria (-) Educación indígena (+).	Incremento en 4.8 a 5.6 puntos en matemáticas por año en primaria, y una disminución en la brecha entre escuelas CONAFE y no CONAFE de entre 9 y 30% por año, dependiendo la especificación econométrica. CONAFE incrementó los puntajes de español entre 0.8 y 3.4 puntos y los de matemáticas entre 0.02 y 1.4 puntos por año en secundaria.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.3. Evaluaciones de impacto a programas federales en educación en México (Cont.)

Autor y año	Programa evaluado	Método	Impacto	Magnitud
Murnane, Willet y Cardenas, (2006)	Programa Escuelas de Calidad (PEC).	Diferencias en diferencias.	Deserción (-).	Esto representa una disminución en la brecha de 24% en matemáticas y 38% en español por año. Disminución en la brecha de repitencia de 6% por año, y de reprobación de 0.3% por año. Tras 3 años en el PEC, las escuelas redujeron su tasa de deserción en casi 6%.
Shapiro y Skoufias (2006)	PEC.	Aparejamiento por marcador de propensión y diferencias en diferencias.	Deserción (-), reprobación (-) y repitencia (-).	-0.24 puntos porcentuales en deserción, -0.24 puntos porcentuales en reprobación y -0.31 puntos porcentuales en repitencia.
Santibáñez, Martínez, Datar, McEwan, Setodji y Basurto-Dávila, (2007)	Carrera Magisterial.	Regresión discontinua.	Puntajes alumnos de secundaria en las pruebas del programa Carrera Magisterial (+); tiempo de preparación de los profesores (+).	Ganancia de 3 a 15% de una desviación estándar, en los puntajes de los exámenes de alumnos de secundaria. Solo para los grupos con profesores más proclives a recibir los beneficios del programa.
Ángeles, Gadsden, Galiani, Gertler, Herrera, Kariger y Seira (2011)	Programa de Estancias Infantiles (PEI).	Método <i>pipeline</i> .	Probabilidad de la madre de encontrar un empleo (+); horas trabajadas de la madre (+). Desarrollo infantil (+).	Aumento en probabilidad de 18% de las titulares beneficiarias de encontrar un empleo. 8% más de permanencia laboral de la madre. Menores de 30 meses: 30% más de prevalencia de enfermedades.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.3. Evaluaciones de impacto a programas federales en educación en México (Cont.)

Autor y año	Programa evaluado	Método	Impacto	Magnitud
			Diversidad de la dieta de los menores (+). ²⁰	Mayores de 30 meses: 0.3 desviaciones estándar más de desarrollo infantil (impacto pequeño-mediano). 17.4% de reducción de prevalencia de enfermedades después de 6 meses en el programa. Hijos de madres que ya trabajaban: 5% más de diversidad en los grupos de alimentos que consumen.
Gertler y Patrinos (2012)	Programa “Apoyos a la Gestión Escolar”.	Diseño experimental aleatorio.	Reprobación (-), repitencia (-).	Reducción de la reprobación en 7.4% y de la repitencia en 5.5% entre primer y tercer grado en el nivel de primaria.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Otros programas, como el Programa Escuelas de Calidad (PEC) y Apoyos a la Gestión Escolar (AGE), parecen tener efectos heterogéneos sobre las poblaciones tratadas; siendo más efectivos en algunos contextos específicos que en otros. PEC, en particular, ha demostrado tener incidencia sobre poblaciones en menor desventaja socioeconómica -de entre los beneficiarios- mientras que las intervenciones de Conafe tienen mayor incidencia sobre los más pobres. También obsérvese que, en estricto sentido, ni las investigaciones sobre el Programa de Becas de Capacitación para Desempleados (PROBECAT), ni la investigación sobre estancias infantiles incluyen como indicador del impacto una variable que mida aprendizajes.

En general, estas investigaciones recurren a levantamientos de datos previamente realizados, o a bases de datos nacionales en los casos de los diseños cuasiexperimentales. Una excepción lo constituye Ángeles *et. al.* (2011), al recurrir a un levantamiento propio a madres de familia para estimar el efecto del Programa de Estancias Infantiles (*ibid.* P. 7).

²⁰ Los autores destacan el hecho de que estos resultados se den en el contexto de la crisis internacional de 2009-2011. Calculan que los impactos serían aún mayores en un contexto internacional de crecimiento económico.

Por su parte, en la Tabla 1.4 se puede apreciar la variedad de programas federales estudiados fuera del ámbito educativo; con un incremento, a partir de 2007, en la proporción de documentos que usan diseños experimentales. En cuanto a los autores de estas investigaciones, aquellos adscritos a la academia encabezan la lista; seguidos de los pertenecientes a organizaciones intergubernamentales como el Banco Mundial, y gubernamentales como el Instituto Nacional de Salud Pública.

Respecto a la exploración de abordajes metodológicos, están más condicionados por el tipo de información que se posee para llevar a cabo la evaluación, que por el método para la estimación de los impactos. Algunas excepciones proponen nuevos abordajes de evaluación surgidos a partir de PROGRESA (por ejemplo, Behrman, Sengupta y Todd, 2005; Behrman, Parker y Todd, 2011 o Attanasio, Meghir y Santiago, 2008 y 2011).

Es interesante observar que la mitad de las evaluaciones a programas federales no educativos utilizan levantamientos de información propios para sus estimaciones. Esta es una proporción mayor que la de los programas educativos, al igual que la de las investigaciones en torno a PROGRESA, cuyos levantamientos de información están a cargo de Sedesol. La mayor proporción se debe al financiamiento recibido para las investigaciones ya que, normalmente, en estos programas la evaluación se hizo de forma obligatoria por la autoridad; y el tipo de evaluación era establecido en los lineamientos a través de un contrato donde se estipulaba el levantamiento de información (Coneval, 2010). En la referida Ley se establece la necesidad de una evaluación periódica (Art. 72 y 78), para “emitir las sugerencias y recomendaciones que considere pertinentes al Ejecutivo Federal y hacerlas del conocimiento público” (Art. 80).

Tabla 1.4. Evaluaciones de impacto de otros programas federales en México

Autor y año	Programa evaluado	Método	Impacto	Magnitud
Ruiz-Arranz, Davis, Stampini, Winters y Handa (<i>op. cit.</i>)	PROCAM-PO.	Variables instrumentales.	Consumo de alimentos (+) e ingesta de calorías (+) Consumo de carne y vegetales (+) Diversidad de alimentos para el consumo (+ Cuando, adicional a PROCAMPO, un hogar recibe PROGRESA).	Propensión marginal de 30 centavos más a consumir alimentos por cada peso recibido de PROCAMPO. 7 centavos más en vegetales, 15 centavos más en cereales, y 60 centavos más en carne (a costa de otros alimentos). Ello se traduce en un consumo de 2.8 calorías más por cada peso transferido a los beneficiarios de PROCAMPO.
Cattaneo, Galiani, Gertler, Martinez y Titunuk (2006, 2009b)	Programa de Ahorro, Subsidio y Crédito para “Tu Casa”.	Aparejamiento por marcador de propensión.	Modalidad “iniciamos tu casa”: calidad de viviendas (+), acceso a drenaje y agua (+), inversión en la vivienda (+), servicios en la comunidad (-), entorno social (-) y situación laboral (-), depresión y estrés (+). Modalidad “Mejoramos tu casa” y “Crecemos tu casa”: ambientes nuevos en la vivienda (+); mejoras en ambientes ya existentes (+). Calidad de pisos (+), calidad de techos (+), satisfacción con la calidad de la vivienda (+) y de vida (+).	Iniciamos tu casa: 61% menos de superficie de la vivienda. 23% más acceso a sanitarios con conexión de agua; 1% más acceso a drenaje, 2% más acceso a servicio de agua en el terreno y 15% más conexión al agua en el interior de la vivienda. 2% más calidad de ambientes (pisos, paredes y techos), y hasta 27% más ventanas en distintos espacios. 431% más de inversión en la vivienda. 33% más dice haber mejorado su calidad de vida derivado de los cambios en su vivienda, y reportan menor presencia de pandillas (-77%). En contraparte, el entorno de los hogares tratamiento tiene menos acceso a servicios como: teléfonos públicos (-21%), aceras (-26%), pavimento (-25%), alcantarillado (-11%), alumbrado público (-2%); clínicas (-86%),

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.4. Evaluaciones de impacto de otros programas federales en México (Cont.)

Autor y año	Programa evaluado	Método	Impacto	Magnitud
Campuzano, Levy y Zamudio (2007) Cruz-Aguayo (2007)	Programa Hábitat. PROBECAT -SICAT.	Aparejamiento por marcador de propensión. Aparejamiento por marcador de propensión.	Acceso a drenaje (+). Probabilidad de obtener empleo (+), beneficios laborales (+), horas trabajadas a la semana (+), informalidad en las condiciones laborales (+ para mujeres, - para hombres).	dispensarios médicos (-82%), parteras (-76%), hospitales privados (-98%), farmacias (-66%), peor calidad del aire (-49%), y más inundaciones (23%). Finalmente, reportan más días laborales perdidos por enfermedad (38%). “Mejoramos tu casa” y “Crecemos tu casa”: 13% más ambientes nuevos; 28% de mejoras en ambientes ya existentes. Entre 3 y 4% de mejoras a pisos, y entre 1 y 7% a techos, en distintas secciones de la vivienda. Entre 10 y 20% más de satisfacción con la calidad de las viviendas y la satisfacción con sus vidas 3% más hogares con acceso a drenaje. Mujeres: 27 puntos porcentuales (pp) en la probabilidad de obtener empleo. 9 horas más de trabajo a la semana. 22 pp más de probabilidad de contar con servicios de salud en el empleo. Hombres: 20 pp en la probabilidad de obtener empleo. 5 horas más de trabajo a la semana. 27 pp más de probabilidad de contar con servicios de salud en su lugar de trabajo. Otros beneficios (como crédito para vivienda y estímulos monetarios),

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.4. Evaluaciones de impacto de otros programas federales en México (Cont.)

King, Gakidou, Imai, Lakin, Moore, Nall, Ravishankar, Vargas, Téllez-Rojo, Hernández, Hdez. y Hdez-Llamas (2007)	Seguro popular.	Diseño experimental aleatorio, y diferencias en diferencias con efectos fijos y variables instrumentales.	Gasto catastrófico en salud (-), gasto de bolsillo en salud (-)	aunque en todos los casos los porcentajes son mayores para los hombres. Adicionalmente, las mujeres tienen una probabilidad 7 pp mayor de que su contrato sea verbal, y 3 pp más de ser autoempleada o no remunerada (este último indicador es negativo en el caso de los hombres en -3 pp). De -3.1% hasta -6.7% en gasto catastrófico en salud; dependiendo del modelo econométrico aplicado. -30% en gasto de bolsillo en salud.
Soto Romero (Coord.) (2007)	Programa Microrregiones.	Regresión discontinua, apareamiento por marcador de propensión.	A nivel de municipios: agua entubada (+ con Oaxaca), piso de tierra (- sin Oaxaca). A nivel de localidad: Piso de tierra (-).	-3.7% de ocupantes de viviendas sin agua entubada (al incluir Oaxaca) y -2.87% de viviendas con piso de tierra (sin incluir Oaxaca). Disminución de las condiciones estructurales de marginación en aproximadamente 0.411 puntos porcentuales.
BUAP (2008)	Fondo Nacional para Empresas en Solidaridad (FONAES)	Regresión discontinua.	Consolidación de las empresas (+).	No se presentan impactos puntuales en el estudio, por realizarse a través de índices agregados. Adicionalmente, se concluye que empresas productivas que reciben apoyo de otras fuentes diferentes a FONAES tienen una consolidación con mayor rapidez.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.4. Evaluaciones de impacto de otros programas federales en México (Cont.)

Autor y año	Programa evaluado	Método	Impacto	Magnitud
González de Cossio, Gutiérrez, González-Castell, Rodríguez, Unar, Leroy, Gadsden, Hernández-Licona y Gertler (2008)	Programa de Apoyo Alimentario.	Diseño experimental aleatorio.	Crecimiento en niños menores de 5 años (+), aumento de peso en mujeres (+), bienestar económico (+).	0.54 cms. de crecimiento en niños de entre 3 y 4 años, en la modalidad de apoyo en efectivo. Consumo de alimentos entre 14 y 16% superior, con un gasto en alimentos entre 17 y 22% mayor.
Instituto Nacional de Ciencias Médicas y Nutrición “Salvador Zubirán” (INCMN) (2008)	Sistema de Protección Social en Salud.	Diseño experimental por conglomerados y diferencias en diferencias.	Gasto catastrófico en hogares (- para 2006) Gastos hospitalarios (-).	-1.7% en gasto catastrófico ($p < 0.1$); -6.9% de gastos hospitalarios para la primera cohorte en 2006, y -18.5% para la segunda cohorte en 2008.
Mundo, Rivera, Shamah y Villalpando (2008)	Programa Abasto Social de Leche (LICONSA). Leche fortificada	Diseño experimental aleatorio.	Anemia (-) Deficiencia de hierro (-).	-40% de anemia, -85% de deficiencia de hierro.
Cattaneo, Galiani, Gertler, Martínez y Titunik (2009a)	Programa Piso Firme.	Diseño experimental a partir de un experimento natural.	Pisos de cemento (+), infecciones parasitarias (-), diarrea (-), anemia (-) y desarrollo cognitivo (+) en niños; bienestar de adultos (+).	28% más pisos de cemento; -19.6% presencia de parásitos; -12.8% episodios de diarrea; -20.1% de incidencia de anemia (niños). 9% más de desarrollo cognitivo (medido por el <i>McArthur test</i>). 42.8% más de satisfacción con los pisos, 15.1% más con la calidad general de la vivienda y 18.7% más de satisfacción con la calidad de vida (reportada por las madres de los menores).

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.4. Evaluaciones de impacto de otros programas federales en México (Cont.)

Autor y año	Programa evaluado	Método	Impacto	Magnitud
Fuentes Castro (2009)	Fondo PyME.	Aparejamiento por marcador de propensión.	Productividad (+), ventas anuales (+) y pago al trabajo (+).	Incremento en ventas anuales por \$243,286 ($p < 0.1$). Incremento en pago al trabajo por \$67,029 ($p < 0.1$).
Galiani y Gertler (2009)	Programa 70 y más.	Regresión discontinua, diferencias en diferencias.	Necesidad de trabajar (-) Ingreso laboral del hogar (-) salud física y mental de adultos mayores (+), pobreza alimentaria (-).	Para el grupo tratado: 18% de sustitución entre trabajo por pago y actividades familiares. Reducción de 27% en el ingreso laboral del hogar. Incremento de 24% en el gasto del hogar. 25% de reducción de la severidad de la pobreza alimentaria. Para el grupo no tratado, en comunidades de tratamiento (efecto anticipación): 15% de sustitución entre trabajo por pago y actividades familiares. Reducción de 11% en el ingreso laboral del hogar. Incremento en el gasto del hogar en 10%.
Calderón-Madrid (2011)	Programa de Apoyo al Empleo.	Aparejamiento y diferencias en diferencias.	Periodo de salida del desempleo (- para individuos que recibieron el SAEBE).	No hay impactos significativos sobre el salario por hora de los participantes. 6.6% más probable que jefes de familia encuentren empleo en el sector formal.
Aterido, Hallward-Driemeier y Pagès (2011)	Seguro Popular.	Diferencias en diferencias.	Empleo (+), empleo formal (-), participación en el mercado laboral (+).	Entre 0.4 y 0.7 puntos porcentuales menos de formalidad laboral. Reducción de 3.1 puntos porcentuales en el flujo de trabajadores en el sector formal.
Campos-Vázquez (2011)	Programa de Rescate de Espacios Públicos.	Aparejamiento por marcador de propensión.	Percepción de seguridad (+).	Aumento de 16.6 puntos porcentuales en mejoría de condiciones de seguridad percibidas por los hogares como buenas o muy buenas.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.4. Evaluaciones de impacto de otros programas federales en México (Cont.)

Autor y año	Programa evaluado	Método	Impacto	Magnitud
Grogger, Arnold, Leon y Ome (2011)	Seguro Popular.	Diseño experimental aleatorio.	Gasto excesivo en salud (-), gastos de maternidad (-), gasto de población en áreas rurales con acceso a unidades de salud grandes (-).	-1 punto porcentual en gasto en salud; -2.2 puntos porcentuales en gasto excesivo en salud (<i>treatment on the treated</i>). -1.9 puntos porcentuales en gasto catastrófico. Todos efectos en el largo plazo.
CONEVAL (2012)	Estrategia 100 X 100.	Aparejamiento, diferencias en diferencias y regresión discontinua.	Población sin derechohabiencia a servicios de salud (-), viviendas con piso de tierra (-), viviendas sin agua entubada (-).	Entre 12 y 27% más de derechohabiencia, dependiendo la estimación econométrica. -12.8% viviendas con piso de tierra, -0.56% viviendas sin agua entubada, -2.61 % viviendas sin energía eléctrica.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

1.4.3. Otras evaluaciones de impacto

La Tabla 1.5 presenta el resto de las investigaciones encontradas. Al igual que las intervenciones previas, los diseños experimentales se concentran en los años más recientes. En el caso de los programas federales no educativos, la heterogeneidad respecto a sus objetos de estudio domina a las evaluaciones de este apartado.

Este grupo de evaluaciones son las que utilizan en mayor proporción levantamientos propios de información y diseños experimentales. Cabe mencionar el relativamente pequeño número de estas investigaciones, al igual que el hecho de que solo cuatro de ellas recurran a un diseño aleatorio como punto de partida para la evaluación.

Tabla 1.5. Otras intervenciones en México que utilizan evaluación de impacto

Autor y año	Intervención	Método	Impacto	Magnitud
Tan y López-Acevedo (2005)	Programa Comprensivo de Modernización y calidad (CIMO, por sus siglas en inglés).	Diferencias en diferencias.	Crecimiento de la productividad (+), Inversión en capacitación de empleados (+), tasas de uso de la capacidad (+), empresas con mecanismos de control de calidad (+).	Entre 6 y 11% más de productividad por trabajador. 170-190 pesos más de inversión en capacitación por trabajador (a pesos constantes de 1994). 23% más empresas adoptaron medidas de control de calidad.
Cull, McKenzie y Woodruff (2007)	Activos a empresas tras shock de capital.	Diseño experimental aleatorio.	Beneficios de las empresas (+).	20-33% de tasa de retorno al capital.
Urquieta, Tepichin y Téllez (2009)	Proyectos Productivos para Mujeres en el área rural.	Aparejamiento por marcador de propensión.	Actividades en agricultura (+).	14 puntos porcentuales de incremento en actividades de agricultura o ganadería.
Aguirre y Orihuela (2010)	Educación sobre bienestar animal a niños de primero de primaria.	Diseño experimental aleatorio.	Conocimientos de los niños (+), riqueza de conceptos (+).	29% más respuestas correctas; 78% más ricas en conceptos.
Beleche (2010)	Duración del año escolar.	Cuasi-experimento natural.	Desempeño académico (+).	El efecto acumulado de diez días adicionales de instrucción es de 1.37 a 2.42 puntos porcentuales en matemáticas; y de 0.78 a 2.58 puntos porcentuales en español.
López-Acevedo y Tinajero (2010)	Conjunto de programas para pequeñas y medianas empresas.	Aparejamiento por marcador de propensión; diferencias en diferencias.	Valor agregado de los productos (+) y empleo (+).	5% más de valor agregado y 6% más de empleo.

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Tabla 1.5. Otras intervenciones en México que utilizan evaluación de impacto (Cont.)

Autor y año	Intervención	Método	Impacto	Magnitud
Campero, Walker, Atienzo y Gutiérrez (2011)	Intervención en educación sexual para padres.	Diseño cuasiexperimental. Aparejamiento por marcador de propensión.	Retraso en iniciación sexual de adolescentes (+) Uso de condón en adolescentes (+).	6.87% menos adolescentes hombres que iniciaron su actividad sexual. Casi el doble de adolescentes recibió condones por parte de sus padres.
Perez-Arce (2011)	Cupo en la universidad.	Cuasiexperimento.	Permanencia en la escuela (+).	Individuos admitidos de forma inmediata en la universidad tuvieron 19% más probabilidad de continuar que quienes fueron admitidos tras una espera de un año. Cada espacio adicional en la universidad incrementa la permanencia escolar de al menos 0.3 individuos entre los elegibles.
Solovieva, Quintanar y Ortiz (2012)	Método para la enseñanza del concepto de número.	Diseño experimental aleatorio.	Aprendizaje del concepto de número (+).	75% más de alumnos mejoró en su comprensión del concepto de número.
Galiani, Gertler, Undurraga, Cooper, Martínez y Ross (2013)	Distribución de casas prefabricadas en zonas de alta marginación.	Diseño experimental aleatorio.	Bienestar de los hogares (+) Satisfacción con la calidad de vida (+) salud de los niños (+).	20% más de satisfacción con la calidad de los pisos, 28% más de satisfacción con la calidad de vida. Disminución de 27% en la prevalencia de diarrea para los menores ($p < 0.1$, resultado aglutinado para México y El salvador).

Notas: el signo indica el sentido del efecto encontrado. Fuente: cálculos propios.

Una explicación para este bajo número puede encontrarse en el financiamiento. La principal impulsora ha sido la academia; seguida de agencias intergubernamentales extranjeras, tales como el *Abdul Latif Jameel Poverty Action Lab (J-PAL)* o la fundación *Bill & Melinda Gates*. Este es un indicio de que, excluyendo los programas sociales, el uso de la MEI en México es limitado, siendo la cuestión presupuestal y la ausencia de información las principales limitantes para potenciar su uso entre la comunidad de investigación del país.

Aunque el uso de la MEI se ha acelerado en los últimos años, sigue resultando insuficiente para dotar de evidencia sobre el impacto que tienen muy diversas políticas públicas; en particular sociales. Desde la academia también hay otra serie de intereses que sirven para entender el comportamiento económico y que a su vez informan política. Con ello se incrementa el acervo de conocimiento sobre la efectividad de nuevas y más novedosas propuestas.

Como puede desprenderse de un análisis de los estudios antes mostrados, la mayoría de éstos atiende programas sociales. A su vez, los programas en educación, o en componentes educativos, son mayores en número al resto. Las intervenciones en este rubro dan cuenta de impactos positivos y significativos sobre el aprendizaje, cuando la intervención se realiza directamente sobre las escuelas y los alumnos; ya sea a través de becas, atención de profesores, o recursos.

Otro gran tema son los programas e intervenciones con impactos positivos sobre la salud de la población; con políticas sobre nutrición, atención médica, y a grupos vulnerables. Destaca el programa *Piso Firme* (Cattaneo *et. al.* 2008; 2009), que tiene impactos indirectos sobre la nutrición y salud intestinal de los individuos.

El tercer gran tema es el empleo; a través de programas que buscan, directa o indirectamente, mejorar la situación laboral de la población. A través de intervenciones que incorporan, por ejemplo, capacitación educativa o incentivos a la docencia son recurrentes los programas que

inciden sobre la empleabilidad de los individuos (por ejemplo, PROBECAT, Chambanet o Prosoft) o que ayudan a la población a mejorar la calidad de su labor.

1.5. Limitaciones

La revisión llevada a cabo en el presente capítulo no está exenta de limitaciones. En primer lugar, si bien se utilizaron los buscadores más comunes y de mayor divulgación (*Google Scholar, ERIC, EconLit, NBER, JSTOR, Science Direct, ProQuest, Coneval*) es posible que haya un margen para omitir investigaciones que cumplan con los criterios definidos, e incluirlas en la presente revisión. Existe, sin embargo, una disyuntiva entre realizar una búsqueda más amplia hacia investigaciones con menor divulgación y la validez y confiabilidad de dichos resultados. El criterio de selección definido en la presente revisión sí permitió contar con documentos de una calidad mínima (documentos publicados y validados por revisores; con una utilización directa de la MEI).

Otra limitación consiste en que algunas de las investigaciones aquí presentadas -expuestas en términos de los resultados más importantes obtenidos utilizando evaluación de impacto- hacían uso de metodologías que enriquecían los resultados obtenidos por medio de la MEI (véase, por ejemplo, Parker, Todd y Wolpin, 2006; o Attanasio, Meghir y Santiago, 2008). La decisión de excluir otros resultados tuvo como propósito centrar la atención en el uso, difusión y evolución de la metodología aquí presentada.

En tercer lugar, el enfoque aquí utilizado seleccionó intervenciones en nuestro país; perdiendo la oportunidad de enriquecer los resultados de estas intervenciones con comparaciones de intervenciones similares en otros contextos. Esa es una veta de investigación que excede los alcances del presente capítulo.

Una limitación más se halla en que la presente revisión no clasifica la calidad de las investigaciones aquí presentadas. Los criterios que dieron origen a la selección de estos estudios

permiten afirmar que cuentan con criterios científicos rigurosos en su elaboración; sin embargo, sí existen diferencias en la precisión de los datos a los que cada una de estas investigaciones recurre; en las pruebas a las que someten sus hipótesis y en la validación que hacen de sus hallazgos.

De la mano con la limitación anterior, se encuentra el hecho que -en el contenido de estas investigaciones- la MEI es solo una de distintas metodologías de evaluación a las que someten la mayoría de las políticas estudiadas; dejando de lado hallazgos complementarios que enriquecen el conocimiento sobre los efectos de las distintas intervenciones de política pública. Finalmente, el marco conceptual bajo el que se ha concebido la presente revisión no ha permitido analizar aspectos que podrían ser potencialmente relevantes para el estudio de la MEI. Preguntas como la relación que guarda el lenguaje utilizado por las distintas evaluaciones; el enfoque epistemológico que da origen a cada una, o las aportaciones en términos de instrumentos específicos de análisis no son contestadas; siendo propicias a responderse ya sea con técnicas etnográficas, de análisis del discurso, o estadísticas y econométricas; con un objetivo distinto al que se ha propuesto alcanzar con el presente análisis.

1.6. Conclusión: ¿qué pueden aportar las evaluaciones de impacto en México?

El presente capítulo tuvo como propósito presentar una revisión exhaustiva de las investigaciones que utilizan la MEI en México. Si bien la perspectiva aquí abordada se enfoca exclusivamente en la metodología de evaluación, se ha mostrado la relevancia que puede cobrar al estudiar el impacto de diferentes programas y políticas.

A partir de los documentos recopilados, fue posible observar que la investigación en nuestro país siguió una trayectoria particular, detonada por las investigaciones en torno a *PROGRESA/Oportunidades/Prospera* (pese a la existencia de trabajos previos para otros programas).

Basta observar que uno de cada tres documentos encontrados para México, en el periodo estudiado, pertenecen a evaluaciones en torno a este Programa. Una vez descontada dicha literatura, observamos una escasa implementación en México, dominada por diseños cuasiexperimentales o de primera generación frente a los diseños experimentales. Estos últimos han ganado preponderancia sobre el resto con el transcurrir del tiempo; aunque ello no obsta para concluir la necesidad de impulsar una cultura de evaluación basada en evidencia robusta, como la que ofrece la MEI. Esto porque -en el universo de evaluaciones de programas sociales en México- los estudios de esta naturaleza permanecen extremadamente escasos.

Para dimensionar esta carencia, en el periodo 2007-2014 Coneval reconoce 1735 evaluaciones a programas y políticas de desarrollo social, mientras que únicamente 17 corresponden a evaluación de impacto; esto es, una de cada cien (Coneval, *op. cit.*, p. 56). Más aún, si descontamos las evaluaciones de programas federales (generalmente auspiciadas por el gobierno mexicano), la lista de documentos de investigación no excede en número a diez.

Es verdad que las evaluaciones de impacto pueden no ser pertinentes para todos los programas e intervenciones en nuestro país. El trabajo de Leeuw y Vaessen (2009) señala cuatro circunstancias en las que el uso de la evaluación de impacto podría no ser apropiada: 1) Cuando otro tipo de evaluación es más útil para apoyar la decisión que se busca tomar en torno a una política o programa; 2) Cuando desvía la necesidad de contar con un abanico de enfoques de evaluación; 3) Cuando se enfrentan restricciones para ofrecer una evaluación de impacto robusta; y 4) Cuando las circunstancias (en particular políticas) permiten vislumbrar que la evaluación de impacto no servirá para orientar la toma de decisiones (p. 52).

Adicionalmente, aunque la evidencia aquí mostrada indica que existe un uso cada vez más extendido de la MEI -y en particular de los diseños experimentales- esta tendencia no debe adoptarse como un fin en sí mismo. Hay que recordar que el fin último de toda evaluación de

políticas públicas debe ser ayudar a mejorar los programas y políticas que se ofrecen a la población; y que la MEI no es aplicable a todas las circunstancias.

Si bien la metodología aquí estudiada tiene una mayor difusión en Estados Unidos de América²¹ -a través del esfuerzo de universidades y organizaciones- su uso ha crecido progresivamente alrededor del mundo; particularmente entre países desarrollados (Baiz *et. al.*, 2020) y en nuestro país²²; lo que permite vislumbrar un área de oportunidad para medir el impacto de distintas políticas y orientar la toma de decisiones.

Existen lecciones sobre buenas prácticas que podemos extraer de la implementación de la MEI en los países desarrollados. Por ejemplo, en Canadá, Estados Unidos, Francia, Reino Unido, Suecia y Alemania el poder legislativo tiene la misión de evaluar las políticas públicas (Baiz *et. al.*, 2020, *op. cit.*, p. 11). En todos estos países, con excepción de Alemania, los principales productores de evaluaciones de impacto son las grandes universidades nacionales en economía. Y en Alemania, las evaluaciones son mayoritariamente realizadas por institutos de investigación (*ibid*, p. 14). Todos estos países tienen al gobierno como el principal financiador de evaluaciones de impacto, además que el principio de obligatoriedad en la publicación de los resultados se dirige hacia la universalidad. Las conclusiones que se extraen de la implementación de la MEI en estos países son directamente aplicables a nuestro contexto: 1) debe existir una mejor articulación entre la demanda y la producción de evaluación de impacto; 2) producción de las evaluaciones que debe apoyarse sobre principios y valores compartidos: independencia de la evaluación, transparencia, nivel de exigencia metodológica, disponibilidad de fondeo,

²¹ Para una revisión de evaluaciones de impacto, tan solo en el ámbito educativo, véase la página del Banco Mundial: <http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluation/Home.aspx?sD=E> consultado el 13 de octubre de 2018.

²² Con esfuerzos en México como la creación de la sede regional de los *Center for Learning on Evaluation and Results* (CLEAR); o la alianza entre J-PAL para Latinoamérica y el Caribe y el Instituto Tecnológico Autónomo de México (ITAM).

experimentación antes que generalización, atención a la transferibilidad de resultados, y un lugar preponderante del conocimiento científico en la toma de decisiones públicas. 3) reflexionar en torno a una alianza que reúna una gran comunidad de actores interesados. ((Baiz *et. al.*, *ibid*, pp. 43-44).

En la actualidad, esfuerzos como *The Campbell Collaboration* (Davies y Boruch, 2001) se enfocan a generar evidencia respecto a la efectividad de intervenciones y sus efectos de política pública. El objetivo de estos esfuerzos es acercar a los tomadores de decisión evidencia que sea producto de estudios rigurosos; generados con altos estándares científicos alrededor del mundo. Quizá la limitación más importante consista en que la MEI es una metodología que estudia los efectos de intervenciones específicas sobre las variables de interés; pero no es capaz de describir el mecanismo a través del cual dichas variables tienen un impacto. Este tipo de intervenciones han sido conocidas en la literatura como “intervenciones de caja negra” (*black box interventions*, descritas en Green, Ha y Bullock, 2009; e Imai *et. al.*, 2011) ya que se conocen los insumos y el resultado; pero el proceso a través del cual se alcanza ese resultado no es investigado. Queda en manos de autoridades, promotores y académicos discernir las circunstancias en las que ha de implementarse esta metodología, y cuánto fomentar su difusión en un contexto como el nuestro.

1.7. Referencias

- Abdul Latif-Jameel Poverty Action Lab (J- PAL) (2010). “¿Por qué aleatorizar?” La evaluación de programas sociales. Universidad de los Andes. Bogota, Colombia.
- Aguilar, Luis F. (1994) “La Reforma del Estado Mexicano”, en *Gestión y Análisis de Políticas Públicas*, Vol. 47, No. 1, pp. 121-144.
- Aguirre, Virginio y Agustín Orihuela (2010) “Assessment of the Impact of an Animal Welfare Educational Course with First Grade Children in Rural Schools in the State of Morelos, Mexico”, en *Early Childhood Education Journal*, Vol. 38, pp. 27-31.

- Ángeles, Gustavo; Paola Gadsden, Sebastián Galiani, Paul Gertler, Andrea Herrera, Patricia Kariger y Enrique Seira (2011) *Evaluación de Impacto del Programa Estancias Infantiles para Apoyar a Madres Trabajadoras*. Instituto Nacional de Salud Pública (INSP), México.
- Angrist, Joshua D. y Víctor Lavy (1997) *Using Maimonides' rule to estimate the effect of class size on student achievement*, No. w5888. National Bureau of Economic Research.
- Angrist, Joshua D. y Jörn-Steffen Pischke (2008) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- (2014) *Mastering Metrics: The path from cause to effect*. Princeton University Press.
- Aterido, Reyes; Mary Hallward-Driemeier y Carmen Pagés (2011) “Does Expanding Health Insurance beyond Formal Sector Workers Encourage Informality? Measuring the Impact of Mexico’s *Seguro Popular*” The World Bank. *Policy Research Working Paper 5785*.
- Attanasio, Orazio, y Valérie Lechene (2014) “Efficient Responses to Targeted Cash Transfers”, en *Journal of Political Economy*, Vol. 122, No. 1, pp. 178-222. Disponible en <http://www.jstor.org/stable/10.1086/674968>.
- Attanasio, Orazio, Costas Meghir, Ana Santiago y Andrew Shepard (2008) “Better Coating for the Silver Bullet: Improving Conditional Cash Transfers in Urban Mexico” *Inter-American Development Bank Education Division – SCL Working Paper No. 1*.
- Attanasio, Orazio, Costas Meghir y Ana Santiago (2011) “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresá”, en *Review of Economic Studies*, No. 79, pp. 37-66.
- Bagby, Emilie Teresa y Wendy Cunningham (2011) “Risk and Protective Factors for School Dropout in Mexico and Chile”, en *Essays on Human Capital Investments in Developing Countries*, UMI Dissertation No. 3496257, Michigan, EUA.

- Baird, Sarah; Craig McIntosh y Berk Özler (2011) “Cash or Condition? Evidence from a Cash Transfer Experiment”, en *The Quarterly Journal of Economics*, Oxford University Press, Vol. 126, No. 4, pp. 1709-1753.
- Baird, Sarah; Michael Kremer, Joan Hamory Hicks y Edward Miguel (2012) “Worms at Work: Long-Run Impacts of Child Health Gains”, mimeo. Disponible en <https://www.poverty-action.org/sites/default/files/worms-work-long-run-impacts-child-health-gains.pdf>.
- Baker, Judy L. (2000) *Evaluación de impacto de los proyectos de desarrollo en la pobreza*. Banco Mundial.
- Bando, Rosángela (2010) *The Effect of School Based Management on Parent Behavior and the Quality of Education in Mexico*, UMI Dissertation No. 3413310, Michigan, EUA.
- Banerjee, Abhijit y Esther Duflo (2008) “The Experimental Approach to Development Economics” *NBER Working Paper 14467*, Cambridge, Massachusetts.
- Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, y Stuti Khemani (2010) “Pitfalls of Participatory Programs: Evidence from a randomized evaluation in education in India”, en *American Economic Journal: Economic Policy*, Vol. 2, No. 1, pp. 1-30.
- Barceinas Paredes, Fernando (2001) *Capital humano y rendimientos a la educación en México*. Tesis doctoral, Universidad Autónoma de Barcelona, Departamento de Economía Aplicada.
- Bardach, Eugene (1998) *Getting Agencies to Work Together. The Practice and Theory of Managerial Craftsmanship*. Brookings Institution Press, Washington D.C.
- Barzelay, Michael (1992) *Breaking Through Bureaucracy. A New Vision for Managing in Government*. University of California Press, EUA.
- Becker, Gary S. (1964) *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education*. 3a Ed. NBER, The University of Chicago Press, Chicago.

- Becker, Sascha O. y Andrea Ichino (2002) "Estimation of average treatment effects based on propensity scores", en *Stata Journal*, Vol. 2, No. 4, pp. 358-377.
- Behrman, Jere R., y John Hoddinott (2000) "An evaluation of the impact of PROGRESA on preschool child height." *International Food Policy Research Institute*, Washington, DC.
- Behrman, Jere R., Susan W. Parker, y Petra E. Todd (2009) "Schooling impacts of conditional cash transfers on young children: Evidence from Mexico.", en *Economic development and cultural change*, Vol. 57, no. 3, pp. 439-477.
- Behrman, Jere R., Jorge Gallardo-Garcia, Susan W. Parker, Petra E. Todd y Viviana Vélez-Grajales (2012) "Are conditional cash transfers effective in urban areas? Evidence from Mexico", en *Education economics*, Vol. 20, No. 3, pp. 233-259.
- Behrman, Jere R., Piyali Sengupta y Petra Todd (2005) "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico", en *Economic Development and Cultural Change*, Vol. 54, No. 1, pp. 237-275.
- Behrman, Jere R., Susan W. Parker y Petra Todd (2011) "Do School Subsidy Programs Generate Long Lasting Benefits? A Five Year Follow-up of Oportunidades Participants", en *Journal of Human Resources*, Vol. 46, No. 1, pp. 93-122.
- Beleche, Trinidad (2010) "Estimating the Effects of School Year Length on Student Performance in Mexico", en *Three Essays on Institutional Reforms in Mexico*, UMI Dissertation No. 3412769, Michigan, EUA.
- Benemérita Universidad Autónoma de Puebla (2008) *Evaluación Integral 2008-2009 del Fondo Nacional de Apoyos para Empresas en Solidaridad*. BUAP, México.
- Bobba, Matteo, y Jérémie Gignoux (2014) "Policy Evaluation in the Presence of Spatial Externalities: Reassessing the ProgresA Program". *PSE Working Papers* n2011-37.

- Bobonis J. Gustavo, Frederico Finan (2009) "Neighborhood Peer Effects in Secondary School Enrollment Decisions", en *The Review of Economics and Statistics*, Vol.91, pp. 695-716.
- Bono, Pierre Henry ; Sandra Debu, Rozenn Desplatz, Maxime Hayet, Clément Lacouette-Fougère y Alan Trannoy (2018) Vingt ans d'évaluations d'impact en France et à l'étranger. Analyse quantitative de la production scientifique. *France Stratégie. Document de travail*, (2018-06).
- Borja-Aburto, Victor H., José A. González-Anaya, Javier Dávila-Torres, Ramón A. Rascón-Pacheco, y Margot González-León (2015) "Evaluation of the impact on non-communicable chronic diseases of a major integrated primary health care program in Mexico.", en *Family practice*, Vol. 33, no. 3, pp. 219-225.
- Bovens, Mark (2007) "Analysing and Assessing Public Accountability. A Conceptual Framework", en *European Law Journal*, No. 13, Vol. 4, pp. 447-468.
- Bruhn, Miriam, Gabriel Lara Ibarra y David McKenzie (2014) "The minimal impact of a large-scale financial education program in Mexico City", en *Journal of Development Economics*, Vol. 108, pp. 184-189.
- Bruns, Barbara; Deon Filmer y Harry Anthony Patrinos (2011) *Making schools work: New evidence on accountability Reforms*. The World Bank, Washington D.C.
- Bryman, Alan (2001) *Social Research Methods*, 3a Ed., Oxford University Press, Oxford.
- Buddlemyer, Hielke y Emmanuel Skoufias (2003) "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA" *IZA Discussion Paper* No. 827.
- Burtless, Gary (ed.) (1996) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, The Brookings Institution, Washington D.C., EUA.
- Calderón-Madrid, Angel (2011) "Retos para la medición de impacto en remuneración, empleabilidad y tiempo de salida al empleo de beneficiarios de los programas de apoyo a

- desempleados en México SICAT Y SAEBE”, en *Bienestar y política social*, Vol. 5, No. 2, pp. 87-110.
- Calderon-Madrid, Angel y Belem Trejo (2001) “The Impact of the Mexican Training Program for Unemployed Workers on Re-employment Dynamics and on Earnings.”, *Research Network Papers*. Inter-American Bank. Washington, D.C.
- Campero, Lourdes; Dilys Walker; Erika E. Atienzo, y Juan Pablo Gutiérrez (2011) “A Quasi-Experimental Evaluation of Parents as Sexual Health Educators Resulting in Delayed Sexual Initiation and Increased Access to Condoms”, en *Journal of Adolescence*, Vol. 34, pp. 215-223.
- Campos Vázquez, Raymundo M. (2012) *Informe Final para la Medición de Impacto del Programa de Rescate de Espacios Públicos*. Sedesol/El Colegio de México. Disponible en http://www.rescatedelaesperanza.sedesol.gob.mx/work/models/SEDESOL/Resource/3162/1/images/PREP_EvaluacionImpacto.pdf. Consultado el 15 de marzo de 2019.
- Campuzano, Larissa, Dan Levy y Andrés Zamudio (2007) “The Effects of Habitat on Basic Infrastructure”, en *Mathematica Policy Research*. Princeton, EUA.
- Canton, Erik y Andreas Blom (2004) “Can Student Loans Improve Accessibility to Higher Education and Student Performance? An Impact Study of the case of SOFES, Mexico”, *Policy Research Working Paper 3425*, World Bank.
- Cárdenas Denham, Sergio (2011) “Escuelas de doble turno en México. Una estimación de diferencias asociada con su implementación”, en *Revista Mexicana de Investigación Educativa*, Vol. 16, No. 50, pp. 801-827.
- Cattaneo, Matias, Sebastián Galiani, Paul Gertler, Sebastian Martinez y Rocio Titiunik (2009a) “Housing, Health and Happiness”, en *American Economic Journal: Economic Policy*. Vol. 1, No. 1, pp. 75-105.

- (2009b) *Evaluación de resultados de impacto del programa de ahorro, subsidio y crédito para la vivienda progresiva 'Tu Casa' Modalidad 'Iniciamos Tu Casa'*.
- Caudillo, Monica L. y Florencia Torche (2014) "Exposure to local homicides and early educational achievement in Mexico.", en *Sociology of education*, Vol. 87, No. 2, pp. 89-105.
- Cejudo, Guillermo M. (2003) "Public Management Policy Change in Mexico, 1982-2000", en *International Public Management Journal*, Vol. 6, No. 3, pp. 309-325.
- Cling, Jean Pierre, Mireille Razafindrakoto y François Roubaud (2008) L'évaluation d'impact des politiques publiques : enjeux, méthodes, résultats. *Nouvelles approches méthodologiques appliquées au développement (2)*, pp. 15-35.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (2012) *Evaluación de impacto de la Estrategia 100 X 100*. CONEVAL, México.
- (2015) *La política de evaluación en México: 10 años del CONEVAL*. CONEVAL, México.
- Cook, Thomas D. y Monique R. Payne, (2002) "Objecting to the objections to using random assignment in educational research", en Mosteller, F. y R. Boruch (eds.) *Evidence Matters: Randomized Trials in Education Research*, Brookings Institution, Washington, D.C. pp. 150–178.
- Crato, Nuno y Paolo Paruolo (Eds.) (2018) *Data-driven policy impact evaluation: How Access to microdata is transforming policy design*. Springer.
- Cruz Aguayo, Yyannú (2010) *An Evaluation of active Labor Market Policies in Developing Economies: The Mexican Case*, UMI Dissertation No. 3277463, Michigan, EUA.

- Cull, Robert; David McKenzie y Christopher Woodruff (2007) “Experimental Evidence on Returns to Capital and Access to Finance in Mexico”, en *The World Bank Economic Review*. Vol. 22, No. 3, pp. 457-482.
- Darney, Blair G., Marcia R. Weaver, Sandra Sosa-Rubi, Dilys Walker, Edson Servan-Mori, Sarah Prager y Emmanuela Gakidou (2013) “The Oportunidades Conditional Cash Transfer Program: Effects on Pregnancy and Contraceptive Use Among Young Rural Women in Mexico”, en *International Perspectives on Sexual and Reproductive Health*, Vol. 39 No. 4, pp. 205-214.
- Davies, Phillip y Robert Boruch (2001) “The Campbell Collaboration”, en *BMJ*, Vol. 323, pp. 294-295.
- Desplat, Rosen y Marc Ferracci (2016) “Comment évaluer l’impact des politiques publiques?”, en *Évaluation*.
- Dhaliwal, Iqbal; Esther Duflo, Rachel Glennerster, y Caitlin Tulloch (2012) “Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education”, *Abdul Latif Jameel Poverty Action Lab (J-PAL)*, MIT.
- De Braw, Alan y John Hoddinott (2008) “Must Conditional Cash Transfer Programs Be Conditioned to Be Effective? The Impact of Conditioning Transfers on School Enrollment in Mexico” *International Food Policy Research Institute*.
- De Hoyos, Rafael, Vicente A. Garcia-Moreno y Harry Anthony Patrinos (2015) *The impact of an accountability intervention with diagnostic feedback: Evidence from Mexico*. The World Bank.
- De La O, Ana. (2013) “Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico”, en *American Journal of Political Science*, Vol. 57, No. 1, pp. 1-14. Disponible en <http://www.jstor.org/stable/23496539>.

- Del Angel-Mobarak, Gustavo A. (2003) *Paradoxes of Financial Development: The Construction of the Mexican Banking System, 1941–1982*. UMI Dissertation No. 3085279, Michigan, EUA.
- Dhaliwal, Iqbal; Esther Duflo, Rachel Glennerster y Caitlin Tulloch (2012) *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries*. Abdul Latif Jameel Poverty Action Lab (J-PAL), MIT.
- DiNardo, John (2008) “Natural Experiments and Quasi-Natural Experiments”, en Steven N. Durlauf y Lawrence E. Blume (Eds.) *The New Palgrave Dictionary of Economics*, Palgrave-McMillan.
- Díaz, Juan José y Sudhanshu Handa (2006) “An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator: Evidence from Mexico’s PROGRESA Program”, en *The Journal of human Resources*, Vol. 41, No. 2, pp. 319-345.
- Dubois, Pierre; Alain de Janvry y Elisabeth Sadoulet (2003) “Effects on School Enrollment and Performance of a Conditional Transfers Program in Mexico”, en *CUDARE Working Papers*, Department of Agricultural and Resource Economics, UCB, UC Berkeley.
- Duflo, Esther (2000) *Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment* (No. w7860). National Bureau of Economic Research.
- Duflo, Esther; Pascaline Dupas y Michael Kremer (2007) “Peer Effects, Pupil-Teacher Ratios, and Teacher Incentives: Evidence from a Randomized Evaluation in Kenya”, en *American Economic Review*, Vol. 91, No. 4, pp 795-813.
- (2009) “[Additional Resources versus Organizational Changes in Education: Experimental Evidence from Kenya](#)”, *Poverty Action Lab Working Paper*, MIT Department of Economics, Cambridge, Massachusetts.

- Duflo, Esther; Rachel Glennerster y Michael Kremer (2007) “Using Randomization in Development Economics Research: A Toolkit”, *Discussion Paper No. 6059*, Center for Economic Policy Research, London, U.K.
- Duflo, Esther y Michael Kremer (2005) “Use of randomization in the evaluation of development effectiveness”, en *Evaluating Development Effectiveness*, Vol. 7, pp. 205-231.
- Durlauf, Steven N. y Lawrence E. Blume (Eds.) (2008) *The New Palgrave Dictionary of Economics*, Palgrave-McMillan.
- Edwards, Brent y Claudia Elizabeth Loucel Urquilla (2016) “The EDUCO program, impact evaluations, and the political economy of global education reform”, en *Education Policy Analysis Archives*, Vol. 24, No. 92.
- Escalante, Roberto (2010) *Análisis del Programa de Subsidio a la Prima del Seguro Agropecuario*. Universidad Nacional Autónoma de México (UNAM), México.
- Escudero Muñoz, Juan M. (2012) “Evaluación de las políticas educativas: cuestiones perennes y retos actuales”, en *Revista Fuentes*, Vol. 10, pp. 8-31.
- Fagan, Jeffrey A. (1990) “Natural Experiments in Criminal Justice”, en Kimberly L. Kempf (Ed.) *Measurement Issues in Criminology*, Springer-Verlag, New York, pp. 108-137.
- Flores-Lima, Roberto (2007) “Innovaciones en evaluación de Impacto del Servicio de Intermediación Laboral en México” Disponible en http://www.sisi.org.mx/jspsi/documentos/2009/seguimiento/00014/0001400008209_065.pdf. Consultado el 4 de marzo de 2019.
- Fuentes Castro, Hugo Javier (Coord.) (2009) *Evaluación Integral 2008-2009 del Fondo de Apoyo a la Micro, Pequeña y Mediana Empresa (Fondo PyME)*. Centro de Estudios Estratégicos, Tecnológico de Monterrey, México.

- Galárraga, Omar, Sandra G. Sosa-Rubí, César Infante, Paul J. Gertler, and Stefano M. Bertozzi (2014) "Willingness-to-accept reductions in HIV risks: conditional economic incentives in Mexico.", en *The European Journal of Health Economics* 15, no. 1: 41-55.
- Galiani, Sebastián y Paul Gertler (2009) *Primer Seguimiento a la Evaluación de Impacto del Programa de Atención a Adultos Mayores de 70 Años y Mas en Zonas Rurales. Informe Final Sobre los Cambios del Programa 70 y Más*. Instituto Nacional de Salud Pública/Sedesol, México.
- Galiani Sebastian; Paul Gertler, Raimundo Undurraga, Ryan Cooper, Sebastian Martinez y Adam Ross (2013) "Shelter from the Storm: Upgrading Housing Infrastructure in Latin American Slums", en *CEGA Working Paper Series No. WPS 028*, Center for Effective Global Action, UC Berkeley.
- García, Sandra y Juan E. Saavedra (2017) "Educational Impacts and Cost-Effectiveness of Conditional Cash Transfer Programs in Developing Countries: A Meta-Analysis", en *Review of Educational Research*, Vol. 87, No. 5, pp. 921-965.
- García Pérez, Iván Felipe (2010) *Factores escolares y extraescolares en el desempeño académico de alumnos de 6° de primaria en México. Un análisis mediante modelos de ecuaciones estructurales*. Tesina de maestría, CIDE, mimeo.
- Gertler, Paul (2004) "Final Report: The Impact of PROGRESA on Health", en *International Food Policy Research Institute*.
- Gertler, Paul, Alan Fuchs y Jennifer Sturdy (2007) *Evaluación de impacto del Componente Energético del Programa de Desarrollo Humano Oportunidades*. Disponible en <http://www.transparenciapresupuestaria.gob.mx/work/models/PTP/SED/Evaluaciones/CH/PF2009/20s072ei07.pdf>. Consultado el 15 de marzo de 2019.

- Gertler, Paul; Sebastian Martinez, Patrick Premand Laura B. Rawlings y Kristel M.J. Vermeersch (2011) *Impact Evaluation in Practice*. The World Bank.
- Gertler, Paul; Harry A. Patrinos y Marta Rubio-Codina (2008) “Empowering Parents to Improve Education: Evidence from Rural Mexico”, *World Bank Policy Research Working Paper*, No. 3935.
- Gertler, Paul y Harry Anthony Patrinos (2012) “Empowering Parents to Improve Education: Evidence from Rural Mexico”, en *Journal of Development Economics*, Vol. 99, No. 1, pp. 68-79.
- Glewwe, Paul (1999) *The economics of School Quality Investments in Developing Countries. An Empirical Study of Ghana*. Macmillan Press, UK.
- y Michael Kremer (2006) “Schools, Teachers and Education Outcomes in Developing Countries”, en Eric Hanushek y Finis Welch (eds.) *Handbook of the Economics of Education*, Vol. 2, North Holland, Amsterdam, pp. 943-1017.
- , Michael Kremer y Sylvie Moulan (2007) “Many Children Left Behind? Textbooks and Test Scores in Kenya”, en *American Economic Journal, Applied Economics*.
- , Nauman Ilias y Michael Kremer (2009) “Teacher Incentives.” *mimeo*.
- , Eric A. Hanushek, Sarah D. Humpage y Renato Ravina (2011) “School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010”, *NBER Working Paper No. 17554*, Cambridge, EUA.
- González de Cossío, Teresa; J Rivera Dommarco, Gladys López Acevedo y Gloria M. Rubio Soto (eds.) (2008) *Nutrición y Pobreza: política pública basada en evidencia*. Banco Mundial, Sedesol. México
- González de Cossío, Teresa; Juan Pablo Gutiérrez, Luz Dinorah González Castell, Sonia Concepción Rodríguez Ramírez, Mishel Unar-Munguía, Jef L. Leroy, Paola Gadsden,

- Gonzalo Hernández-Licona y Paul J. Gertler (2008) “Evaluación de impacto del Programa de Apoyo Alimentario”, en Teresa González de Cossío, Juan Rivera Dommarco, Gladys López Acevedo y Gloria M. Rubio Soto (eds.) (2008) *Nutrición y Pobreza: política pública basada en evidencia*. Banco Mundial, Sedesol. México.
- Grogger, Jeffrey, Ana Sofía León y Alejandro Ome (2011) *Seguro Popular and Health Spending: Evidence from the National Survey of Household Income and Expenditures*, Harris School of Public Policy Studies, University of Chicago, EUA.
- Grogger, Jeffrey; Tamara Arnold, Ana Sofía León y Alejandro Ome (2011) *Efectos del Seguro Popular sobre el gasto en salud, utilización de servicios médicos, y nivel de salud, y resultados a largo plazo del experimento del Seguro Popular en México y evidencia de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)*. Centro de Investigación y Docencia Económicas (CIDE). México.
- Handa, Sudhanshu, Mari-Carmen Huerta, Raúl Pérez y Beatriz Straffon (2000) *Poverty, Inequality, and Spillover in Mexico's Education, Health and Nutrition Program*, International Food Policy Research Institute, Washington D.C., EUA.
- Hanushek, Eric A. (1981) “Throwing Money at Schools”, en *Journal of Policy Analysis and Management*, Vol. 1, No. 1, pp. 19-41.
- (1986) “The Economics of Schooling”, en *Journal of Economic Literature*, No. 24, pp. 1141-1177.
- (1994) *Making Schools Work. Improving Performance and Controlling Costs*, The Brookings Institution, Washington, D.C.
- y Dennis D. Kimko (2000) “Schooling, Labor Force and the Growth of Nations”, en *The American Economic Review*, Vol. 90 No. 5, pp. 1184-1208.

- y Ludger Wößmann (2010) *The High Cost of Low Educational Performance. The Long-run Economic Impact of Improving PISA Outcomes*, OCDE.
- Heckman, James J. (1974) “Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship”, en *Journal of the American Statistical Association*, Vol. 69, No. 346, pp. 350-354.
- (1979) “Sample Selection Bias as Specification Error”, en *Econometrica*, Vol. 47, No. 1 pp. 53-161.
- Anne Layne-Farrar y Petra E. Todd (1996) “Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship”, en Gary Burtless (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, The Brookings Institution, Washington D.C., EUA.
- , Lance J. Lochner y Petra E. Todd (2005) “Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond”. *Working Paper 11544*, National Bureau of Economic Research.
- Hevia de la Jara, Felipe (2009) "De Progres a Oportunidades: efectos y límites de la corriente cívica en el gobierno de Vicente Fox.", en *Sociológica* No. 24 Vol. 70, pp. 43-81.
- Hoddinott, J. y Emmanuel Skoufias (2004) “The Impact of PROGRESA on Food Consumption”, en *Economic Development and Cultural Change*, No. 53, pp. 37–61.
- Holla, Alaka y Michael Kremer (2008) “Pricing and Access: Lessons from Randomized Evaluations in Education and Health”, mimeo.
- Hunter, Jennifer B.; Jill Guernsey de Zapien, Mary Papenfuss, María Lourdes Fernández, Joel Meister y Anna R. Giuliano (2004) “The Impact of a *Promotora* on Increasing Routine Chronic Disease Prevention Among Women Aged 40 and Older at the U.S.-Mexico Border”, en *Health Education & Behavior*, Vol. 31, No. 4, pp. 18S-28S.

- Instituto Nacional de Ciencias Médicas y Nutrición “Salvador Zubirán” (INCMN) (2008) *Evaluación de Efectos del Sistema de Protección Social en Salud*, INCMN, México.
- Instituto Nacional para la Evaluación de la Educación (2015). *Código de ética del evaluador de desempeño*. Disponible en https://www.inee.edu.mx/wp-content/uploads/2019/01/C%C3%B3digo_de_Etica290915_004_DGAJ.pdf. Consultado el 23 de noviembre de 2021.
- Kempf, Kimberly L. (Ed.) (1990) *Measurement Issues in Criminology*, Springer-Verlag, New York.
- Khandker, Shahidur R., Gayatri B. Koolwal y Hussain A. Samad (2010) *Handbook on Impact Evaluation. Quantitative Methods and Practices*. The World Bank.
- King, Gary; Emmanuela Gakidou, Kosuke Imai, Jason Lakin, Ryan Moore, Clayton Nall, Nirmala Ravishankar, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández, Mauricio Hernández y Héctor Hernández-Llamas (2009) “Public Policy for the Poor? A Randomized Assessment of the Mexican Universal Health Insurance Programme”, en *Lancet* Vol. 373, No. 9673. pp. 1447-54.
- Lalive; Rafael y M. Alejandra Cattaneo (2009) “Social Interactions and Schooling Decisions”, en *The Review of Economics and Statistics*, Vol. 91, pp. 457-477.
- Le Velly, Gwenolé, Alexandre Sauquet, y Sergio Cortina (2015) "Evaluating PES impact and leakages in Yucatan, Mexico: A spatial approach.
- Leeuw, Frans y Jos Vaessen (2009) “NONIE Guidance on Impact Evaluation”; Draft Version for Discussion at the Cairo conference March-April, 2009. Network on Networks on Impact Evaluation.
- Leroy, Jef. L.; Armando García-Guerra, Raquel García, Clara Dominguez, Juan Rivera y Lynnette M. Neufeld (2008) “The Oportunidades Program Increases the Linear Growth of

- Children Enrolled at Young Ages in Urban Mexico”, en *The Journal of Nutrition*, No. 138, Pp. 793-798.
- López-Acevedo, Gladys (1999) “Learning Outcomes and School Cost-Effectiveness in México.”, *Policy research Working Paper No. 2128*, World Bank.
- y Mónica Tinajero (2010) “Impact Evaluation of SME Programs Using Panel Firm Data”, en *World Bank Policy Research Working Paper 5186*.
- y Mónica Tinajero (2013) “Evaluating Different Types of Enterprise Support Programs Using Panel Firm Data: Evidence from the Mexican Manufacturing Sector”, en *Economía*, Vol. 14, No. 1, pp. 1-26.
- López-Calva, Luis F. y Luis D. Espinosa (2006) “Efectos diferenciales de los Programas Compensatorios del CONAFE en el Aprovechamiento Escolar”, *mimeo*.
- Maluccio, John; John Hoddinott, Jere R. Behrman, Reynaldo Martorell y Agnes R. Quisumbing (2009) “The Impact of Nutrition during Early Childhood on Education Among Guatemalan Adults”, en *The Economic Journal*, Vol. 119, No. 537, pp. 734-763.
- Merino, Mauricio; Guillermo Cejudo; David Arellano Gault; Teresa Bracho; María Amparo Casar; José Ramón Gil-García; Claudia Maldonado; Judith Mariscal; Lucrecia Santibáñez y Laura Sour (2010) *Problemas, decisiones y soluciones: enfoques de política pública* (No. 35). Fondo de Cultura Económica/CIDE.
- Meyer, Bruce D. (1995) “Natural and Cuasi-experiments in Economics”, en *Journal of Business and Economic Statistics*, Vol. 13, pp. 151-161.
- Miguel, Edward y Michael Kremer (2004) “Worms: Identifying Impacts on education and Health in the Presence of Treatment Externalities”, en *Econometrica*, Vol. 72, No. 1.

- Miller Flores, Dinorah (2009) *La equidad en la universidad. PRONABES y la condición de juventud de los estudiantes. Una mirada desde la UAM*. Colección Biblioteca de la Educación Superior. Serie Investigadores, ANUIES, Ciudad de México.
- Mincer, Jacob (1974) *Schooling. Experience and Earnings*, Columbia University Press, Columbia.
- Mosteller, Frederick y Robert Boruch (eds.) (2002) *Evidence Matters: Randomized Trials in Education Research*, Brookings Institution, Washington, D.C.
- Murnane, Richard J., John B. Willet y Sergio Cárdenas Denham (2006) “¿Ha contribuido el Programa Escuelas de Calidad (PEC) a mejorar la educación pública en México?”, en Fernando Reimers (coord.) *Aprender más y mejor. Políticas, programas y oportunidades de aprendizaje en educación básica en México*. Fondo de Cultura Económica, México.
- Organización para la Cooperación y el Desarrollo Económico (OCDE) (2010) *The High Cost of Low Educational Performance: The Long-Run Economic Impact of Improving PISA Outcomes*, OCDE, París.
- (2011) *Panorama de la educación 2011*. Nota de país-México. Disponible en <http://www.oecd.org/mexico/48667648.pdf>. Consultado el 15 de marzo de 2019.
- Ortega-Hesles, María Elena (2012) “Lessons from the Pilot Study of a Cluster Randomized Trial: Summer Reading Intervention Targeting Third Grade Students in Mexico”, *Doctoral dissertation*, Harvard Graduate School of Education.
- Parker, Susan; Luis Ruvalcaba y Graciela Teruel (2002) “Schooling Inequality among the Indigenous: A Problem of Resources or Language Barriers?”, en *Inter-American Development Bank. Research Network Working Paper #R-438*.
- Parker, Susan, Petra E. Todd y Kenneth I. Wolpin (2006) “Within-Family Program effect estimators: The Impact of Oportunidades on Schooling in Mexico”, mimeo.

- Parker, Susan, Jehre R. Behrman Luis Ruvalcaba y Graciela Teruel (2007) “[The quality of education in Latin America and the Caribbean Region: The Mexican case](#)” En *Inter-American Development Bank*.
- Perez-Arce, Francisco (2011) “Is A Dream Deferred a Dream Denied? College Enrollment and Labor Market Search”, en *Labor and Population Working Paper WR-843*, RAND Corporation.
- (2017) “The effect of education on time preferences”, en *Economics of Education Review*, Vol. 56, pp. 52-64.
- Pritchett, Lant; Salimah Samji y Jeffrey Hammer (2012) “It’s All About MeE. Using Structured Experiential Learning (‘e’) to Crawl the Design Space”, *CID Working Paper No. 249*, Center for International Development at Harvard University.
- PROGRESA (1997) *PROGRESA: Programa de Educación, Salud y Alimentación*. México.
- Ramírez-Silva, Ivonne; Juan A. Rivera, Jef L. Leroy y Lynnette M. Neufeld (2013). “The Oportunidades Program’s Fortified Food Supplement, but Not Improvements in the Home Diet, Increased the Intake of Key Micronutrients in Rural Mexican Children Aged 12–59 Months”, en *The Journal of Nutrition*, pp. 656-663.
- Ravallion, M. (1999) *The mystery of the vanishing benefits: Ms Speedy analyst's introduction to evaluation* (Vol. 2153). World Bank Publications.
- (2015) *The economics of poverty: history, measurement, and policy*. Oxford University Press.
- Reimers, Fernando y Sergio Cárdenas Denham (2006) “Who Benefits from School-Based Management in Mexico”, en *Prospects*, Vol. 37, No. 1, pp. 37-56.

- Revenga, Ana; Michelle Riboud y Hong Tan (1994) "The Impact of Mexico's Retraining Program on Employment and Wages.", en *The World Bank Economic Review*, Vol. 8, No. 2, pp.247-278.
- Rivera, Juan A.; Daniela Sotres-Álvarez, Jean-Pierre Habitch, Teresa Shamah y Salvador Villalpando (2004) "Impact of the Mexican Program for Education, Health and Nutrition (Progresa) on Rates of Growth and Anemia in Infants and Young Children", en *Journal of the American Medical Association*, Vol. 291, No. 21 Pp. 2563-2570.
- Rosati, Furio C. y Maria Rossi (2007) "Impact of School Quality on Child Labor and School Attendance: The Case of CONAFE Compensatory Education Program in Mexico" *Understanding Children's Work working Paper Series*, Roma, Italia.
- Rosenbaum, Paul y Donald Rubin (1983) "The Central Role of the Propensity Score in Observational Studies of Causal Effects", en *Biometrika*, Vol. 70 No. 1, pp 41-55.
- Rubalcava, Luis y Graciela Teruel (2006) "Guía de Usuario de la Encuesta Nacional sobre Niveles de Vida de los Hogares Primera Ronda".
- (2008) "Guía de Usuario de la Encuesta Nacional sobre Niveles de Vida de los Hogares Segunda Ronda".
- Ruiz-Arranz, Marta; Benjamin Davis, Marco Stampini, Paul Winters y Sudhanshu Handa (2002) "More calories or more diversity? An econometric evaluation of the impact of the PROGRESA and PROCAMPO transfer programs on food security in rural Mexico". Mimeo. Harvard University.
- Salinas-Rodríguez, Betty; Karla Manrique Espinoza, Margarita Moreno Tamayo, Pilar Torres Pereda, Vanessa De la Cruz Góngora, Gustavo Ángeles Tagliaferro, Aarón Rodríguez, y Martha María Téllez-Rojo Solís (2014) *Impact evaluation of the non-contributory social pension programme 70 y más in Mexico*. 3iE Impact Evaluation Report No. 5.

- Sánchez González, José Juan (2005) *Reforma, modernización e innovación en la historia de la administración pública en México*. 2º ed. Porrúa-Grijalbo, México.
- Santibañez, Lucrecia; José Felipe Martínez; Ashlesha Datar; Patrick J. McEwan; Claude Messan Setodji y Ricardo Basurto-Dávila (2007) "Breaking Ground: Analysis of the Assessment System and Impact of Mexico's Teacher Incentive Program *Carrera Magisterial*. Technical Report". Summary. RAND Corporation (2007).
- Schultz, T. Paul (2000) *School Subsidies for the Poor: Evaluating a Mexican Strategy for Reducing Poverty*. Report submitted to PROGRESA. International Food Policy Research Institute, Washington D.C., EUA.
- Secretaría de Desarrollo Social (2008) *Evaluación Externa del Programa Oportunidades 2008. A Diez Años de Intervención en Zonas Rurales (1997-2007)*: Secretaría de Desarrollo Social, México.
- Secretaría de Economía (2009) *Evaluación de impacto del Programa para el Desarrollo de la Industria del Software (evaluación integral 2008-2009 del PROSOFT)*. Secretaría de Economía/Instituto Tecnológico Autónomo de México (ITAM)/Centro de Estudios de Competitividad (CEC). México.
- Seol, Sunmi; Aaron Sharp y Paul Kim (2011) "Stanford Mobile Inquiry-based Learning Environment (SMILE): using mobile phones to promote student inquires in the elementary classroom", en *Proceedings of the 2011 International Conference on Frontiers in Education: Computer Science & Computer Engineering*, pp. 270-276.
- Shadish, William R.; Donald T. Campbell y Thomas D. Cook (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Company, Boston, E.U.A.

- Shavelson, Richard J., y Lisa Towne (2002) *Scientific research in education*. National Academy Press, Washington, DC.
- Shapiro, Joseph y Jorge Moreno Treviño (2004) “Compensatory Education for Disadvantaged Mexican Students: An Impact Evaluation Using Propensity Score Matching”, *World Bank Policy Research Working Paper 3334*.
- Sianesi, Barbara (2004) “An evaluation of the Swedish system of active labor market programs in the 1990s”, en *Review of Economics and Statistics*, Vol. 86, No. 1, pp. 133-155.
- Siegel, Jacob S. y David A. Swanson (Eds.) (2004) *The methods and materials of demography*. Elsevier/Academic Press. San Diego.
- Skoufias, Emmanuel y Susan Parker (2001) “Conditional Cash Transfers and Their Impact on Child Work and Schooling: Evidence from the PROGRESA Program in Mexico”, en *Food Consumption and Nutrition Division Discussion Papers*, No 123, Washington D.C., USA.
- Skoufias, Emmanuel and Joseph Shapiro (2006) “Evaluating the Impact of Mexico’s Quality Schools Program: The Pitfalls of Using Nonexperimental Data,” *SSRN eLibrary*.
- Skoufias, Emmanuel (2001) *PROGRESA and its Impacts on the Human Capital and Welfare of Rural Households in Mexico: A Synthesis of the Results of an Evaluation by IFPRI*. International Food Policy Research Institute, Washington D.C., EUA.
- Skoufias, Emmanuel; Susan W. Parker, Jere R. Behrman y Carola Pessino (2001) “Conditional Cash Transfers and Their Impact on Child Work and Schooling: Evidence from the PROGRESA Program in Mexico”, en *Economía*, Vol. 2, No. 1, pp. 45-96.
- Solovieva, Yulia; Luis Quintanar y Gerardo Ortiz (2012) “The Formation of Initial Components of Number Concepts in Mexican Children”, en *Psychology Research*, Vol. 2, No. 7, pp. 422-430

- Tan, Hong y Gladys López-Acevedo (2005) “Evaluating Training Programs for Small and Medium Enterprises”, en *World Bank Policy Research Working Paper 3760*.
- Todd, Petra y Kenneth I. Wolpin (2006) “Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling: Assessing the Impact of a School Subsidy Program in Mexico”, en *American Economic Review*, Vol. 96, No. 5, pp. 1384–1417.
- Toulmin, Stephen E. (2003) *The Uses of Argument*. Cambridge University Press, New York.
- Urquieta, José E.; Ana María Tepichin y Martha María Tellez (2009). “Poverty and Gender Perspective in Productive Projects for Rural Women in Mexico: Impact Evaluation of a Pilot Project”, en *Evaluation Review*, Vol. 33, No. 7, pp. 27-53.
- Vilalta, Carlos (2008) “Relaciones indirectas y variables asociadas a la duración de las penas y las apelaciones en México: evidencia y comentarios”, en *Estudios Sociológicos*, Vol. XXVI, No. 2, pp. 409-433.
- Villalpando, Salvador; Teresa Shamah Levy, Verónica Mundo y Juan Ángel Rivera Dommarco (2008) “La fortificación de la leche Liconsa con micronutrientes: un ejemplo del uso de la investigación para la toma de decisiones en políticas sociales”, en Teresa González de Cossío, J Rivera Dommarco, G López Acevedo y GM Rubio Soto, editores. *Nutrición y Pobreza: política pública basada en evidencia*. Banco Mundial, Sedesol. México.
- White, Howard y David A. Raitzer (2017) *Impact evaluation of development interventions: A practical guide*. Asian Development Bank.
- Wooldridge Jeffrey M. (2008) *Introductory Econometrics: A Modern Approach*, 4a Ed. South-Western College Pub, Chicago.
- Zaga, Daniel (2014) *The Impact of Three Mexican Nutritional Programs: The Case of Df-Puebla*. No. 09-2015. Centre for Finance and Development, The Graduate Institute.

Zeisel, Hans (1973) "Reflections on Experimental Techniques in the Law", en *Journal of Legal Studies*, Vol. 2, No. 1, pp. 107-124.

CAPÍTULO 2

Escuelas de doble turno: una evaluación con métodos no experimentales

“Ah! Señor cura, usted no ama las rudezas del mundo real. Cristo sí las amaba. (...) Cuando escribió: *Sinite parvulos...*, no distinguía entre los niños pequeños. No se molestó con la cercanía del descendiente de Barrabás o del descendiente de Herodes. Señor, su inocencia es en sí misma una corona. La inocencia no hace sino enaltecerlos. Es igual de augusta aún desgarrada y sin flores.”

Victor Hugo. *Los miserables*.

2.1. Introducción

Las escuelas de doble turno comprenden un conjunto de políticas que los países siguen; a fin de maximizar las oportunidades de acceso a educación para los alumnos en un contexto de recursos escasos. De acuerdo con el trabajo de Bray (2008, p. 17), en un sistema de doble turno o bisesional “las escuelas atienden dos grupos completamente distintos de alumnos a lo largo del día. (...) Cada grupo utiliza los mismos edificios, equipamiento y otras instalaciones”. Este es el tipo de arreglo más común al ofrecer diferentes calendarios de trabajo en una misma escuela, a pesar de existir sistemas con más de dos turnos escolares en el mismo plantel (Bray, 1989b, pp. 45-49).

La investigación sobre turnos escolares da cuenta de al menos 45 países que están implementando o han implementado esta estrategia (Lusher y Yassenov, 2015; Bray, 2008, *op. cit.*; Cárdenas Denham, 2009; Kea, 2007; Kleinhans, 2008 y Sheryn, 2011), mostrando lo difundida que se encuentra esta política en el mundo. En México existe evidencia de que este

sistema presenta segregación, siendo los estudiantes que no se encuentran en el turno matutino quienes tradicionalmente están en desventaja (Cárdenas, 2011; 2009, *op. cit.*).

Este capítulo explora una política pública, llevada a cabo por autoridades de una entidad en México, que consiste en una lotería para decidir qué niños asisten a cada turno escolar en el nivel de primaria. El objetivo es evaluar el efecto que una redistribución como ésta tiene sobre los alumnos; medido a través de las brechas en reprobación, deserción, repitencia, extraedad y rezago educativo, además de las brechas en desempeño en los exámenes ENLACE de español y matemáticas. Utilizando una metodología de apareamiento y diferencias en diferencias se encuentra que la política tiene efectos que incrementan las brechas de desigualdad entre turnos, a pesar de reducir la brecha en desempeño.

El capítulo se estructura de la siguiente manera: la sección 2 presenta las principales conclusiones a las que llega la literatura sobre escuelas de doble turno a nivel internacional; subrayando las especificidades de este tipo de escuelas en México. La sección 3 describe la intervención que tuvo lugar en un estado de la república mexicana, así como las características de esta política pública en la entidad. La sección 4 contiene el diseño de investigación y la metodología del estudio. La sección 5 lleva a cabo las estimaciones presentando los principales hallazgos, y la sección 6 concluye.

2.2. El doble turno

Siendo el objeto de estudio de este capítulo los efectos de implementaciones diferenciadas sobre la escolaridad de doble turno (en particular, si los estudiantes fueron seleccionados aleatoriamente a un turno específico o asignados de acuerdo con algún criterio), la Tabla 2.1 muestra literatura reciente acerca de la forma en que esta política se lleva a cabo alrededor del orbe. Aquí se hace énfasis en tres insumos relevantes: estudiantes, personal e infraestructura;

para mostrar cómo asocian los autores citados las particularidades en la implementación con los resultados observados.²³

De las políticas de doble turno se pueden obtener algunas lecciones, independientemente de los diferentes contextos en los que se lleva a cabo su implementación. Primero, tanto los alumnos como los padres de familia tienden a considerar el turno vespertino menos deseable que el matutino; a pesar de la evidencia que muestra -para algunos países- que no hay diferencia en el desempeño de los estudiantes que atienden uno u otro. La jornada vespertina, por su parte, continúa siendo una opción deseable y costo-efectiva para diferentes gobiernos; porque las naciones que enfrentan recursos humanos, materiales o de espacio físico insuficientes para lidiar con la demanda de maestros y escuelas, necesitan maximizar el uso de los activos existentes.

Razones que justifican la implementación de escuelas de doble turno incluyen ahorros en materia de construcción de escuelas y uso de tierra; recursos que pueden ser redirigidos hacia políticas educativas más efectivas y eficientes, tales como: más espacio y equipamiento en los planteles existentes, mejores salarios y condiciones de trabajo para profesores y personal administrativo; e incluso una reducción en el número de alumnos por salón de clase (Bray, 1989a).

Otro argumento es que los turnos -distintos al matutino- expanden la oferta para los estudiantes; lo que permite hacer frente a labores que los alumnos no tendrían tiempo de realizar de otra manera. Por ejemplo, en Kea (2007) se presenta evidencia de que la escuela vespertina se ha convertido en una ventana de oportunidad para que las niñas con jornadas de trabajo al inicio del día asistan a la escuela.

²³ Para efectos de este trabajo, el enfoque se centra en la forma de implementar las jornadas matutina y vespertina, aún cuando los países representados utilicen más de dos turnos en sus sistemas educativos. No obstante, cuando un tercer turno tenga incidencia sobre los primeros dos, se hará explícito en el documento.

Tabla 2.1. Consecuencias de la implementación de escuelas de doble turno alrededor del mundo

Autor	País	Particularidades en la implementación del doble turno	Conclusiones
Fiallo-Billini (2003)	República Dominicana	<p><i>Estudiantes.</i> Los turnos matutino y vespertino en educación básica están reservados para los niños y adolescentes, mientras que los turnos nocturnos se reservan usualmente para adultos y trabajadores. Hay evidencia de que los menores con mayor desventaja educativa no están en el turno matutino.</p> <p><i>Personal.</i> Las escuelas pueden compartir profesores y directivos entre turnos, pudiendo trabajar dos jornadas en el mismo día.</p> <p><i>Infraestructura.</i> La jornada vespertina tiene mayores problemas en lo relativo a electricidad y disponibilidad de insumos.</p>	<p>El sistema adolece de duración suficiente para las clases en cada turno; lo que se agrava por el uso ineficiente del tiempo de instrucción. Más importante aún, los maestros y directivos son menos proclives a realizar sus actividades en los turnos vespertino y nocturno, debido al cansancio. La evidencia muestra también una falta de infraestructura en el turno vespertino; pero también beneficios al reducir jornadas de trabajo y estudiantes en cada salón de clases; y permitir a los profesores trabajar un solo turno con salarios adecuados.</p>
Tedla (2003) en Bray (2008)	Eritrea.	<p><i>Estudiantes.</i> Asignados a cada turno de acuerdo con varios aspectos, incluyendo el grado que cursan. Adicionalmente, algunas escuelas cambian cada semestre la asignación a cada jornada escolar.</p> <p><i>Personal.</i> Los profesores en los turnos dobles tienden a dar tutorías privadas.</p> <p><i>Infraestructura.</i> No hubo un esfuerzo por dotar a las escuelas bisesionales de mejor infraestructura en comparación con aquellas de un solo turno.</p>	<p>El sistema de doble turno parece fomentar la proliferación de tutorías privadas por parte de los profesores, dado que la currícula no se cubre totalmente debido a menos horas de instrucción. Ello fomenta un sistema perverso de incentivos para los docentes, quienes se benefician de no cubrir el temario en horas lectivas para orillar a los alumnos a pagar por cursos remediales.</p>
Kleinhans (2004) en Bray (2008)	Uganda.	<p><i>Estudiantes.</i> Se utilizan diferentes criterios para asignar a los estudiantes.</p>	<p>La sociedad en Uganda tiene un concepto negativo acerca del doble turno, considerándolo de calidad</p>

Nota: elaboración propia.

Tabla 2.1. Consecuencias de la implementación de escuelas de doble turno alrededor del mundo (Cont.)

Autor	País	Particularidades en la implementación del doble turno	Conclusiones
Kea (2007)	Gambia.	<p>Algunas escuelas colocan a los estudiantes de primaria en la mañana y a los de secundaria en la tarde, mientras otras enseñan a los mismos grados tanto en la mañana como en la tarde con diferentes estudiantes en cada turno.</p> <p><i>Personal.</i> El Gobierno ofreció un 25% adicional del salario base de los profesores para cubrir los dos turnos.</p> <p><i>Infraestructura.</i> Debido al incremento sustantivo de la matrícula, se llevaron a cabo grandes proyectos de construcción, paralelos a la difusión del doble turno.</p> <p><i>Estudiantes.</i> El doble turno ha aminorado las tensiones derivadas de las obligaciones de las estudiantes de cumplir con el <i>kafo</i>²⁴ y las labores domésticas, por una parte, y la escuela, por otra. Ha ayudado a la intensificación y apropiación de las ganancias del trabajo para estas jóvenes. No obstante, la precarización de la calidad educativa (particularmente en el turno vespertino) y el agotamiento propio de las dobles jornadas de trabajo y estudio, han propiciado la sobreexplotación de las menores.</p> <p><i>Infraestructura.</i> No solo la construcción de infraestructura se ha incrementado en las escuelas de doble turno; otros aspectos también han mejorado, como la separación de baños para niños y niñas.</p>	<p>inferior. Las promesas incumplidas referentes a los pagos extra de los profesores han exacerbado dicha opinión negativa.</p> <p>La agresiva política de atraer niños a la escuela ha sido ampliamente exitosa en Gambia, especialmente en el caso de las niñas. El doble turno ha ayudado a paliar este enorme incremento en la matrícula, y ha generado externalidades positivas para que los niños en labores agrícolas o domésticas asistan a la escuela. Sin embargo, los profesores se quejan de que la calidad educativa se ve comprometida, siendo el turno vespertino el que tiene las mayores desventajas y los profesores con mayor desgaste físico. Adicionalmente, las alumnas enfrentan un esquema de sobreexplotación de su labor que las conduce al agotamiento.</p>

Nota: elaboración propia.

²⁴ El *kafo* son grupos de trabajo de las gambianas, particularmente jóvenes, arraigados en una fuerte cultura comunitaria.

Tabla 2.1. Consecuencias de la implementación de escuelas de doble turno alrededor del mundo (Cont.)

Autor	País	Particularidades en la implementación del doble turno	Conclusiones
Bray (2008) (1990a) (1990b) ²⁵	Hong Kong.	<p><i>Estudiantes.</i> Ambos turnos reciben las mismas horas de instrucción, y un número comparablemente similar de asignaturas, que otras escuelas de un solo turno alrededor de mundo. Esto se logra asistiendo a clases los sábados.</p> <p><i>Personal.</i> El personal administrativo cubre primero el turno matutino, y acomoda el “sobrecupo” por la tarde. Hay diferentes profesores para las sesiones de la mañana y la tarde; y a ciertos profesores se les permite enseñar en ambas jornadas, aunque no de tiempo completo. Los directivos de turnos diferentes dentro de una escuela se organizan para realizar diferentes tareas, a fin de no duplicar funciones.</p> <p><i>Infraestructura.</i> Las escuelas de un solo turno normalmente operan de lunes a viernes, mientras que aquellas de doble turno también lo hacen los sábados. Comparten casi toda la infraestructura en ambas jornadas, excepto las oficinas y aulas para el personal de apoyo y, en menor medida, los libros.</p>	Una fuerte cultura gerencial permitió que la implementación de turnos escolares fuera eminentemente exitosa, sin diferencias notables entre jornadas; con un desempeño que sobresale en casi todo el mundo.
	Indonesia	<p><i>Estudiantes.</i> La asignación entre turnos escolares incluye distinción por grado, para hacer un uso más eficiente del tiempo.</p>	La principal desventaja con el traslape de turnos en Indonesia han sido los planteles saturados en horas pico, con consecuencias como el desorden y la confusión en la escuela.

Nota: elaboración propia.

²⁵ Se eligieron países donde se reportase información para dos o más de las particularidades enunciadas. La investigación se llevó a cabo a partir de Bray (2008), aunque las brechas se llenaron con información de Bray (1990a, 1990b).

Tabla 2.1. Consecuencias de la implementación de escuelas de doble turno alrededor del mundo (Cont.)

Autor	País	Particularidades en la implementación del doble turno	Conclusiones
Bray (2008) (1990a) (1990b) ²⁶	Senegal	<p><i>Estudiantes.</i> La inscripción de los estudiantes se incrementó debido a los turnos escolares, haciendo la asistencia a la escuela más corta y flexible.</p> <p><i>Personal.</i> A los profesores se les permite trabajar en más de un turno debido a la escasez en trabajo calificado. Se les paga solo 25% más cuando enseñan en ambos turnos.</p>	<p>Las escuelas de doble turno que experimentaron una razón alumno-profesor menor reportaron un mejor desempeño. No hubo diferencias explícitas entre turnos; a pesar de incentivos adversos para la jornada vespertina.</p>
	Malasia	<p><i>Personal.</i> Los turnos traslapados permiten compartir profesores entre turnos, sin extender sus horarios de trabajo.</p> <p><i>Infraestructura.</i> Los planteles con buena infraestructura (librería, laboratorios, y talleres) permiten traslapar turnos haciendo un uso eficiente de todos los recursos disponibles.</p> <p><i>Estudiantes.</i> No hay diferencias aparentes entre turnos ni tipos de escuela en alumnos con nivel medio de inglés quienes, en general, estudian en mejores condiciones que el resto de los alumnos malayos.</p> <p><i>Infraestructura.</i> Pocas escuelas de doble turno (usualmente de alumnos con nivel medio de inglés) se diseñaron teniendo en cuenta el calor del turno vespertino de la escuela. El resto no.</p>	<p>A pesar de ello, este sistema democratiza el acceso a los recursos de instrucción, y reduce el ausentismo.</p> <p>Beebout (1972, en Bray, 1989) halló un mayor desempeño de las escuelas de un solo turno que en aquellas con dos turnos, aunque las diferencias se concentraron entre los alumnos que solo hablan malayo. Las explicaciones incluyen diferencias en infraestructura y apoyo en el hogar.</p>

Nota: elaboración propia.

²⁶ Se eligieron países con información para dos o más de las particularidades enunciadas. La investigación se llevó a cabo a partir de Bray (2008), aunque las brechas se llenaron con información de Bray (1990a, 1990b).

Tabla 2.1. Consecuencias de la implementación de escuelas de doble turno alrededor del mundo (Cont.)

Autor	País	Particularidades en la implementación del doble turno	Conclusiones
Orko-dashvili (2009)	Países de la África Sub-Sahariana (Etiopía, Gambia, Ghana, Guinea, Senegal, Tanzania, Zambia, Zimbabwe)	<p><i>Estudiantes.</i> En general, el segundo turno recibe menos horas de instrucción.</p> <p><i>Personal.</i> Algunos países prohíben a los profesores trabajar en más de un turno, mientras otros pagan un extra de sus salarios base, que es menos del doble.</p> <p><i>Infraestructura.</i> Insuficiente para cubrir las crecientes demandas por cobertura, a pesar de los enormes esfuerzos por construir más escuelas.</p>	<p>Al igual que Bray (2008) Orkodashvili observa que el aprendizaje no necesariamente se ve afectado por la implementación del doble turno, dado que puede existir una razón alumno-profesor menor y una mejoría en las instalaciones cuando son compartidas (como en Senegal).</p>
Sheryn (2011)	Maldivas.	<p><i>Estudiantes.</i> Criterios diferenciados para la asignación entre turnos; tales como: el grado, el flujo de estudiantes y el género.</p> <p><i>Personal.</i> Algunos profesores trabajan dos jornadas. Generalmente, el resto del personal es diferente en cada turno.</p> <p><i>Infraestructura.</i> Debido a la escasez de territorio, la mayoría de las escuelas son bisesionales. Los salones de clase normalmente están saturados.</p>	<p>La asistencia a la escuela se incrementó entre la población, al igual que las tasas de alfabetismo. No obstante, debido a la menor duración de la jornada escolar, los alumnos destinaron menos tiempo a las actividades académicas. La autora asocia este efecto con un incremento en la delincuencia juvenil, a pesar de aceptar que la disminución en horas de aprendizaje no es la única causa.</p>
Bervell, Cobbina Boadu (2014); Katai (2013)	Ghana.	<p><i>Estudiantes.</i> Los alumnos alternan entre turnos cada dos semanas para ofrecer igualdad de oportunidades.</p> <p><i>Personal.</i> Ghana no permite a los profesores trabajar ambas jornadas. Enseñan en un grupo particular con sesiones alternadas (mañana y tarde) en un periodo de dos semanas a lo largo del año académico.</p> <p><i>Infraestructura.</i> El mobiliario escolar no se diseñó para el calor vespertino. Ambos turnos comparten infraestructura.</p>	<p>El cambio entre turnos sirvió como un mecanismo de compensación académica. El calor vespertino afectó el aprendizaje de los estudiantes, generando cansancio. También se involucraron menos en las tareas, debido a la falta de tiempo.</p> <p>El doble turno permitió a los alumnos elevar su ingreso a través del trabajo cuando no están en la escuela; y a los profesores obtener tiempo extra</p>

Nota: elaboración propia.

Tabla 2.1. Consecuencias de la implementación de escuelas de doble turno alrededor del mundo (Cont.)

Autor	País	Particularidades en la implementación del doble turno	Conclusiones
Sagyn-dykova (2014); Cárdenas (2009)	México.	<p><i>Estudiantes.</i> Los directores tienen poder discrecional para decidir qué estudiantes se asignan a cada turno en la mayoría de las escuelas; decisión que es influenciada por el poder de negociación de los padres. También hay una práctica de asignar el mismo turno a los hermanos menores.</p> <p><i>Personal.</i> Se encontraron diferencias en las características de los profesores y los directivos entre jornadas. Más de la mitad de los planteles tuvieron personal menos capacitado en el turno vespertino frente al matutino.</p> <p><i>Infraestructura.</i> No hay evidencia de mejora en la infraestructura de las escuelas de doble turno frente a las unisionales. Más aún, existe evidencia de que los turnos pueden no compartir todos los materiales al interior de una escuela, quedando en desventaja el turno vespertino.</p>	<p>para preparar las lecciones o para trabajar en su zona escolar.</p> <p>Adicionalmente, hay un déficit de 40 minutos en la sesión vespertina, lo que reduce las horas de contacto en la misma duración, a lo largo del año académico. Los alumnos que procrastinan en los alrededores de la escuela generalmente ponen como excusa el pertenecer a un turno distinto al que está en labores.</p> <p>Incremento en la carga de trabajo para los directivos y cansancio para los profesores. Existe evidencia de que las escuelas del turno vespertino tienen menor acceso a equipamiento y recursos. También se encontró que, en promedio, las escuelas durante la jornada vespertina tienen mayores tasas de deserción, extraedad, y reprobación que las escuelas del turno matutino, aún cuando compartan infraestructura.</p> <p>Sagyndykova (2014, <i>op. cit.</i>) encontró que la mayoría de las diferencias observadas en cada turno pueden ser atribuibles a las características de los alumnos y los profesores presentes en cada uno de éstos.</p>

Nota: elaboración propia.

Aún en países con altos ingresos per capita, donde el trabajo infantil no está permitido, esta política podría expandir las opciones para los estudiantes de mayor edad; en niveles educativos como la secundaria o la preparatoria.

En el contexto mexicano, la investigación de Cárdenas (2011) evidencia una diferenciación en la calidad y cantidad de recursos educativos que se brindan entre jornadas escolares, concentrando el turno vespertino una mayor cantidad de estudiantes en situación de pobreza, resultados académicos bajos, menor calidad de insumos escolares y mayores tasas de deserción, extraedad y reprobación que las escuelas en el turno matutino con las que comparten plantel. En Sagyndykova (2014), no obstante, se muestra que *prima facie* asistir a la escuela por la mañana o por la tarde no es una causa natural del bajo desempeño. Tras controlar por las características del profesor y los estudiantes, dicho trabajo muestra que el turno al que se asiste no afecta el logro académico.

No obstante, el enfoque que presenta este documento difiere del tomado por Sagyndykova, quien señala que “las escuelas de doble turno en México sirven a su propósito proveyendo igualdad de oportunidades educativas para todos los estudiantes” (p. 12, traducción propia). La hipótesis que aquí se explora, dado que la distribución de alumnos es una decisión principalmente de directivos y autoridades escolares (Cárdenas, 2009, op. cit., pp. 20-36), es que las diferencias observadas en logro no surgen naturalmente del turno al que asisten los estudiantes, dadas sus características. Por el contrario, se busca probar que la asignación implementada tiene efectos en las brechas de desempeño; y explorar si estas diferencias pueden aminorarse para mejorar el *statu quo*.

La evidencia internacional respalda la idea que una distribución más equitativa de insumos al interior de las escuelas puede conducir a cerrar las brechas de desempeño entre turnos; con resultados deseables. Por ejemplo, una de las sociedades más exitosas en términos de logro

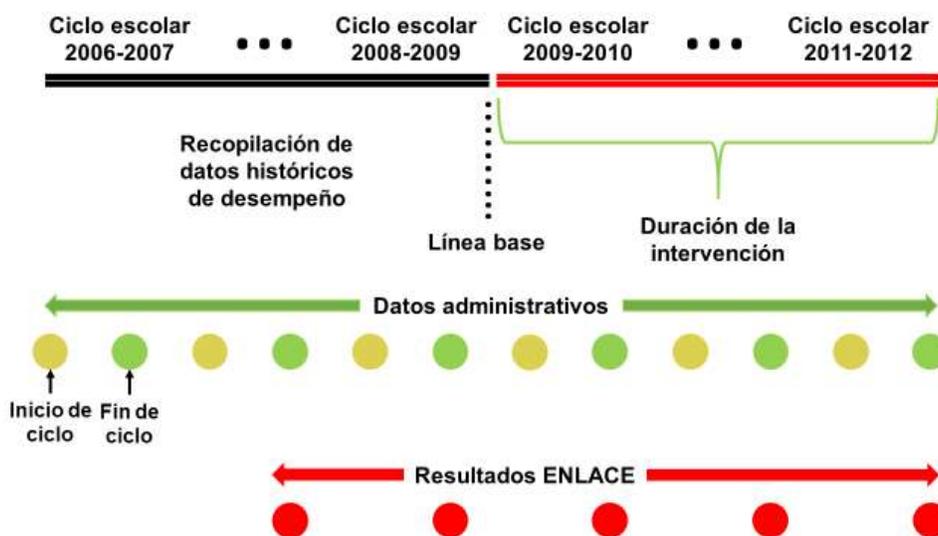
académico (Hong Kong), ha aplicado la modalidad de doble turno hace más de medio siglo; sin diferencias observables entre mañana y tarde, fatiga extrema (*burnout*) de los profesores u horarios de trabajo dobles o triples (Bray, 1989a; 1990b). Otros países del oriente de Asia han seguido esta estrategia con resultados similares (Bray, 2008, *op. cit.*).

Para explorar la hipótesis de una reducción en la brecha en desempeño -a partir de una mejor distribución de los recursos educativos en las escuelas de México- necesitamos contrastar estudiantes en los planteles tradicionales con otros alumnos, similares en sus características, que asisten a escuelas más equitativas. Esta posibilidad se presenta al estudiar la asignación que tuvo lugar en un estado de la república mexicana, a través de una lotería que propició una distribución de alumnos a ambos turnos con una probabilidad similar de pertenecer a uno u otro, al ingresar a la primaria. Se busca probar que esta diferente asignación tuvo consecuencias sobre la equidad en los resultados educativos y la prueba ENLACE.

2.3. Descripción de la intervención

Durante el periodo comprendido entre 2009 y 2011, las autoridades de un estado en la república mexicana llevaron a cabo una lotería para asignar a los estudiantes que ingresaron al primer grado en el nivel de primaria al turno matutino o al vespertino en las escuelas públicas de la región. La entidad en la que se implementó la política es una de las más pequeñas en el país. Este estado fue el único en hacer una reforma así a sus leyes escolares en el periodo; de acuerdo con información provista por las unidades de transparencia de las 32 entidades federativas del país. Del mismo modo, ningún otro gobierno reportó llevar a cabo una lotería, como política estatal, para asignar a los estudiantes entre turnos. A fin de capturar los efectos sobre la entidad, se diseñó un estudio a partir de los registros administrativos de las autoridades escolares; y de las pruebas estandarizadas que se aplicaron a los alumnos. La Figura 2.1 presenta la línea de tiempo de este diseño de investigación.

Figura 2.1. Línea de tiempo de la intervención



Fuente: elaboración propia.

El objetivo del estudio fue medir, a partir del cambio que constituyó esta nueva política, el impacto de una asignación aleatoria de estudiantes a las jornadas escolares matutina y vespertina. Este método de asignación de alumnos dista del tradicional, que selecciona a los alumnos con base en características específicas (como el grado que cursan o el parentesco), o discrecionalmente por parte del director (como se describe en Cárdenas, 2009, *op. cit.*).

Dado que esta política de aleatorización entre turnos escolares dio inicio al comenzar el ciclo escolar 2009, se utilizó información preintervención para aproximar un grupo contrafactual de escuelas que representa lo que hubiera sucedido con las escuelas tratadas de no haber participado en la lotería. Se recopiló información de los datos administrativos de las escuelas, tanto al inicio como al final del año académico. Asimismo, se tomaron como medidas de desempeño de los alumnos los resultados de un examen estandarizado, aplicado a nivel nacional, conocido como

prueba ENLACE (acrónimo de Examen Nacional del Logro Académico en Centros Escolares), de alcance universal a partir del ciclo escolar 2007-2008.

En 2009, año de la intervención, poco más de 31% de todas las escuelas primarias con financiamiento público -en la entidad donde tuvo lugar esta política- contaban con la modalidad de doble turno, llegando al 35% en 2011; el último año observado. Adicionalmente, el número de planteles con esta modalidad se mantuvo constante entre el ciclo escolar 2007-2008 y el ciclo escolar 2010-2011. Estas escuelas constituyen el grupo de tratamiento.

La intervención se llevó a cabo cuando los estudiantes ingresaron a las escuelas de primaria y secundaria; existiendo la posibilidad de que los estudiantes cambiaran el turno que resultó de la lotería en periodos subsecuentes; lo que podría diluir los impactos. Asimismo, las autoridades en la entidad reconocieron que no tenían registros para confirmar que todos los planteles en la muestra utilizaron esta asignación aleatoria, siendo el efecto aquí estimado un mínimo de los impactos que esta intervención podría lograr. Es plausible, además, que algunas familias decidieran -tras el primer año- cambiar a sus hijos a otro plantel, en caso de que no desearan seguir asistiendo al mismo turno escolar, por lo que se trata de una intervención con cumplimiento imperfecto.

2.4. Metodología

El tratamiento recibido (la lotería mandatada) afectó a todos los alumnos pertenecientes a las escuelas públicas de doble turno en la entidad estudiada. Esto generó dos efectos. En primer lugar, hizo comparables en sus características a los alumnos del turno matutino con aquellos del turno vespertino en las escuelas que efectivamente llevaron a cabo la aleatorización; pero los registros administrativos solo permitieron garantizar que la legislación fue aplicable para todas las escuelas antes mencionadas; no que todas las escuelas la llevaron a cabo.

La focalización de este programa público sobre un grupo específico de la población permitiría que la lotería, de haberse aplicado universalmente, hiciera indistinguibles a los alumnos inscritos en el programa para el caso de los turnos matutino y vespertino. Las diferencias encontradas entre el ciclo 2009-2010 y el ciclo 2011-2012 evidencian que no fue así.

La Tabla 2.2 presenta indicadores relacionados a las características de estas escuelas en el año 2008. En ella se muestran diferencias estadísticamente significativas entre turnos antes de la implementación del programa (resultado similar al de Cárdenas, 2009, *op. cit.*).

Entre estas diferencias es posible identificar valores con significancia estadística en indicadores asociados a las condiciones de las escuelas; tales como el porcentaje de estudiantes indígenas, o la proporción de alumnos que han recibido beneficios de *Oportunidades* (considerados la población con mayores carencias en el país). Como se observa en dicha tabla, el número de salones de clase, estudiantes y profesores es mayor en la jornada matutina. Aún en términos relativos, el número de estudiantes por salón de clase es mayor en este turno (33.34 frente a 27.87, en promedio). No obstante, el vespertino tiene un mayor porcentaje de alumnos en el Programa *Oportunidades*, asociado a un mayor nivel de pobreza.

En el turno vespertino existe además un porcentaje mayor de estudiantes indígenas; así como una fracción menor de niños que previamente han estudiado al menos dos años de preescolar. Los porcentajes de alumnos discapacitados -así como los menores que son beneficiarios de los servicios USAER²⁷- también son mayores en el turno vespertino. Con respecto a los puntajes de

²⁷ Acrónimo de Unidad de Servicios a la Atención Regular, ayudando a menores con necesidades educativas especiales. (<https://usaer17bcs.wordpress.com/2011/01/16/definicion-de-la-u-s-a-e-r/>, consultado el 17 de septiembre de 2013).

Tabla 2.2. Valores promedio y desviaciones estándar, escuelas primarias en 2008

Variable	Matutino	Vespertino	Diferencia entre turnos matutino - vespertino
Inscripción total	408.26 (198.18)	232.86 (154.28)	175.40*** [232.53]
Salones de clase	13.38 (4.79)	12.36 (5.63)	1.01*** [38.14]
Profesores frente al aula	12.40 (4.90)	8.67 (4.50)	3.73*** [181.85]
Director	0.92 (0.26)	0.83 (0.38)	0.09*** [47.43]
Profesores en <i>Carrera Magisterial</i>	6.99 (4.59)	4.70 (3.35)	2.29*** [113.66]
Oportunidades (%)	6.18% (0.1)	8.60% (0.15)	-2.42%*** [-51.66]
Indígenas (%)	0.68% (0.11)	0.91% (0.08)	-0.23%*** [-7.20]
Preescolar (%)	4.58% (0.06)	3.65% (0.06)	0.93%*** [23.97]
Beneficiarios de USAER (%)	1.91% (0.04)	2.83% (0.06)	-0.93%*** [-27.29]
Menores con habilidades sobresalientes	0.68% (0.06)	0.91% (0.06)	-2.34*** [-7.20]
Menores con necesidades especiales (%)	1.25% (0.03)	1.93% (0.06)	-0.68%*** [-27.39]
Prueba <i>ENLACE</i> de Español, nivel 0	13.37% (10.78)	18.12% (15.63)	-4.75%*** [-53.52]
Prueba <i>ENLACE</i> de Español, nivel 1	42.86% (13.88)	45.29% (18.05)	-2.43%*** [-21.76]
Prueba <i>ENLACE</i> de Español, nivel 2	35.64% (14.54)	28.93% (18.45)	6.71%*** [62.29]
Prueba <i>ENLACE</i> de Español, nivel 3	7.48% (10.46)	5.17% (11.13)	2.31%*** [0.08]
Prueba <i>ENLACE</i> de Matemáticas, nivel 0	16.09% (12.01)	21.27% (17.03)	-5.17%*** [0.11]
Prueba <i>ENLACE</i> de Matemáticas, nivel 1	44.34% (13.41)	44.59% (17.46)	-0.25%** [0.12]

Notas: entre paréntesis se presentan las desviaciones estándar, y entre corchetes los estadísticos t. Códigos de significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01. Fuente: cálculos propios con base en la Estadística 911 SEP.

Tabla 2.2. Valores promedio y desviaciones estándar, escuelas primarias (Cont.)

Variable	Matutino	Vespertino	Diferencia entre turnos matutino - vespertino
Prueba ENLACE de Matemáticas, nivel 2	29.51% (12.38)	24.71% (15.88)	4.80%*** [0.11]
Prueba ENLACE de Matemáticas, nivel 3	9.40% (10.88)	6.95% (12.01)	2.45%*** [0.09]

Notas: entre paréntesis se presentan las desviaciones estándar, y entre corchetes los estadísticos t.

Códigos de significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01.

Fuente: cálculos propios con base en la Estadística 911 SEP.

la prueba ENLACE, la jornada vespertina presenta menos estudiantes en los niveles de desempeño considerados bueno o excelente.

Dentro del estado donde se realizó la intervención, no hay un grupo de planteles que no hayan sido potencialmente afectados por el programa. Por otro lado, utilizar planteles de otras entidades como grupo de control no resulta en un grupo de comparación directo. Más aún; la imposibilidad para observar si existió una característica distintiva en la entidad que implementó la lotería, obliga a utilizar un diseño cuasiexperimental. Adicionalmente, la lotería modificó la respuesta de las autoridades respecto a los bienes y servicios provistos en cada turno en las escuelas, como más adelante se muestra. Esto provocó que las jornadas matutina y vespertina tras la intervención fuesen distintas a las jornadas preintervención, en características que van más allá de los alumnos inscritos. Por ello, los efectos de tratamiento sobre los tratados no pueden ser derivados a partir de las cohortes pertenecientes a los ciclos escolares previos, pues los estudiantes y las escuelas no son directamente comparables entre sí.

2.4.1. Datos

Para estimar el efecto de la intervención se utilizaron tres diferentes fuentes de información.

Una primera fuente de información es el Censo Nacional de Escuelas para el nivel de primaria

(conocido como *Formato 911*), de la Secretaría de Educación Pública (*SEP*). Este cuestionario, cuya información proveen los directores de las escuelas al inicio y fin de cada ciclo escolar, presenta información relacionada a cada centro educativo en el país (*centro de trabajo*)²⁸. Ofrece una descripción de los principales indicadores dentro de cada plantel, incluyendo alumnos repetidores, reprobados y que han abandonado la escuela; así como el número de profesores por nivel de instrucción.

La segunda fuente de información es la base de datos del programa de transferencias condicionales conocido como *Oportunidades*, de la Secretaría de Desarrollo Social (*Sedesol*), con información concerniente a la proporción de estudiantes que han recibido beneficios de éste -el programa para combatir la pobreza que tenía mayor cobertura en el país- desagregado por escuela.

La base de datos de *Oportunidades* constituye una referencia cuando se trata de comparar la estratificación social entre turnos escolares, dado que los beneficiarios de este programa son elegidos en un proceso de dos etapas: en la primera fase se identifican las comunidades pobres del país y posteriormente, al interior de estas comunidades, se seleccionan las familias con mayor precariedad económica. Por ello, los centros de trabajo que compartiesen el mismo plantel tendrían la misma oportunidad de incorporar a un estudiante inscrito en *Oportunidades*; si las probabilidades de que dicho estudiante asistiese a cualquier turno fueran iguales.

Finalmente, la tercera fuente de información es la proporción de alumnos en cada uno de los cuatro niveles de desempeño en la prueba ENLACE. Este es un examen nacional, llevado a cabo al final del ciclo escolar tanto en escuelas primarias como en secundarias. Para efectos de este

²⁸ Se utiliza el término escuela no en términos de su infraestructura; sino en función de cada turno escolar -al interior de un plantel- asociado a un código único de Centro de Trabajo (conocido, por su acrónimo, como CCT).

análisis, solo las asignaturas de matemáticas y español de tercer grado en el nivel de primaria son evaluadas (dado que la prueba no está disponible para grados previos).²⁹

Dos conjuntos distintos de resultados se eligieron para estimar los impactos de la intervención. La primera es la eficiencia interna (tasas de reprobación, deserción, extraedad y repitencia de los alumnos) utilizando información de la Base 911. Para esta pregunta de investigación, el primer resultado medible será la “tasa de deserción estudiantil intraciclo”, para el ciclo escolar t , donde la tasa anual de deserción para el grado k en cada ciclo escolar se define como $100 * [1 - (\text{número de estudiantes inscritos en el grado } k \text{ al } \textit{final} \text{ del ciclo escolar} / \text{dividido por el número total de estudiantes inscritos en el grado } k \text{ al inicio del ciclo escolar})]$.³⁰

El segundo resultado para medir la eficiencia es el porcentaje de alumnos con extraedad para el grado en el que están inscritos, en el ciclo escolar t . La tasa de extraedad para el grado k se define como: $100 * [\text{número de estudiantes inscritos al grado } k \text{ al final del ciclo escolar que presentan extraedad para dicho grado} / \text{dividido entre el número total de estudiantes inscritos al grado } k \text{ al final del ciclo escolar}]$.

El tercer resultado es la tasa promedio de reprobación en el ciclo escolar t . Al computar los valores de esta variable, la tasa de reprobación en el grado k para cada año se define como: $100 * [\text{número de estudiantes en el grado } k \text{ que no recibieron un promedio aprobatorio al final del ciclo escolar} / \text{dividido entre el número de estudiantes inscritos en el grado } k \text{ al final del ciclo escolar}]$. Finalmente, la tasa de repitencia en el grado k se define como: $100 * [\text{número de estudiantes repetidores en el grado } k \text{ al inicio del ciclo escolar} / \text{dividido entre el número de estudiantes inscritos en el grado } k \text{ al inicio del ciclo escolar}]$.

²⁹ La construcción de la base de datos partió del procedimiento descrito en Cárdenas (2009, pp. 82-84), añadiendo a los estadísticos allí presentados los puntajes de la prueba ENLACE.

³⁰ La información es provista a partir de la encuesta de final de ciclo; completada por el Director de cada escuela.

Adicionalmente, los resultados de las diferencias en estratificación social al interior de las escuelas de doble turno será el porcentaje de estudiantes en *Oportunidades* inscritos en cada escuela. El porcentaje de alumnos en *Oportunidades* en la escuela “j” se define como: $100 * \left[\frac{\text{número de estudiantes inscritos en la escuela “j” al final del ciclo escolar}}{\text{número total de estudiantes reportados como beneficiarios de Oportunidades en la escuela “j” al final del ciclo escolar}} \right]$.

El segundo conjunto de resultados permite una aproximación al desempeño escolar. Dado que la escala de evaluación de la prueba ENLACE tuvo cambios año con año durante el periodo de la intervención, para la medición se utilizarán niveles de desempeño, que son comparables entre sí. La medición se realizó a través del porcentaje de estudiantes en los niveles más alto y bajo de desempeño en la prueba ENLACE; en las asignaturas de español y matemáticas. El motivo para concentrarse en estos dos resultados se debe a las múltiples interpretaciones de los cambios para los niveles intermedios; que puede deberse tanto a mejoras en el aprendizaje de los alumnos en los dos últimos niveles como a un menor desempeño de aquellos en los niveles más altos.

Para llevar a cabo el análisis, se delimitó el universo de escuelas a las de doble turno. El panel de datos comprendió los ciclos escolares 2007-2008 a 2011-2012 para las variables seleccionadas³¹.

Comenzando con más de 17,000 escuelas públicas de doble turno de todo el país, incluyendo a la entidad donde ocurrió la intervención, se utilizó un mecanismo de identificación consistente en las direcciones de los planteles, para mantener una muestra de tipo panel de 14,196 centros de trabajo con 7,098 planteles para los ciclos escolares 2006-2009 a 2011-2012; siendo los últimos tres años en los que tuvo lugar la intervención en el estado tratado.

³¹ La elección de dichos años estuvo sujeta a la disponibilidad de la prueba ENLACE, cuya aplicación comenzó en el ciclo escolar 2007-2008.

La muestra incluye escuelas que han estado operando bajo el modelo de doble turno desde el ciclo escolar 2006-2007; esto es, tres años antes de que la intervención tuviera lugar. Esto controla por el hecho que una escuela de turno vespertino normalmente inicia operaciones cuando la escuela de turno matutino lleva uno o más ciclos escolares operando en ese lugar; situación que podría resultar en diferencias no observables que sesgarían los resultados (Cárdenas, 2009, *op. cit.*, p. 72).

2.4.2. Análisis descriptivo

Para evaluar las diferencias en los años previos al tratamiento, se llevó a cabo una regresión por mínimos cuadrados ordinarios (MCO) usando los datos de las escuelas en las que la intervención tuvo lugar, para evaluar las diferencias entre turnos escolares, con y sin controles, para la escuela i ; en los ciclos escolares 2007 a 2011. La especificación para cada ciclo escolar es la siguiente:

$$Y_i = \alpha_0 + \alpha_1 SHIFT_i + \delta X_i + \varepsilon_{ij} \quad (I)$$

Donde:

- i denota a cada escuela.
- Y_i representa el indicador de diferencias de los planteles entre turnos escolares, compuesto por tasas de reprobación, deserción y extraedad, así como puntajes en la prueba *ENLACE*.
- $SHIFT_i$ es una variable *dummy*, cuyo valor es cero en el turno matutino, y uno en el turno vespertino.
- X_i es un conjunto de variables de control de características observables al interior de cada escuela, incluyendo: proporción de estudiantes indígenas, proporción de estudiantes que han cursado los 3 años de preescolar; la proporción de estudiantes que pertenecen al programa

USAER, la proporción de estudiantes sobredotados, y la proporción de estudiantes con alguna discapacidad.

- $\alpha = (\alpha_0, \alpha_1, \delta)$ son los parámetros de la estimación.
- ε es el término de error.

Tanto la especificación econométrica sin controles, como la especificación con los controles arriba descritos se reportan. La Tabla 2.3.1 muestra los coeficientes estimados α_{ijk} , que se interpretan como la diferencia entre los alumnos de primer grado en los indicadores de reprobación, extraedad, deserción y repitencia en el turno vespertino, con respecto a sus contrapartes en el turno matutino. La Tabla 2.3.2, a su vez, muestra los resultados de la prueba ENLACE. Dado que la prueba fue diseñada para los alumnos de tercer grado en adelante, los efectos de la intervención pueden observarse a partir del ciclo escolar 2011-2012.

Los efectos encontrados indican mayores tasas de reprobación, extraedad, deserción y repitencia en el turno vespertino en los planteles de la entidad que llevó a cabo la aleatorización. Además, estas escuelas tienen un porcentaje mayor de estudiantes en el nivel más bajo de desempeño -y un porcentaje menor en el nivel más alto- comparadas con aquellas en el turno matutino.

A partir de este análisis no se observan diferencias de magnitud relevante entre los años previos a la lotería y los años posteriores; con la excepción de una ligera disminución en la magnitud del coeficiente promedio para la brecha en reprobación y el nivel más bajo de desempeño en la prueba ENLACE de español.

Tabla 2.3.1. Regresión MCO de la diferencia entre turnos en escuelas primarias en el grupo de tratamiento
(primer grado, ciclo escolar 2007-2008 a 2011-2012)

Variable	2007		2008		2009		2010		2011	
	Sin controles	Con controles								
<i>Oportunidades</i>	5.21*** (1.38)	4.50*** (1.72)	5.13*** (1.40)	2.73 (1.83)	4.36*** (1.21)	3.34** (1.59)	4.61*** (1.27)	3.91** (1.57)	4.14*** (1.33)	4.07*** (1.53)
Reprobación	1.89** (0.88)	2.11* (1.17)	2.45** (1.03)	1.16 (1.68)	1.39* (0.74)	1.57 (0.97)	4.02*** (0.81)	2.02 (1.06)	2.94*** (0.87)	2.45** (1.05)
Extraedad	13.70*** (1.49)	13.36*** (1.96)	13.67*** (1.58)	10.23*** (2.29)	12.30*** (1.43)	9.09*** (1.76)	13.20*** (1.29)	10.79** (1.66)	14.52*** (1.25)	11.59*** (1.49)
Deserción	11.00*** (1.36)	8.76*** (1.80)	13.22*** (1.35)	13.66*** (2.14)	13.38*** (1.42)	10.63*** (1.83)	12.98*** (1.35)	12.25*** (1.90)	12.90*** (1.06)	11.25*** (1.39)
Repitencia	8.79*** (1.23)	7.91*** (1.63)	8.15*** (1.37)	5.20* (3.17)	5.11** (2.05)	5.11*** (1.12)	7.13*** (1.12)	5.91*** (1.39)	8.95*** (0.98)	6.99*** (1.09)

Notas: entre paréntesis se presentan los errores estándar. Códigos de significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01.
Fuente: cálculos propios con base en la Estadística 911 SEP.

Tabla 2.3.2. Regresión MCO de la diferencia entre turnos en escuelas primarias en el grupo de tratamiento
(tercer grado, ciclo escolar 2011-2012)

Variable	2007		2008		2009		2010		2011	
	Sin controles	Con Controles	Sin controles	Con Controles	Sin controles	Con Controles	Sin controles	Con Controles	Sin controles	Con Controles
Prueba <i>ENLACE</i> , español nivel 0	9.59*** (2.03)	4.50*** (1.72)	7.40*** (2.03)	7.00** (3.02)	2.12 (1.86)	3.13 (2.48)	5.98*** (1.77)	4.31* (2.50)	2.48* (1.51)	3.35* (1.91)
Prueba <i>ENLACE</i> , español nivel 3	-1.92*** (0.45)	-0.42 (0.56)	-1.61** (0.63)	-1.41 (0.97)	-4.32*** (0.93)	-3.84** (1.35)	-4.39*** (1.16)	-2.00 (1.69)	-3.96*** (1.36)	-6.69 (4.08)
Prueba <i>ENLACE</i> , matemáticas nivel 0	10.63*** (2.10)	6.49*** (2.68)	6.98** (2.36)	5.11 (3.51)	5.21*** (1.76)	5.70** (2.30)	4.34** (2.09)	2.14 (2.93)	4.92*** (1.80)	5.96*** (2.24)
Prueba <i>ENLACE</i> , matemáticas nivel 3	-2.34*** (0.74)	0.27 (0.89)	-3.18*** (0.85)	-2.24* (1.30)	-4.14*** (1.01)	-3.96*** (1.44)	-5.06*** (1.35)	-2.37 (1.93)	-4.68*** (1.58)	-3.56** (1.84)

Notas: entre paréntesis se presentan los errores estándar. Códigos de significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01.
Fuente: cálculos propios con base en la Estadística 911 SEP.

El Anexo 2A muestra el grado de variabilidad entre el estado que compone el tratamiento y el resto de los estados en el país. Estos efectos placebo estadísticamente significativos podrían deberse a sesgos específicos que existen respecto a los indicadores estudiados. La debilidad de estos resultados radica en el hecho que las escuelas en el grupo de tratamiento se contrastan con un grupo heterogéneo de escuelas de todo el país. La probabilidad de que variables no observables afecten las tendencias entre ambos grupos es alta.

Para contar con un grupo de control más comparable y estimar los efectos con precisión, se presenta y describe un diseño de diferencias en diferencias combinado con aparejamiento en la siguiente sección.

2.4.3. Metodología de aparejamiento por marcador de propensión

La aproximación al estudio de esta política pública se lleva a cabo mediante un diseño cuasiexperimental. A continuación, se presenta la metodología de aparejamiento y diferencias en diferencias para estimar el efecto que tuvo la lotería para incidir sobre los resultados de desempeño de los alumnos. Para este propósito, se emplea información del resto de los estados para construir un grupo de comparación. El método de aparejamiento se utiliza para seleccionar un grupo de control que se parezca lo más posible a la población intervenida que estamos estudiando, en ausencia de una aleatorización preintervención. El aparejamiento es una forma de minimizar las diferencias entre las escuelas tratadas y las no tratadas (también llamadas grupo de comparación) previo a su exposición al tratamiento; en este caso la lotería. Lo que necesitamos es definir un conjunto de variables condicionantes W sobre las cuales elegir un grupo de comparación -similar en estas características- al grupo que fue intervenido. Es importante destacar que la elección del grupo comparable necesita ser independiente del estatus del tratamiento:

$$Y_0 \perp\!\!\!\perp D | W \quad (\text{III.1})$$

Donde el signo $\perp\!\!\!\perp$ denota independencia. Asumiendo que existe una probabilidad entre cero y uno de participar, condicionamos dicho tratamiento a las características de las escuelas:

$$0 < Pr(D = 1|W) < 1 \quad (\text{III.2})$$

Con ello es posible comparar las escuelas en el grupo de tratamiento con escuelas que tengan características similares, así como una probabilidad similar de ser seleccionadas en la lotería. El objetivo es encontrar un grupo de escuelas que se parezcan a las escuelas del estado donde ocurrió la intervención en términos de las características W . Dado que el aparejamiento en un conjunto grande de características puede ser muy difícil (especialmente si dichas características son continuas), este capítulo parte de la metodología desarrollada en Rosenbaum y Rubin (1983) que demuestra -partiendo del supuesto de que sea válido aparejar sobre un conjunto W de características- que es equivalente aparejar usando las características W directamente que aparejar sobre el puntaje del marcador de propensión $Pr(D = 1|W)$. Este es el principio sobre el que se basa lo que se conoce como aparejamiento por marcador de propensión (*propensity score matching*).

A partir de esta metodología, sea Y_1 el resultado para una escuela que fue expuesta a la lotería, y Y_0 el resultado si no hubiese sido expuesta. En principio el impacto del programa (o efecto del tratamiento) para esa escuela sería la diferencia entre estos dos valores, o $I = Y_1 - Y_0$. No obstante, no es posible observar a la misma escuela al mismo tiempo en condiciones normales de operación y bajo la asignación aleatoria. El enfoque que se sigue para aproximar esta diferencia es estimar la diferencia de las diferencias (DID) tras la intervención para estimar el efecto de la política.

Sea t' un periodo previo a la intervención, y t el periodo posterior a la intervención, el estimador DID está dado por

$$DID = E(Y_{1t} - Y_{1t'} | D = 1) - E(Y_{0t} - Y_{0t'} | D = 0) \quad (III.3)$$

donde D denota la exposición al tratamiento, siendo 1 cuando las escuelas han sido expuestas y 0 cuando no lo han sido, y E es el promedio de cada diferencia. Para que el estimador DID sea un estimador insesgado del impacto del programa, el supuesto necesario es que las escuelas que no están en el grupo de tratamiento se comporten como las escuelas tratadas, en ausencia del tratamiento. Una forma de expresar esto es:

$$E(Y_{0t} - Y_{0t'} | D = 1) = E(Y_{0t} - Y_{0t'} | D = 0) \quad (III.4)$$

un supuesto que resulta restrictivo, dado que no existe evidencia de que las escuelas no tratadas sean comparables con las escuelas que recibieron la intervención. De hecho, como se muestra en las Gráficas 2.1a-i, las características de las escuelas comparadas -previo a la aleatorización- difieren de las características de las escuelas en el estado donde se llevó a cabo la lotería.

Una mirada a las gráficas de los patrones pre-intervención en el ciclo escolar 2008-2009 revela qué tan diferentes son estas escuelas del resto de escuelas que sirven de comparación, para: la proporción de estudiantes en el programa Oportunidades (2.1a), el porcentaje de reprobación (2.1b), extraedad (2.1c), deserción (2.1d) y repitencia (2.1e). Así como el porcentaje de estudiantes en el nivel más bajo (2.1f) y más alto (2.1h) de la prueba Enlace de español; y el porcentaje en el nivel más bajo (2.1g) y más alto (2.1i) de matemáticas en dicha prueba.

Así, el aparejamiento de diferencias en diferencias (DIDM) usado para evaluar las diferencias parte del siguiente supuesto:

$$E(Y_{0t} - Y_{0t'} | P(X), D = 1) = E(Y_{0t} - Y_{0t'} | P(X), D = 0) \quad (III.5)$$

Esto es, que el valor promedio de las diferencias antes y después del programa es similar, condicionado con respecto al marcador de propensión, para el grupo de tratamiento y el de comparación; si no hubiese sido tratado. Este supuesto es menos restrictivo que el supuesto III.4;

ya que previamente se controló que las características entre los grupos de tratamiento y control -antes de la aplicación del programa- fuesen más parecidas entre los grupos. Una perspectiva distinta desde la cual ver este supuesto, es asumir que la tendencia que siguen las escuelas en el grupo de comparación sería la misma que la de aquellas en el grupo de tratamiento; en ausencia de la lotería.

2.4.3.2. Procedimiento para el aparejamiento

A pesar del relativamente bajo número de centros de trabajo que siguieron la política de selección aleatoria (en comparación con el total de planteles del país), un grupo de comparación suficientemente grande está disponible al utilizar la muestra completa de escuelas a nivel nacional. La Tabla 2.4 muestra la distribución de escuelas disponible en el panel de datos, dividida por turno.

Tabla 2.4. Escuelas en los grupos de tratamiento y control, por turno

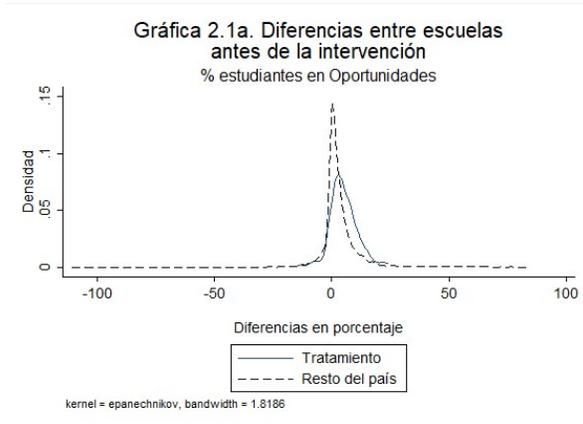
Grupo	Turno matutino	Turno vespertino	%
Control	7,014	7,014	98.82
Tratamiento	84	84	1.18

Nota: cálculos propios.

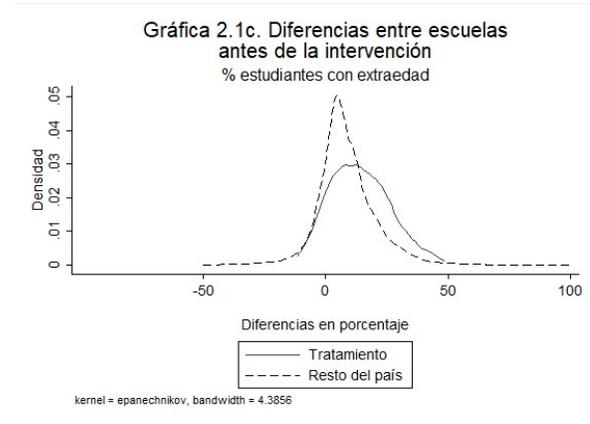
El Anexo 2B muestra la lista de variables condicionantes seleccionadas como determinantes clave de la participación, incluyendo los resultados estudiados y los puntajes. La Tabla 2.5, a su vez, muestra los resultados de estimar el marcador de propensión, utilizando un modelo logístico binomial.

11 de las 24 variables son estadísticamente significativas para predecir la asignación al grupo de tratamiento, al menos a un nivel de significancia del 10%. Las estimaciones muestran que las variables que incrementan la probabilidad de participar en el tratamiento son: el número total de directivos en la escuela, el porcentaje de estudiantes de primer grado que no terminaron el

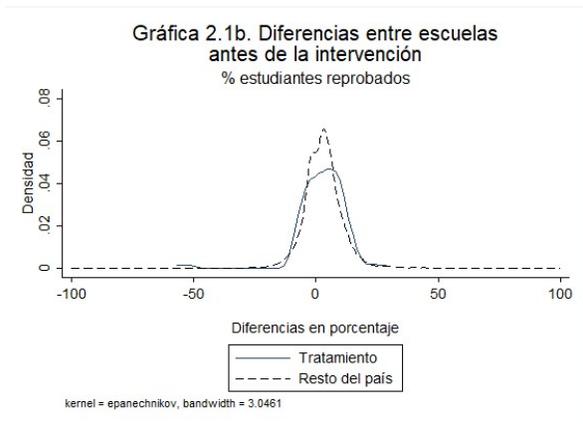
Gráfica 2.1a-i. Distribución de las diferencias entre turnos; resultados seleccionados. Escuelas primarias.



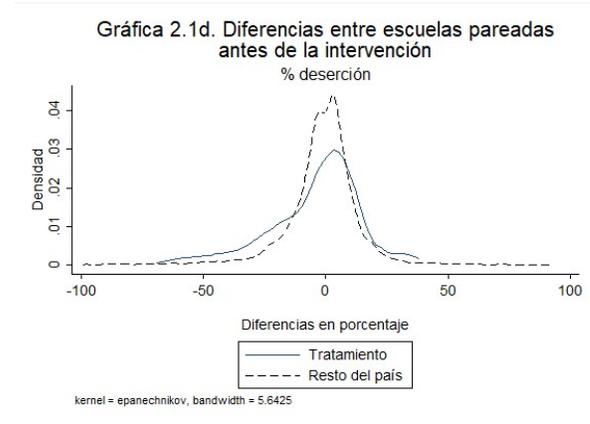
Fuente: cálculos propios.



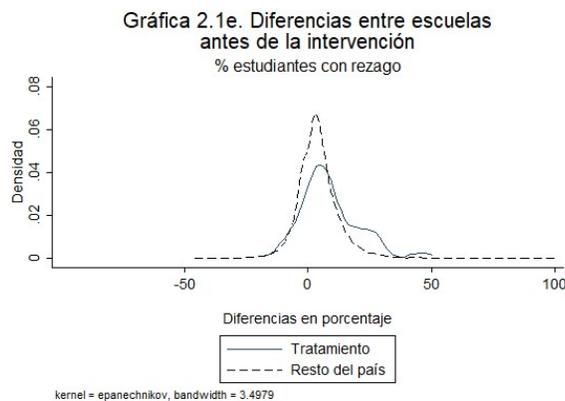
Fuente: cálculos propios.



Fuente: cálculos propios.

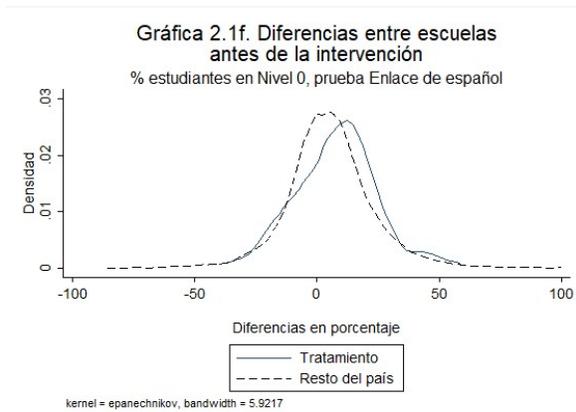


Fuente: cálculos propios.

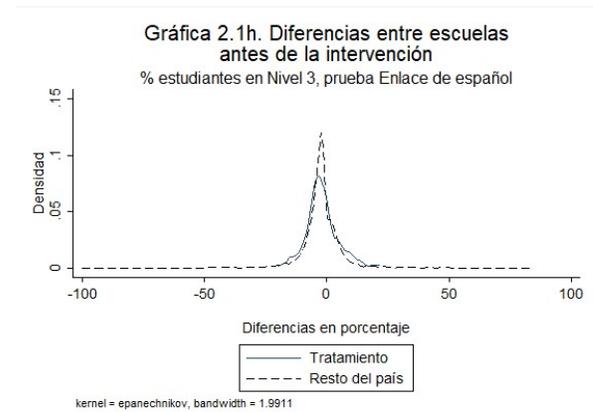


Fuente: cálculos propios.

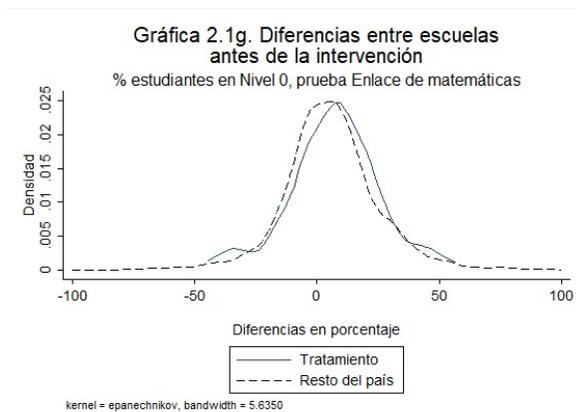
Gráfica 2.1a-i. Distribución de las diferencias entre turnos; resultados seleccionados. Escuelas primarias (Cont.)



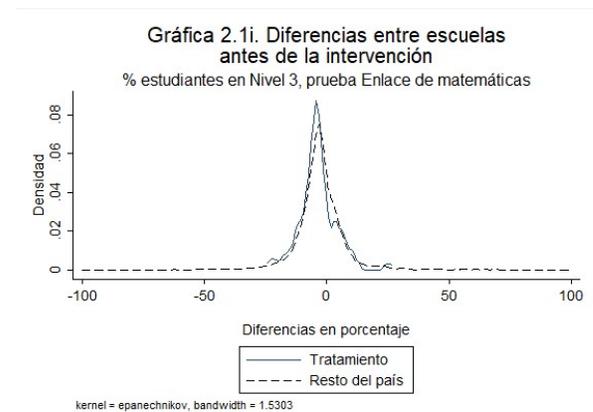
Fuente: cálculos propios.



Fuente: cálculos propios.



Fuente: cálculos propios.



Fuente: cálculos propios.

Tabla 2.5. Modelo logístico de participación en el tratamiento
D = 1 Elegible para la asignación aleatoria (Tratamiento), D = 0 No elegible para la asignación aleatoria (Resto del país)

Variable	Escuelas primarias		
	Coefficiente ¹	Error estándar ¹	Estadístico T ²
Número total de directivos en la escuela	3.626	1.577	2.30**
Número total de administrativos en la escuela	-0.277	0.139	-1.99**
Total de salones de clase disponibles	-0.250	0.070	-3.60***
Porcentaje de estudiantes que no terminaron el año al que asistieron; primer grado	0.132	0.014	9.23***
Inscripción a primer grado	-0.030	0.022	-1.35
Inscripción total a la escuela	-0.003	0.005	-0.54
Gasto total anual en cuotas escolares	-0.005	0.002	-3.21***
Número total de grupos de primer grado en la escuela	1.137	0.534	2.13**
Número total de profesores en el Programa <i>Carrera Magisterial</i>	0.044	0.052	0.86
<i>Dummy</i> para escuela multigrado (0=regular, 1=multigrado)	-0.299	0.267	-1.12
Porcentaje de estudiantes en nivel 0 en la prueba ENLACE de español	-0.007	0.017	-0.39
Porcentaje de estudiantes en nivel 0 en la prueba ENLACE de matemáticas	0.010	0.015	0.70
Porcentaje de estudiantes en nivel 3 en la prueba ENLACE de español	-0.005	0.032	-0.16
Porcentaje de estudiantes en nivel 3 en la prueba ENLACE de matemáticas	-0.014	0.030	-0.49
Porcentaje de estudiantes con algún grado de extraedad; primer grado	-0.059	0.020	-2.96***

Notas: códigos de significancia: *P<0.1, **P<0.05, ***P<0.01. Fuente: cálculos propios.

Tabla 2.5. Modelo logístico de participación en el tratamiento (Cont.)

Variable	Escuelas primarias		
	Coficiente	Error estándar	Estadístico T
Porcentaje de estudiantes que han repetido al menos un grado.	0.116	0.021	5.47***
Porcentaje de estudiantes que no aprobaron el grado al que asisten	-0.018	0.025	-0.71
Número total de profesores frente a grupo	0.497	0.248	2.01**
Número total de grupos en la escuela	-0.266	0.277	-0.96
Porcentaje de alumnos en el programa <i>Oportunidades</i> en la escuela	2.798	1.151	2.43**
Porcentaje de alumnos que cursaron los 3 años de preescolar	-3.108	2.693	-1.15
Variable dicotómica para turno escolar al que se asiste (0=mañana, 1=tarde)	-0.804	0.348	-2.31**

Notas: códigos de significancia: *P<0.1, **P<0.05, ***P<0.01. Fuente: cálculos propios.

año al que asistieron, el número total de grupos de primer grado, el porcentaje de estudiantes que han repetido al menos un grado, el número total de profesores frente a grupo y el porcentaje de alumnos pertenecientes al programa *Oportunidades*.

Por el contrario, las variables que disminuyen la probabilidad de participación son: el número total de administrativos en la escuela, el total de salones de clase disponibles, el gasto total anual en cuotas escolares, el porcentaje de estudiantes con extraedad en primer grado; y pertenecer al turno vespertino.

Para construir un grupo comparable al de tratamiento, el primer paso fue utilizar aparejamiento mediante la técnica de vecinos más cercanos en términos del marcador de propensión estimado, para asignarle escuelas no tratadas del grupo de comparación a las escuelas tratadas. Se utilizó

un aparejamiento con 3 vecinos más cercanos³². Una vez que se determinaron los vecinos más cercanos, se seleccionaron ambas escuelas (matutino y vespertino) de cada plantel.

Se construyó, para cada escuela tratada, un contrafactual definido como el promedio de las diferencias matutino-vespertino de las escuelas aparejadas. Los vecinos se hallaron para cada una de las 168 escuelas de la entidad estudiada.

Tabla 2.6. Escuelas en los grupos pareados de tratamiento y control, por turno

	Tratamiento	Control (NN3)
Turno matutino	84	454
Turno vespertino	84	454

Fuente: cálculos propios.

Las estimaciones se realizaron a través de modelos de tres vecinos. Se utilizaron los planteles con los puntajes más cercanos a cada escuela en el grupo de tratamiento, para constituir las contrapartes de las escuelas aparejadas. Con las muestras aparejadas, se calcularon las diferencias entre los turnos matutino y vespertino para las nueve variables de interés.

Para el cálculo de las diferencias por año se promediaron las diferencias matutino-vespertino de las escuelas aparejadas a cada escuela en el grupo de tratamiento, siendo los efectos promedio de estos promedios las diferencias por año en el grupo de comparación. Los errores estándar se calcularon utilizando un procedimiento bootstrap con 1000 repeticiones. El bootstrap se utiliza para calcular la distribución empírica de un estimador a partir de un conjunto de repeticiones de estadísticos obtenidos a partir de muestras aleatorias de la base de datos original.

³² Adicional a esta estimación, se realizó un aparejamiento con 5 vecinos, y una estimación de 3 y 5 vecinos imputando valores de cero en los valores perdidos. En todos los casos los resultados siguieron el patrón que se describe en este capítulo.

Un defecto de los datos es la presencia de múltiples valores perdidos concentrados en muy pocas variables relevantes; lo que provoca, no obstante, que los marcadores de propensión no se puedan calcular.

En el anexo 2C se presenta un procedimiento similar, pero llevando a cabo una imputación de ceros en las variables seleccionadas para el cálculo del marcador de propensión (que se describe con mayor detalle en dicho apartado); a fin de perder la menor cantidad de información relevante en el cálculo del estimador. Los impactos presentados, no obstante, utilizaron únicamente los datos originales reportados en las distintas bases de datos.

La Tabla 2.7 muestra que -con la muestra aparejada y usando los datos de las escuelas y sus pares del resto del país- no existen diferencias estadísticamente significativas en ninguna de las nueve variables estudiadas, a un nivel de significancia del 5%, salvo en una variable. La variable de deserción intraciclo (2.3d y 2.4d) tiene una diferencia estadísticamente significativa al 10%. Los resultados de la Tabla 2.7 nos permiten afirmar que es posible evaluar el efecto promedio del tratamiento sobre los tratados para escuelas bajo la lotería obligatoria comparando los periodos pre y post intervención. En el caso del modelo con imputación de ceros del Anexo 2C (Tabla 2C.1), ninguna variable resulta significativa al 5%.

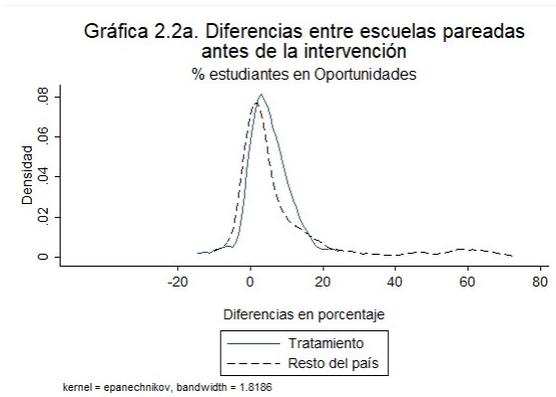
Las Gráficas 2.2a-i muestran la distribución de las diferencias entre las escuelas en el grupo de tratamiento y las escuelas obtenidas mediante esta metodología, complementando la evidencia presentada en la Tabla 2.8. Las gráficas 2C.1a-i del Anexo 2C presenta las tendencias para el caso del modelo NN3 con imputación de ceros.

Tabla 2.7. Diferencias en las medias entre el grupo de tratamiento y el grupo de control, obtenido mediante el apareamiento, previo al tratamiento

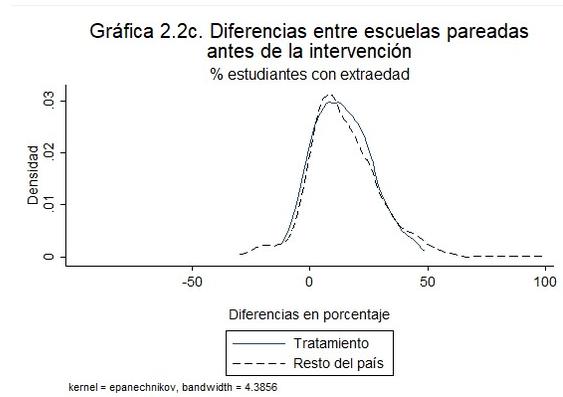
Variable	Tratamiento ¹	NN3		
		Control	Diferencia	Estadístico T
<i>Oportunidades</i>	5.13 (0.44)	6.14 (0.76)	-1.01 (0.87)	-1.15 [0.13]
Reprobación, primer grado	2.45 (0.69)	2.57 (0.40)	-0.12 (0.81)	-0.14 [0.44]
Extraedad, primer grado	13.67 (0.91)	14.24 (0.77)	-0.57 (1.19)	-0.48 [0.32]
Deserción, primer grado	-10.40 (2.82)	-5.48 (1.66)	-4.93 (3.29)	-1.50* [0.07]
Repitencia, primer grado	8.15 (0.88)	8.48 (0.76)	-0.33 (1.17)	0.28 [0.39]
<i>ENLACE</i> español, nivel 0	7.76 (1.26)	7.74 (1.03)	0.01 (1.63)	0.01 [0.50]
<i>ENLACE</i> español, nivel 3	-1.66 (0.46)	-2.33 (0.43)	0.67 (0.63)	1.07 [0.14]
<i>ENLACE</i> matemáticas, nivel 0	7.26 (1.40)	7.19 (1.02)	0.07 (1.73)	0.04 [0.48]
<i>ENLACE</i> matemáticas, nivel 3	-3.23 (0.56)	-4.36 (0.39)	1.12 (0.68)	1.65* [0.05]

Nota: entre paréntesis se presentan las desviaciones estándar, y entre corchetes la significancia de los estadísticos t. Fuente: cálculos propios.

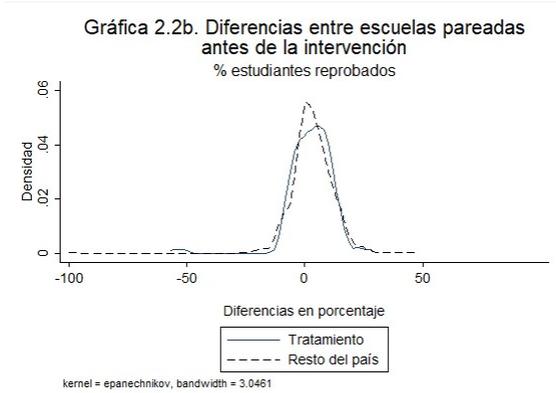
Gráfica 2.2a-i. Distribución de las diferencias entre escuelas pareadas (NN3); variables de interés. Escuelas primarias.



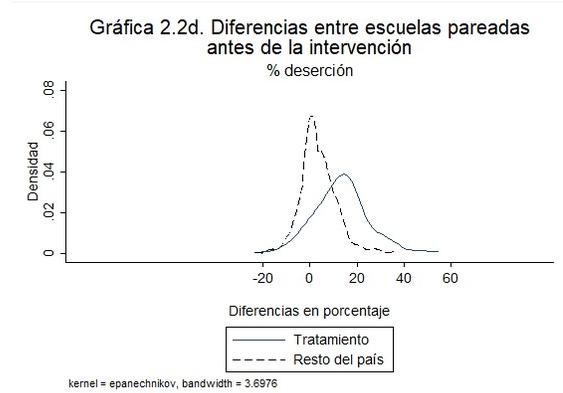
Fuente: cálculos propios.



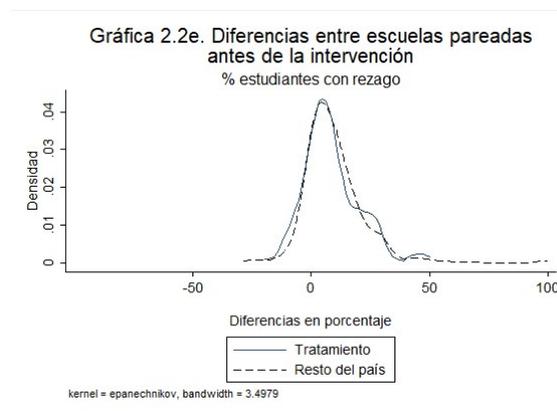
Fuente: cálculos propios.



Fuente: cálculos propios.

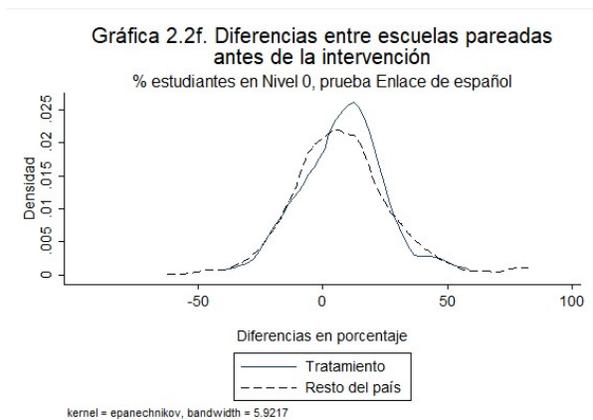


Fuente: cálculos propios.

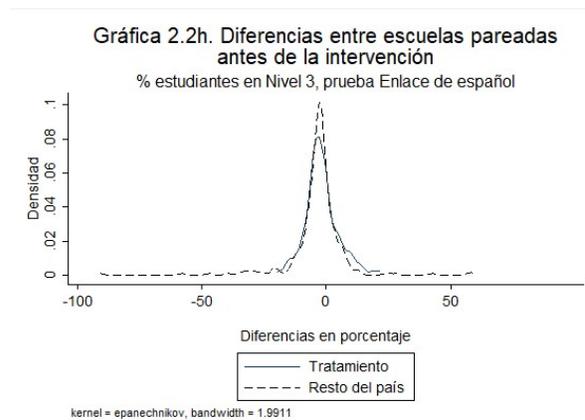


Fuente: cálculos propios.

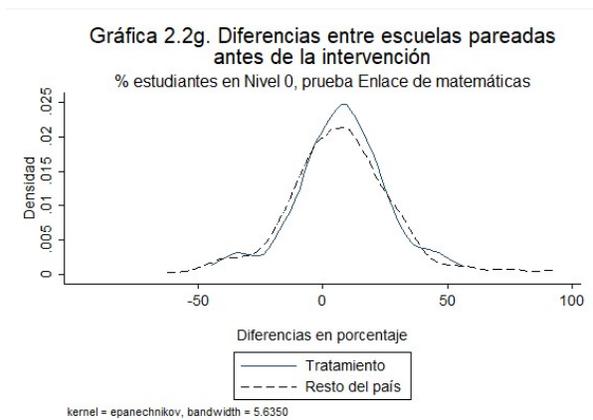
Gráfica 2.2a-i. Distribución de las diferencias entre escuelas pareadas (NN3); variables de interés. Escuelas primarias (Cont.)



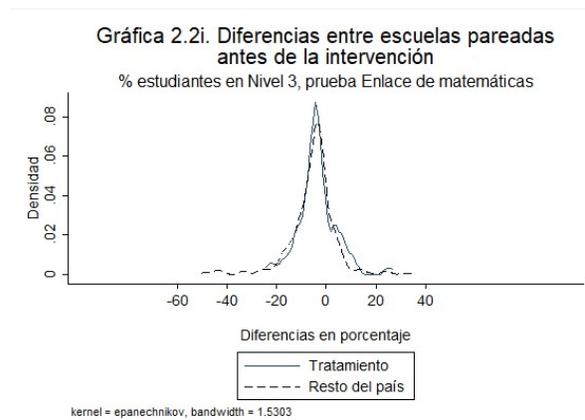
Fuente: cálculos propios.



Fuente: cálculos propios.



Fuente: cálculos propios.



Fuente: cálculos propios.

Otra forma de constatar si el grupo de tratamiento y el de control son similares en sus características, es mirando las tendencias antes y después de que la intervención tuvo lugar. Las gráficas 2.3f-i muestran diferencias en las tendencias pre y post intervención, utilizando los resultados de la prueba ENLACE en el nivel más bajo de español (2.3f) y matemáticas (2.3h); así como los correspondientes cambios en sus niveles más altos (2.3g y 2.3i). Las gráficas 2.3a-e presentan las tendencias del resto de los resultados de interés; y las gráficas 2C.2a-i del Anexo 2C lo hacen para el caso del modelo con imputación de ceros.

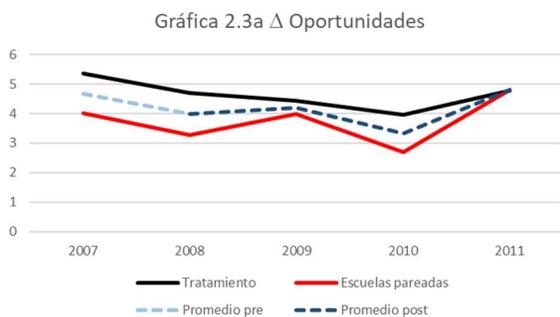
2.5. Resultados del diseño cuasiexperimental

Con el conjunto de datos panel del grupo de tratamiento y el de comparación, obtenidos a través de la técnica descrita para los ciclos escolares 2008-2009 a 2011-2012, las Tablas 2.9.1 a 2.10.2 muestran el efecto obtenido mediante diferencias en diferencias; comparando cada uno de los años posteriores a la intervención con el periodo anterior a la misma, sobre las variables de interés.

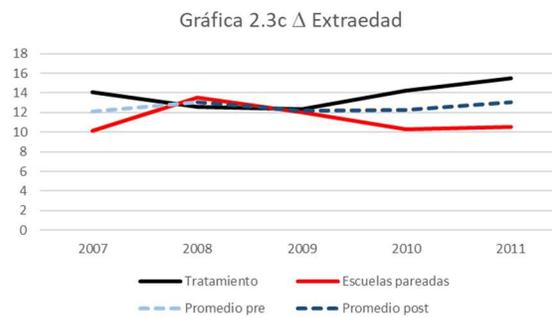
Como muestra la Tabla 2.9.1 (NN3), hay un incremento en las brechas de extraedad y repitencia para las cohortes de estudiantes de primer grado en los ciclos escolares 2009 a 2011. Estos efectos oscilan entre el 2 y el 7% de incremento en la brecha, en el caso de extraedad, y entre el 2.5 y el 6%, en el caso de repitencia. Incluso, la brecha de reprobación se incrementa para la cohorte 2010 en 1.91%.

El cálculo de los errores estándar mediante técnica bootstrap se utilizó para calcular la media del error estándar a partir de un número suficientemente grande (1000) de repeticiones de indicadores obtenidos a partir de muestras aleatorias de la base de datos original. Esto permite, mediante ley de grandes números, aproximar a la media del valor correcto del error, minimizando el sesgo.

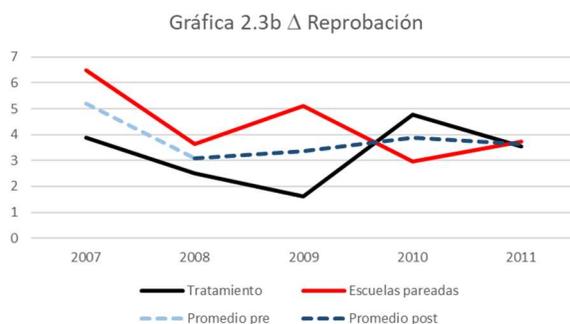
Gráfica 2.3a-i. Diferencias en las tendencias entre grupos;
cambios en resultados seleccionados (NN3). Escuelas primarias.



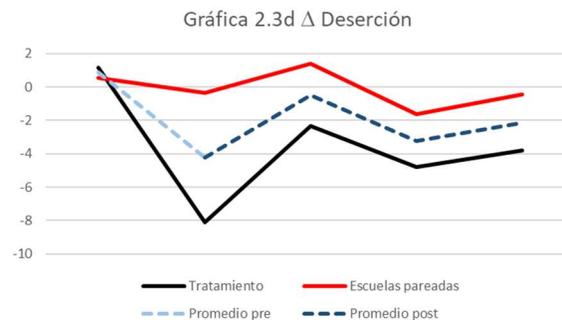
Fuente: cálculos propios.



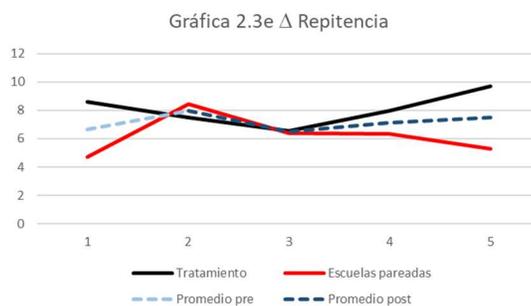
Fuente: cálculos propios.



Fuente: cálculos propios.

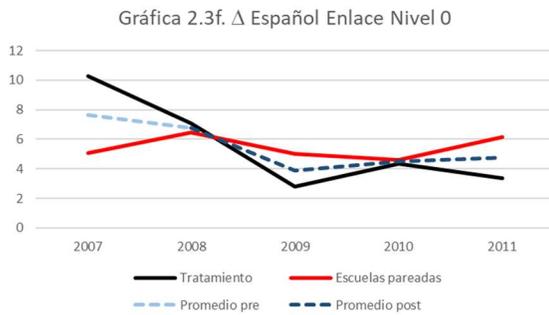


Fuente: cálculos propios.

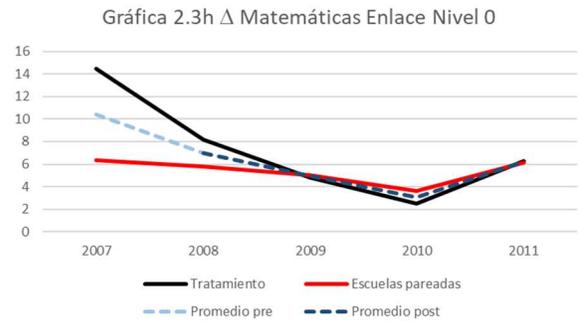


Fuente: cálculos propios.

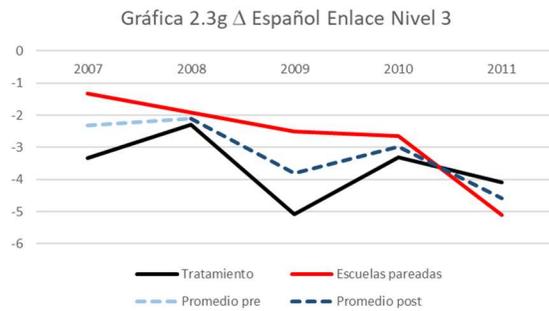
Gráfica 2.3a-i. Diferencias en las tendencias entre grupos; cambios en resultados seleccionados (NN3). Escuelas primarias (Cont.)



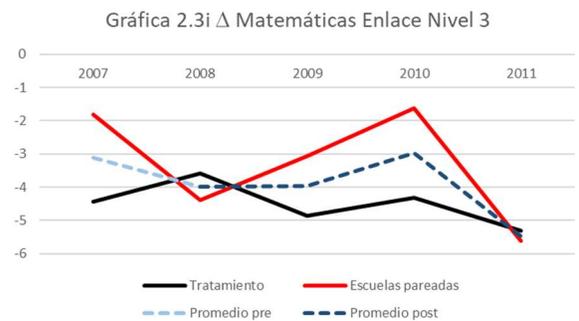
Fuente: cálculos propios.



Fuente: cálculos propios.



Fuente: cálculos propios.



Fuente: cálculos propios.

Los errores estándar se estimaron usando 1,000 repeticiones bootstrap. En cada repetición se generaron muestras aparejadas del mismo tamaño que la muestra original, con un remuestreo aleatorio con reemplazo. La ventaja de utilizar la técnica de cálculo del error estándar de esta manera es que no requiere supuestos respecto a la forma de la distribución de los no observables de los grupos. La desviación estándar calculada a través de los 1000 valores replicados es una buena aproximación del error estándar de la diferencia en diferencias estimada. La Tabla 2.9.2 (NN3), a su vez, muestra los resultados correspondientes a la prueba ENLACE. No se presentan efectos significativos para el nivel más alto de desempeño; así como tampoco para la prueba de Matemáticas. Los resultados no difieren mucho al imputar ceros en las variables estudiadas (tablas 2C.2.1 y 2C.2.2).

2.5.1. Limitaciones

Existen varias limitaciones que deben señalarse en el presente estudio. Primero, la falta de verificación para asegurarse que la lotería se haya llevado a cabo por todas las escuelas, como mandató la ley, podría resultar en una subestimación del impacto. En segundo lugar, el hecho de que los directivos hayan podido distorsionar la asignación tras el primer año del tratamiento -por ejemplo, enviando alumnos del turno vespertino al matutino- también afectaría el impacto estimado de la intervención.

Tercero, la falta de un grupo de control apropiado designado *a priori* impone restricciones. Dada la naturaleza cuasiexperimental de los métodos de evaluación de impacto aquí utilizados -diferencias en diferencias y aparejamiento- es importante notar que los resultados dependen crucialmente de la validez de los supuestos.

Al utilizar diferencias en diferencias, debe recordarse que este método utiliza el supuesto de que no existen variaciones en las tendencias en el tiempo. Como se observa en Khandker, Koolwal y Samad (2010) al implementar esta metodología “la heterogeneidad no observada, variable en

el tiempo, podría llevar a un sesgo de magnitud creciente o decreciente” (P. 77, traducción propia).

Tabla 2.8.1. Impactos estimados utilizando diferencias en diferencias (NN3); resultados seleccionados en escuelas primarias pareadas

Variables	2008		2009		2010		2011	
	promedio		promedio	Efecto DID	promedio	Efecto DID	promedio	Efecto DID
<i>Oportunidades</i>	-1.01 (0.87)		-1.34** (0.74)	-0.33 (1.49)	-0.88 (0.83)	0.12 (1.48)	-0.98 (0.63)	0.03 (1.37)
Reprobación, primer grado	-0.12 (0.81)		-1.83*** (0.63)	-1.71* (0.94)	1.79*** (0.64)	1.91** (0.91)	0.98* (0.64)	1.10 (0.87)
Reprobación, segundo grado	0.33 (0.66)				0.79 (0.66)	0.46 (0.83)	0.67 (0.63)	0.34 (0.56)
Reprobación, tercer grado	-0.12 (0.54)						0.50 (0.59)	0.62 (0.71)
Extraedad, primer grado	-0.57 (1.19)		2.03** (1.05)	2.61 (1.64)	4.03*** (0.96)	4.61*** (1.64)	6.39*** (1.05)	6.96*** (1.61)
Extraedad, segundo grado	3.49*** (1.24)				1.00 (1.05)	-2.49 (1.84)	5.19*** (1.11)	1.70 (1.86)
Extraedad, tercer grado	4.84*** (1.28)						2.19 (1.13)	-2.65 (1.97)
Deserción, primer grado	-4.93* (3.29)		-1.58 (1.35)	3.34 (2.74)	-9.48*** (3.28)	-4.55 (8.19)	3.86 (3.12)	1.07 (3.02)
Deserción, segundo grado	-0.08 (1.49)				1.23 (1.07)	1.31 (3.44)	0.10 (1.19)	0.18 (1.59)
Deserción, tercer grado	2.11* (1.58)						-0.93 (1.48)	-3.04* (1.72)
Repitencia, primer grado	-0.33 (1.17)		2.66*** (0.64)	2.99** (1.23)	2.99*** (0.77)	3.32*** (1.26)	5.39*** (0.81)	5.72*** (1.26)
Repitencia, segundo grado	1.52** (0.70)				-0.63 (0.54)	-2.15** (0.95)	2.29*** (0.73)	0.77 (0.93)
Repitencia, tercer grado	2.73*** (0.74)						1.81*** (0.52)	-0.92 (0.75)

Notas: entre paréntesis se presentan los errores estándar.

Códigos de significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01.

Fuente: cálculos propios con base en la Estadística 911 SEP.

Tabla 2.8.2. Impactos estimados utilizando diferencias en diferencias (NN3); resultados seleccionados en escuelas primarias pareadas

Variables	2008	2011	
	promedio	promedio	Efecto DID
<i>ENLACE</i> español, nivel 0	0.01 (1.63)	-1.75* (1.13)	-1.77 (2.19)
<i>ENLACE</i> español, nivel 3	0.67 (0.63)	0.64 (1.23)	0.66 (1.72)
<i>ENLACE</i> matemáticas, nivel 0	0.07 (1.73)	0.73 (1.48)	-0.04 (2.54)
<i>ENLACE</i> matemáticas, nivel 3	1.12 (0.68)	-0.36 (1.38)	-1.48 (1.79)

Notas: entre paréntesis se presentan los errores estándar.

Códigos de significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01.

Fuente: cálculos propios con base en la Estadística 911 SEP.

Si los cambios en cualquiera de los resultados entre un año escolar y el siguiente se debieran a factores no observables, y esos factores difiriesen entre las escuelas que recibieron la lotería y el grupo de comparación, ello podría conducir a una fuente potencial de error en la estimación. Pese a ello, en este contexto la disponibilidad de la información para las características pre-intervención, así como las del ciclo escolar previo, ayudaron a controlar las diferencias en tendencias entre las escuelas tratadas y el grupo de comparación.³³

Respecto al aparejamiento, la principal restricción es que las características no observadas no deben diferir entre los grupos de tratamiento y control. Si las hubiere, y dichas diferencias incidiesen sobre el resultado, entonces el efecto estimado podría sufrir de un sesgo (Gertler, Martinez *et. al.*, 2011; p.107).

Adicionalmente, los impactos estimados resultaron particularmente sensibles a la elección de variables utilizadas para el cálculo del marcador de propensión. En el modelo que únicamente contenía los nueve indicadores de interés en niveles y diferencias, por ejemplo, se obtuvo una

³³ Pese a ello, cabe recalcar que estas diferencias no podrían dar cuenta de cambios en la tendencia que afectasen solo al grupo de tratamiento. Por ejemplo, producto de políticas locales (que, hasta donde se tuvo conocimiento, no ocurrieron) que afectasen dicho tratamiento.

línea base sin diferencias estadísticamente significativas entre tratados y no tratados; pero impactos nulos sobre desempeño en la prueba ENLACE. En el extremo opuesto, bastó con introducir seis de las once variables significativas en el modelo logit para encontrar impactos significativos y de gran magnitud, pero a costa de una línea base muy distinta entre escuelas tratadas y elegidas para el grupo de comparación.

Al elegir el grupo de comparación adecuado, el tamaño y la afinidad con la muestra relativamente pequeña de escuelas tratadas fue crucial. Como se mostró, utilizar el promedio para las escuelas en el resto del país resultó equivocado, dada la enorme diferencia en las características y tendencias de éstas. El grupo de comparación obtenido a través apareamiento mostró tener una comparabilidad superior.

2.6. Conclusión

El problema de desigualdades entre turnos escolares (que normalmente surgen como producto de la sobrepoblación en el turno matutino) afecta a millones de niños en México; así que los beneficios potenciales de una política que mitigue esas desigualdades resultan sustantivos. Al mismo tiempo, el tamaño de la población potencialmente involucrada hace obligatorio proceder con cautela; siendo indispensable contar con evidencia del impacto de cualquier intervención relacionada.

Este capítulo dio cuenta de una política, llevada a cabo por las autoridades de uno de los 32 estados de la república mexicana, en la cual las nuevas cohortes de escuelas primarias públicas fueron sujetas a una asignación de turnos escolares distinta a la tradicional. El estudio de caso, enriquecido por la información pública disponible para las escuelas, permitió estimar efectos de la aleatorización entre turnos; midiendo la reducción en la brecha de resultados educativos con esta, más equitativa, distribución.

Para la identificación del efecto causal de la política, en este trabajo se emplea un modelo de aparejamiento combinado con diferencias en diferencias y utiliza escuelas de doble turno -en el estado donde la política era obligatoria- como el grupo de tratamiento; y una submuestra del resto de las escuelas del país (obtenida a través del método de vecino más cercano) como grupo de comparación.

Los resultados muestran una reducción en las brechas de reprobación entre los turnos matutino y vespertino el primer año de la implementación; pero posteriormente esta tendencia se revierte; incrementándose. Se incrementan asimismo las brechas de extraedad y repitencia.

La prueba ENLACE, a su vez, señala una reducción en la brecha de desempeño en español para los alumnos de más bajo desempeño, a un nivel de significancia del 10%. Los procesos específicos a través de los cuales se alcanzó este resultado no son conocidos, yendo más allá del propósito de este estudio; haciendo un llamado a la investigación utilizando evaluaciones y metodologías complementarias.

Estos resultados ponen en contexto intervenciones que buscan reducir la discrecionalidad de los administradores educativos para seleccionar a los alumnos a los turnos matutino o vespertino de acuerdo con características como el desempeño escolar o nivel socioeconómico (Cárdenas y Ramírez de la Cruz, 2017). Si bien los resultados de estas decisiones en la actualidad ocasionan desigualdades en términos de equidad para los turnos escolares, el presente estudio demuestra que no resulta sencillo encontrar alternativas que mejoren la situación actual para los alumnos pertenecientes a escuelas de doble turno, ya que las mejoras que se pueden tener en términos de la reducción de las brechas de desempeño tienen una contraparte en cuanto a los indicadores de equidad; particularmente en términos de extraedad y repitencia. Actualmente no parecen existir determinantes en las características de la población de alumnos que incidan en el desempeño *per se* en un turno escolar o en el otro (Arrona-Palacios y Díaz-Morales, 2018). La mitigación

de las diferencias en uno y otro turno, en consecuencia, deben buscarse en la forma como se asignan los alumnos entre turnos escolares; en los recursos que se asignan en el turno vespertino *vis-á-vis* el matutino; y en las características de los profesores cuando enseñan en uno u otro. Por estos antecedentes, es que se recomienda finalmente utilizar un diseño experimental para la evaluación de cualquier política con un propósito similar, a fin de robustecer el cúmulo de evidencia.

2.7. Referencias

- Abadie, Alberto y Guido W. Imbens (2002) "Simple and Bias-corrected Matching Estimators for Average Treatment Effects.", *NBER Technical Working Paper No. 283*.
- (2006) "Large sample properties of matching estimators for average treatment effects.", en *Econometrica*, Vol. 74, No. 1, pp. 235-267.
- Adams, Guy B.; Danny L. Balfour y George Reed (2006) "Abu Ghraib, Administrative Evil, and Moral Inversion: The Value of "Putting Cruelty First."", en *Public Administration Review*, Vol. 66, No. 5, pp. 680-693.
- Angrist, Joshua D. y Jinyong Hahn (2004) "When to control for covariates? Panel asymptotics for estimates of treatment effects.", en *Review of Economics and Statistics*, Vol. 86, No. 1, pp. 58-72.
- Angrist, Joshua D. y Jörn-Steffen Pischke (2008) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- (2014) *Mastering Metrics: The path from cause to effect*. Princeton University Press.
- Arceo Gómez, Eva; Raymundo M. Campos Vázquez (2014a) "Teenage Pregnancy in Mexico: Evolution and Consequences", en *Latin American Journal of Economics*, Vol. 51, No. 1, pp. 109-146.

- (2014b) “Race and marriage in the labor market: A discrimination correspondence study in a developing country”, en *The American Economic Review*, Vol. 104, No. 5, pp. 376-380.
- Arrona-Palacios, Arturo y Juan F. Díaz-Morales (2018) “Morningness–eveningness is not associated with academic performance in the afternoon school shift: Preliminary findings”, en *British Journal of Educational Psychology*, Vol. 88, pp. 480-498.
- Ashong-Katai, Lois N.D. (2013) *Abolition of the Double Shift System of Schooling in Ghana: Policy and Its Implementation in Public Basic Schools*. Dissertation for the grade of Master in Philosophy in Comparative and International Education. Institute for Educational Research. Faculty of Education, Oslo University.
- Attanasio, Orazio, Costas Meghir y Ana Santiago (2008) “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to evaluate Progresá,” *IFS Working Papers*, No. 10, 14. Available at <http://hdl.handle.net/10419/47515>.
- Banerjee, Abhijit y Esther Duflo (2008) “The Experimental Approach to Development Economics” *NBER Working Paper 14467*, Cambridge, Massachusetts.
- Becerril, Javier y Awudu Abdulai (2010) "The impact of improved maize varieties on poverty in Mexico: a propensity score-matching approach.", en *World Development*, Vol. 38, No. 7, pp. 1024-1035.
- Becker, Sascha O. y Marco Caliendo (2007) “Sensitivity analysis for average treatment effects”. *The Stata Journal*, Vol. 7, No. 1, pp. 71-83.
- Becker, Sascha O. y Andrea Ichino (2002) “Estimation of average treatment effects based on propensity scores”, en *Stata Journal*, Vol. 2, No. 4, pp. 358-377.
- Beebout, H.S. (1972) *The production surface for academic achievement: An economic study of West Malaysian secondary schools*, Ph.D. dissertation, University of Wisconsin.

- Bellei, Cristián (2009) “Does lengthening the school day increase students’ academic achievement? Results from a natural experiment in Chile”, en *Economics of Education Review*, Vol. 28, No. 5, pp. 629-640.
- Bertrand, Marianne; Esther Duflo y Sendhil Mullainathan (2004) “How Much Should We Trust Differences-in-Differences Estimates?”, en *Quarterly Journal of Economics* 119, No. 1, pp. 249-275.
- Bray, Marc (2008) *Double Shift Schooling: Design and Operation for Cost Effectiveness*. Paris, International Institute for Educational Planning/UNESCO.
- (1989a) “Bisessional Versus Unisessional Education: Hong Kong Policies and Practice in Comparative Perspective”, en *New Horizons*, Vol. 30.
- (1989b) *Multiple-Shift Schooling: Design and Operation for Cost-Effectiveness*. Commonwealth Secretariat Publications, Marlborough House, Pall Mall, London SW1Y 5HX, England, United Kingdom.
- (1990a) “The Quality of Education in Multiple-Shift Schools: How Far Does a Financial Saving Imply an Educational Cost?” in *Comparative Education*, Vol. 26, No. 1, pp. 73-81.
- Ed. (1990b) *Educational Administration in Hong Kong: Personnel and Schools*. Faculty of Education, University of Hong Kong.
- Bruns, Barbara; Deon Filmer y Harry Anthony Patrinos (2011) *Making schools work: New evidence on accountability Reforms*. The World Bank, Washington D.C.
- Burke, Mary A. y Tim R. Sass (2008) “Classroom peer effects and student achievement”, en *National Center for Analysis of Longitudinal Data in Education Research*. Working Paper 18.

- Cabrera-Hernandez, Francisco (2015) "Does lengthening the school day increase students' academic achievement? Evidence from a natural experiment". University of Sussex, Working Paper Series No. 74-2015.
- Caliendo, Marco y Sabine Kopeinig (2008) "Some practical guidance for the implementation of propensity score matching", en *Journal of economic surveys*, Vol. 22, No. 1, pp. 31-72.
- Campbell, Donald T. (1970) "Considering the case against experimental evaluations of social innovations.", en *Administrative Science Quarterly*, Vol. 15, No. 1, pp. 110-113.
- Cárdenas Denham, Sergio (2009) "Is the Class Schedule the Only Difference between Morning and Afternoon Shift Schools in Mexico?". UMI Dissertation No. 3385009, Harvard Graduate School of Education, EUA.
- (2011) "Escuelas de doble turno en México. Una estimación de diferencias asociada con su implementación", en *Revista Mexicana de Investigación Educativa*, Vol. 16, No. 50, pp. 801-827.
- Cárdenas Denham, Sergio y Edgar Eugenio Ramirez de la Cruz (2017) "Controlling administrative discretion promotes social equity? Evidence from a natural experiment", en *Public Administration Review*, Vol. 77, No. 1, pp. 80-89.
- Carreón Rodríguez, Victor Gerardo, Sonia Di Giannatale y Carlos López Morales (2007) *Mercados formal e informal de crédito en México: Un estudio de caso*. Documento de trabajo No. 386, CIDE, División de Economía.
- (2015) *La política de evaluación en México: 10 años del CONEVAL*. CONEVAL, México.
- DiPrete, Thomas A. y Markus Gangl (2004) "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with

- imperfect instruments”, *WZB Discussion Paper, No. SP I 2004-101*, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin.
- Duflo, Esther (2000) *Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment* (No. w7860). National Bureau of Economic Research.
- Dunning, Thad (2012) *Natural Experiments in the Social Sciences. A Design-Based Approach*. Cambridge University Press, New York.
- Fiallo Billini, José Antinoe. (2003) *La doble tanda: efectos, límites y propuestas*. UNESCO Santo Domingo.
- Gilligan, Michael J. y Ernest J. Sergenti (2008) “Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference”, en *Quarterly Journal of Political Science*, Vol. 3, pp. 89-122.
- Green, Donald P.; Shang E. Ha y John G. Bullock (2010) “Enough already about “black box” experiments: Studying mediation is more difficult than most scholars suppose”, en *The Annals of the American Academy of Political and Social Science*, Vol. 628, No. 1, pp. 200-208.
- Ho, Daniel E., Kosuke Imai, Gary King y Elizabeth Stuart (2007) “Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference”, en *Political Analysis*, Vol. 15, No. 3, pp.199-236. Disponible en <https://dash.harvard.edu/handle/1/4214880>.
- Imai, Kosuke; Luke Keele, Dustin Tingley y Teppei Yamamoto (2011) “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies”, en *American Political Science Review*, Vol. 105, No. 4, pp. 765-789.

- Kea, Pamela (2007) “Girl Farm Labour and Double-Shift Schooling in The Gambia: The Paradox of Development Intervention”, PRUS Working Paper no. 39.
- Khandker, Shahidur R., Gayatri B. Koolwal y Hussain A. Samad (2010) *Handbook on Impact Evaluation. Quantitative Methods and Practices*. The World Bank.
- King, Gary y Richard Nielsen (2016) “Why Propensity Scores Should Not Be Used for Matching”. Disponible en j.mp/PScore. Consultado el 7 de marzo de 2019.
- King, Gary, y Langche Zeng (2006) “The dangers of extreme counterfactuals”, en *Political Analysis*, Vol. 14, pp. 131–159.
- Kleinhans, Godfrey (2008) “Double shift schooling: Motives for implementation- the cases of Namibia and Uganda”. *Southern Africa Review of Education* 14(3), 106-121.
- Lalive; Rafael y M. Alejandra Cattaneo (2009) “Social Interactions and Schooling Decisions”, en *The Review of Economics and Statistics*, Vol. 91, pp. 457-477.
- Lee, Myoung-jae (2016) *Matching, Regression Discontinuity, Difference in Differences, and Beyond*. Oxford University Press.
- Leeuw, Frans y Jos Vaessen (2009) “NONIE Guidance on Impact Evaluation”; Draft Version for Discussion at the Cairo conference March-April, 2009. Network on Networks on Impact Evaluation.
- Linden, Toby (2001) *Double Shift Secondary Schools: Possibilities and Issues*. World Bank.
- Llach, Juan; Cecilia Androgué; María Gigaglia, y Catherine Rodríguez Orgales (2009) “Do Longer School Days Have Enduring Educational, Occupational, or Income Effects? A Natural Experiment in Buenos Aires, Argentina”, in *Economía*, Vol. 10, No. 1, pp. 1-43.
- Lusher, Lester y Vasil Yassenov (2016) “Double-shift schooling and student success: Quasi-experimental evidence from Europe”, en *Economics Letters*, Vol. 139, pp. 36-39.

- Meyer, Bruce D. (1995) "Natural and Cuasi-experiments in Economics", en *Journal of Business and Economic Statistics*, Vol. 13, pp. 151-161.
- Orkodashvili, Mariam (2010) "Double-Shift Schooling and EFA Goals: Assessing Economic, Educational and Social Impacts" MPRA Paper No. 23519.
- Ravallion, Martin (1999) *The mystery of the vanishing benefits: Ms Speedy analyst's introduction to evaluation* (Vol. 2153). World Bank Publications.
- Reimers, Fernando (coord.) (2006) *Aprender más y mejor. Políticas, programas y oportunidades de aprendizaje en educación básica en México*. Fondo de Cultura Económica, México.
- Rojas Valdés, Rubén Irvin (2018) *Migration, Microfinance, and Microentrepreneurship in Rural Mexico: Neighborhood Effects, Dynamic Behavior, and Hope*. UMI Dissertation No. 10846284, The University of California, EUA.
- Sagyndykova, Galiya (2014) "Academic Performance in Double-Shift Schooling", in *Three Essays in Microeconomics of Education*. UMI Dissertation No. 3620725. The University of Arizona, EUA.
- Sesma-Vázquez, Sergio; Julio César Campuzano-Rincón, Víctor Gerardo Carreón-Rodríguez, Felicia Knaul, Francisco Javier López-Antuñano y Mauricio Hernández-Avila (2002) "El comportamiento de la demanda de tabaco en México: 1992-1998.", en *Salud Pública de México*, Vol. 44, pp. s82-s92.
- Shavelson, Richard J., y Lisa Towne (2002) *Scientific research in education*. National Academy Press, Washington, DC.
- Sheryn, Aishath (2011) "Improving Efficiency of Schooling in the Maldives: Is De-shifting a Desirable Policy Direction?" *Thesis for Master of Arts*, Massey University, New Zealand.

- Smith, Jeffrey A. y Petra E. Todd (2005) “Does matching overcome LaLonde's critique of nonexperimental estimators?”, en *Journal of econometrics*, Vol. 125, No. 1-2, pp. 305-353.
- Tedla, Habtezhgi Kidane (2003) *The implementation and organization of double-shift schooling in Eritrea*, Tesis de Maestría, University of the Western Cape.
- Vilalta, Carlos (2005) “Cómo enseñar autocorrelación espacial”, en *Economía, Sociedad y Territorio*, Vol. 5, No. 18, pp. 323-333.
- Yim, Chiu Shiu (1990) “The administration of Bisessional Primary Schools”, in Mark Bray (Ed.) *Educational Administration in Hong Kong: Personnel and Schools*. Faculty of Education, University of Hong Kong.
- Zimmerman, David J. (2003) “Peer effects in academic outcomes: Evidence from a natural experiment”, en *Review of Economics and statistics*, Vol. 85, No. 1, pp. 9-23.

Anexo 2A. Valor promedio de las diferencias entre el estado tratado y el resto del país, un año antes de la intervención

Estado	Oportunidades		Reprobación		Extraedad		Deserción		Repitencia	
	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia
Aguascalientes	2.31 (4.52)	-2.82*** (0.55)	3.58 (5.65)	1.12* (0.81)	6.65 (6.84)	-7.01*** (1.04)	5.87 (6.34)	-7.35*** (1.00)	5.37 (7.27)	-7.84*** (1.04)
Baja California	0.61 (2.60)	-4.52*** (0.46)	0.27 (2.63)	-2.18*** (0.70)	8.65 (9.93)	-5.02*** (1.01)	7.98 (8.73)	-5.24*** (0.97)	0.50 (2.66)	-12.72*** (0.90)
Baja California Sur	1.72 (2.72)	-3.41*** (0.56)	2.72 (5.06)	0.27 (0.95)	6.29 (10.34)	-7.38*** (1.60)	11.79 (7.30)	-1.43 (1.29)	5.70 (8.23)	-7.52*** (1.37)
Campeche	5.22 (6.41)	0.09 (0.67)	1.81 (9.68)	-0.65 (1.04)	16.18 (16.46)	2.51* (1.59)	6.50 (9.13)	-6.72*** (1.15)	4.38 (8.08)	-8.84*** (1.10)
Chiapas	5.83 (10.89)	0.70 (0.98)	4.12 (9.44)	1.66* (1.03)	21.52 (16.02)	7.86*** (1.57)	5.09 (11.18)	-8.13*** (1.26)	4.70 (9.77)	-8.13*** (1.26)
Chihuahua	1.84 (7.69)	-3.29*** (0.76)	0.15 (2.22)	-2.30*** (0.70)	5.27 (7.59)	-8.40*** (0.97)	2.94 (7.53)	-10.28 (0.95)	3.38 (6.74)	-9.84*** (0.94)
Coahuila	1.26 (2.96)	-3.87*** (0.47)	1.48 (4.40)	-0.97* (0.73)	3.66 (5.97)	-10.01*** (0.96)	4.06 (5.87)	-9.16*** (0.94)	1.08 (4.39)	-12.14 (0.92)
Distrito Federal	0.37 (1.52)	-4.76*** (0.45)	0.90 (12.78)	-1.55** (0.88)	6.72 (8.05)	-6.95*** (0.97)	1.53 (7.44)	-11.69*** (0.95)	2.59 (4.87)	-10.63*** (0.92)
Durango	4.79 (6.00)	-0.34 (0.62)	0.69 (4.21)	-1.77** (0.76)	6.46 (11.17)	-7.21*** (1.21)	7.50 (9.90)	-5.72*** (1.14)	1.53 (4.23)	-11.69*** (0.94)
Guanajuato	9.78 (22.17)	4.65*** (1.66)	0.71 (7.04)	-1.75*** (0.72)	3.00 (14.72)	-10.67*** (1.00)	2.51 (7.78)	-10.71*** (0.92)	1.04 (7.82)	-12.18*** (0.92)
Guerrero	7.05 (9.32)	1.92** (1.10)	3.77 (10.16)	1.32 (1.30)	12.63 (13.20)	-1.04 (1.69)	3.24 (7.35)	-9.98*** (1.19)	3.04 (9.11)	-10.18*** (1.33)
Hidalgo	4.51 (6.37)	-0.62 (0.67)	4.59 (6.07)	2.13*** (0.84)	9.14 (7.93)	-4.53*** (1.10)	5.21 (6.15)	-8.01*** (1.01)	4.87 (6.52)	-8.35*** (1.03)
Jalisco	3.04 (7.60)	-2.09*** (0.47)	3.17 (7.95)	0.71 (0.72)	8.47 (11.40)	-5.20*** (0.94)	4.13 (8.15)	-9.09*** (0.91)	4.26 (9.34)	-8.95*** (0.91)
México	1.79 (4.18)	-3.34*** (0.45)	2.41 (5.33)	-0.05 (0.70)	7.74 (8.58)	-5.93*** (0.93)	4.13 (6.46)	-9.09*** (0.90)	4.32 (6.61)	-8.90*** (0.90)

Nota: niveles de significancia: *P<0.1, **P<0.05, ***P<0.01. Fuente: cálculos propios con base en la Estadística 911 de la SEP.

Anexo 2A. Valor promedio de las diferencias entre el estado tratado y el resto del país, un año antes de la intervención (Cont.)

Estado	Oportunidades		Reprobación		Extraedad		Deserción		Repitencia	
	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia
Michoacán	0.61 (5.18)	-4.52*** (0.49)	3.33 (8.83)	0.87 (0.78)	13.47 (12.98)	-0.19 (1.06)	3.38 (7.72)	-9.84*** (0.95)	5.74 (10.11)	-7.48*** (0.98)
Morelos	4.19 (7.34)	-0.94*** (0.33)	2.04 (5.74)	-0.41 (0.77)	11.34 (10.74)	-2.33** (1.10)	3.97 (6.76)	-9.24 (0.97)	3.14 (6.02)	-10.08*** (0.96)
Nayarit	4.29 (7.49)	-0.84 (0.76)	3.50 (9.74)	1.05 (1.47)	10.72 (11.26)	-2.95** (1.76)	8.25 (9.35)	-4.97*** (1.53)	3.25 (5.44)	-9.97*** (1.15)
Nuevo León	0.82 (2.43)	-4.31*** (0.21)	2.10 (6.21)	-0.36 (0.73)	5.53 (8.64)	-8.13*** (0.97)	1.71 (7.22)	-11.50*** (0.93)	3.08 (6.46)	-10.14*** (0.92)
Oaxaca	5.04 (8.57)	-0.09 (0.35)	2.42 (10.92)	-0.03 (0.96)	10.41 (13.06)	-3.26** (1.21)	3.27 (8.55)	-9.95*** (1.03)	2.88 (10.82)	-10.34*** (1.11)
Puebla	4.52 (7.82)	-0.61** (0.27)	3.04 (5.92)	0.59 (0.75)	10.81 (10.80)	-2.86*** (1.04)	3.11 (6.04)	-10.10*** (0.94)	3.47 (6.68)	-9.75*** (0.95)
Querétaro	2.71 (5.18)	-2.42*** (0.42)	3.62 (4.90)	1.17* (0.78)	7.58 (8.18)	-6.09*** (1.09)	4.07 (4.45)	-9.15*** (0.95)	4.66 (5.76)	-8.56*** (0.98)
Quintana Roo	2.29 (4.57)	-2.84*** (0.45)	2.90 (6.29)	0.44 (0.86)	6.14 (9.19)	-7.53*** (1.17)	2.99 (6.25)	-10.23*** (1.02)	1.76 (5.73)	-11.46*** (1.00)
San Luis Potosí	3.87 (6.87)	-1.26*** (0.29)	3.68 (8.27)	1.22* (0.81)	8.63 (12.73)	-5.03*** (1.12)	4.11 (6.49)	-9.11*** (0.95)	4.09 (9.60)	-9.13*** (1.02)
Sinaloa	3.63 (6.52)	-1.50*** (0.33)	5.36 (6.30)	2.91*** (0.78)	14.75 (15.02)	1.08 (1.26)	5.17 (8.39)	-8.05*** (1.01)	3.17 (5.97)	-10.05*** (0.96)
Sonora	2.75 (5.25)	-2.38*** (0.27)	1.77 (4.51)	-0.68 (0.73)	7.93 (7.89)	-5.74*** (0.98)	9.00 (9.25)	-4.22*** (0.99)	3.26 (6.35)	-9.96*** (0.94)
Tabasco	1.38 (3.25)	-3.75*** (0.42)	3.54 (8.37)	1.08 (0.92)	10.25 (10.30)	-3.42*** (1.18)	3.82 (7.58)	-9.40*** (1.05)	4.43 (7.78)	-8.79*** (1.06)
Tamaulipas	2.79 (5.30)	-2.34*** (0.26)	1.70 (6.26)	-0.75 (0.75)	8.84 (10.08)	-4.83*** (1.02)	5.58 (10.08)	-7.64*** (1.01)	3.28 (6.91)	-9.94*** (0.95)
Tlaxcala	4.80 (8.23)	-0.33 (0.51)	1.92 (4.41)	-0.53 (0.80)	6.58 (8.29)	-7.09*** (1.17)	2.87 (5.66)	-10.35*** (1.02)	3.86 (5.62)	-9.36*** (1.02)
Veracruz	4.03 (7.10)	-1.10*** (0.22)	4.62 (8.69)	2.17*** (0.77)	13.19 (13.78)	-0.48 (1.05)	3.54 (8.41)	-9.68*** (0.95)	6.24 (11.03)	-6.98*** (0.99)
Yucatán	3.79 (6.75)	-1.34*** (0.39)	2.66 (6.70)	0.21 (0.83)	5.80 (12.12)	-7.87*** (1.24)	3.10 (5.98)	-10.12*** (0.98)	1.78 (6.67)	-11.43*** (1.00)
Zacatecas	4.49 (7.77)	-0.64* (0.45)	-1.85 (6.75)	-4.31*** (0.87)	-1.46 (10.13)	-15.13*** (1.21)	1.13 (6.81)	-12.09*** (1.04)	-2.15 (8.42)	-15.37*** (1.11)

Nota: niveles de significancia: *P<0.1, **P<0.05, ***P<0.01. Fuente: cálculos propios con base en la Estadística 911 de la SEP.

Anexo 2A. Valor promedio de las diferencias entre el estado tratado y el resto del país, un año antes de la intervención (Cont.)

Variable	ENLACE español nivel 0		ENLACE matemáticas nivel 0		ENLACE español nivel 3		ENLACE matemáticas nivel 3	
	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia
Aguascalientes	5.13 (13.36)	-2.62** (1.59)	4.66 (14.79)	-2.60* (1.76)	-2.02 (4.28)	-0.36 (0.56)	-3.26 (6.17)	-0.02 (0.72)
Baja California	4.77 (14.77)	-2.98** (1.43)	5.37 (15.12)	-1.89 (1.55)	-1.45 (8.59)	0.21 (0.60)	-1.61 (10.23)	1.63** (0.73)
Baja California Sur	3.1 (14.38)	-4.66** (2.22)	3.30 (16.47)	-3.96* (2.51)	-2.00 (3.85)	-0.34 (0.67)	-2.78 (5.37)	0.46 (0.88)
Campeche	6.98 (15.57)	-0.77 (1.77)	11.62 (21.80)	4.36** (2.23)	-4.33 (14.96)	-2.67** (1.28)	-3.53 (15.60)	-0.29 (1.36)
Chiapas	11.77 (30.37)	4.02* (2.86)	13.73 (30.94)	6.47** (2.96)	-4.13 (14.10)	-2.47** (1.28)	-5.74 (18.03)	-2.51* (1.62)
Chihuahua	4.05 (16.58)	-3.71*** (1.45)	3.87 (17.16)	-3.39** (1.58)	-1.57 (10.01)	0.09 (0.63)	-2.91 (10.77)	0.33 (0.72)
Coahuila	7.66 (14.74)	-0.10* (1.46)	7.79 (14.25)	0.53 (1.57)	-2.66 (6.31)	-1.00** (0.56)	-3.97 (7.52)	-0.73 (0.67)
Distrito Federal	5.62 (13.04)	-2.14* (1.38)	5.30 (14.50)	-1.96* (1.53)	-1.70 (4.72)	-0.04 (0.50)	-1.84 (6.61)	1.40** (0.63)
Durango	8.71 (25.38)	0.95 (2.23)	5.76 (27.21)	-1.50 (2.42)	-4.01 (13.10)	-2.35** (1.06)	-4.82 (17.51)	-1.58 (1.39)
Guanajuato	1.62 (24.84)	-6.13*** (1.44)	2.17 (28.11)	-5.10*** (1.60)	-1.32 (9.48)	0.34 (0.53)	-2.14 (9.81)	1.09 (0.62)
Guerrero	6.43 (18.17)	-1.32 (2.33)	8.45 (21.11)	1.18 (2.67)	-3.75 (12.75)	-2.09* (1.45)	-6.29 (10.17)	-3.05*** (1.23)
Hidalgo	7.94 (15.59)	0.18 (1.76)	8.90 (17.15)	-6.08*** (1.41)	-2.39 (10.94)	-0.73 (0.98)	-2.69 (10.92)	0.55 (1.03)
Jalisco	3.36 (15.56)	-4.40*** (1.76)	3.17 (7.95)	-4.09*** (1.44)	-1.22 (9.38)	0.44 (0.51)	-1.85 (12.26)	1.39** (0.62)
México	6.57 (13.81)	-1.19 (1.30)	7.04 (15.70)	-0.22 (1.44)	-1.18 (5.14)	0.48 (0.48)	-1.98 (6.18)	1.25 (0.57)
Michoacán de Ocampo	0 (0)	-7.76 (n.a.)	0 (0)	-7.26 (n.a.)	0 (0)	1.65 (n.a.)	0 (0)	3.23 (0)

Nota: niveles de significancia: *P<0.1, **P<0.05, ***P<0.01. Fuente: cálculos propios con base en la Estadística 911 de la SEP.

Anexo 2A. Valor promedio de las diferencias entre el estado tratado y el resto del país, un año antes de la intervención (Cont.)

Variables	ENLACE español nivel 0		ENLACE matemáticas nivel 0		ENLACE español nivel 3		ENLACE matemáticas nivel 3	
	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia	Promedio	Diferencia
Morelos	9.91 (16.12)	2.15* (1.57)	10.59 (16.93)	3.33** (1.70)	-2.24 (7.14)	-0.58 (0.62)	-3.60 (10.63)	-0.37 (0.83)
Nayarit	7.22 (14.05)	-0.53 (2.26)	6.67 (18.89)	-0.59 (2.88)	-2.90 (6.22)	-1.24* (0.95)	-7.21 (10.19)	-3.97 (1.47)
Nuevo León	3.16 (14.00)	-4.60*** (1.37)	2.84 (12.97)	-4.42*** (1.48)	-1.90 (6.34)	-0.24 (0.52)	-2.46 (9.09)	0.78 (0.65)
Oaxaca	8.10 (25.27)	0.35 (2.05)	7.75 (26.88)	0.49 (2.21)	-1.53 (20.81)	0.13 (1.40)	-1.31 (23.13)	1.92 (1.58)
Puebla	6.83 (17.43)	-0.92 (1.51)	8.04 (19.08)	0.78 (1.66)	-2.28 (8.82)	-0.62 (0.62)	-3.33 (10.44)	-0.10 (0.74)
Querétaro	3.95 (13.58)	-3.80*** (1.60)	4.95 (14.75)	-2.31* (1.76)	-1.33 (5.04)	0.33 (0.59)	-1.48 (6.52)	1.75*** (0.73)
Quintana Roo	3.38 (12.12)	-4.38*** (1.59)	3.55 (15.22)	-3.71** (1.85)	-0.76 (3.56)	0.90** (0.54)	-0.04 (4.39)	3.19*** (0.66)
San Luis Potosí	7.87 (19.37)	0.12 (1.61)	10.41 (21.84)	3.15** (1.79)	-2.58 (4.74)	-0.92** (0.52)	-3.91 (6.91)	-0.68 (0.66)
Sinaloa	9.71 (19.19)	1.95 (1.68)	11.43 (21.19)	4.17** (1.86)	-0.81 (7.99)	0.84* (0.65)	-2.52 (9.47)	0.72 (0.78)
Sonora	3.56 (14.02)	-4.19*** (1.43)	4.05 (18.04)	-3.21** (1.64)	-1.85 (17.06)	-0.19 (0.94)	-2.62 (16.31)	0.61 (0.96)
Tabasco	12.60 (26.65)	4.84*** (2.32)	10.68 (25.03)	3.42* (2.30)	-1.82 (13.87)	-0.16 (1.11)	-1.13 (14.87)	2.11** (1.22)
Tamaulipas	9.00 (16.00)	1.25 (1.46)	9.97 (17.22)	2.71** (1.61)	-2.87 (8.60)	-1.21 (0.61)	-3.28 (10.69)	-0.05 (0.75)
Tlaxcala	6.00 (16.88)	-1.76 (1.48)	7.40 (18.68)	0.14 (2.17)	-1.66 (9.56)	< (-0.01) (0.97)	-2.86 (7.64)	0.38 (0.88)
Veracruz	4.54 (19.11)	-3.22** (1.46)	5.26 (19.51)	-2.00* (1.58)	-1.91 (12.06)	-0.26 (0.65)	-1.80 (13.41)	1.43** (0.76)
Yucatán	4.93 (15.15)	-2.83** (1.64)	3.65 (15.51)	-3.62 (1.76)	-1.09 (5.24)	0.57 (0.59)	-2.39 (7.69)	0.85 (0.77)
Zacatecas	-5.10 (28.64)	-12.85*** (2.67)	4.81 (32.11)	-10.88*** (1.40)	-2.47 (7.82)	-0.81 (0.79)	-2.97 (11.84)	0.27 (1.12)

Nota: niveles de significancia: *P<0.1, **P<0.05, ***P<0.01. Fuente: cálculos propios con base en la Estadística 911 de la SEP.

Anexo 2B. Variables utilizadas en el modelo de marcador de propensión³⁴

Nombre de las variables	Descripción
Oportunidades	Porcentaje de alumnos bajo el programa <i>Oportunidades</i> en la escuela.
n0_e3	Porcentaje de alumnos en el nivel 0 en la Prueba ENLACE de español.
n1_e3	Porcentaje de alumnos en el nivel 1 en la Prueba ENLACE de español.
n2_e3	Porcentaje de alumnos en el nivel 2 en la Prueba ENLACE de español.
n3_e3	Porcentaje de alumnos en el nivel 3 en la Prueba ENLACE de español.
n0_m3	Porcentaje de alumnos en el nivel 0 en la Prueba ENLACE de matemáticas.
n1_m3	Porcentaje de alumnos en el nivel 1 en la Prueba ENLACE de matemáticas.
n2_m3	Porcentaje de alumnos en el nivel 2 en la Prueba ENLACE de matemáticas.
n3_m3	Porcentaje de alumnos en el nivel 3 en la Prueba ENLACE de matemáticas.
Ovtot1	Porcentaje de alumnos con algún grado de extraedad; primer grado.
Ovtot2	Porcentaje de alumnos con algún grado de extraedad; segundo grado.
Ovtot3	Porcentaje de alumnos con algún grado de extraedad; tercer grado.
Repetition1	Porcentaje de alumnos con al menos un ciclo escolar reprobado; primer grado.
Repetition2	Porcentaje de alumnos con al menos un ciclo escolar reprobado; segundo grado.
Repetition3	Porcentaje de alumnos con al menos un ciclo escolar reprobado; tercer grado.
Drp1	Porcentaje de alumnos que no terminaron el ciclo escolar cursado; primer grado.
Drp2	Porcentaje de alumnos que no terminaron el ciclo escolar cursado; segundo grado.
Drp3	Porcentaje de alumnos que no terminaron el ciclo escolar cursado; tercer grado.
Fail1	Porcentaje de alumnos que no aprobaron el ciclo escolar cursado, primer grado.
Fail2	Porcentaje de alumnos que no aprobaron el ciclo escolar cursado, segundo grado.
Fail3	Porcentaje de alumnos que no aprobaron el ciclo escolar cursado, tercer grado.

Fuente: construcción propia con base en la información presente en la Estadística 911 de la SEP.

³⁴ Debido a la falta de convergencia de la función logit, las variables de nivel educativo de los profesores y nivel educativo del director fueron desechadas. Las variables finales en el modelo logit se encuentran en negritas.

Anexo 2B. Variables utilizadas en el modelo de marcador de propensión (Cont.)

Nombre de las variables	Descripción
Indig	Número total de estudiantes indígenas.
Pre1	Porcentaje de alumnos que cursaron 1 año de preescolar.
Pre2	Porcentaje de alumnos que cursaron 2 años de preescolar.
Pre3	Porcentaje de alumnos que cursaron 3 años de preescolar.
Ent	Inscripción total a la escuela.
Ukids	Porcentaje de alumnos atendidos por la USAER.
Disable	Porcentaje de alumnos con alguna discapacidad.
Gifted	Número total de estudiantes sobredotados.
G1	Número total de grupos de primer grado en la escuela.
G2	Número total de grupos de segundo grado en la escuela.
G3	Número total de grupos de tercer grado en la escuela.
Gtot	Número total de grupos en la escuela.
Dir	Número total de directores en la escuela.
Teachers	Número total de profesores frente a grupo.
Ent	Número total de estudiantes inscritos.
Blindtotal	Número total de estudiantes con ceguera.
Visiontotal	Número total de estudiantes con problemas de visión.
Deaftotal	Número total de estudiantes con sordera.
Lohtotal	Número total de estudiantes con problemas de audición.
Lomototal	Número total de estudiantes con alguna discapacidad motora.
Locogtot	Número total de estudiantes con alguna discapacidad mental.
Special	Número de alumnos con necesidades educativas especiales.
Femprincele	Directoras con primaria incompleta.
Maleprincele	Directores con primaria incompleta.
Femprcele	Directoras con primaria completa.
Maleprcele	Directores con primaria completa.
Femprincsec	Directoras con secundaria incompleta.
Maleprincsec	Directores con secundaria incompleta.
Femprcsec	Directoras con secundaria completa.
Maleprcsec	Directores con secundaria completa.

Fuente: construcción propia con base en la información presente en la Estadística 911 de la SEP.

Anexo 2B. Variables utilizadas en el modelo de marcador de propensión (Cont.)

Nombre de las variables	Descripción
Femprinctech	Directoras con carrera técnica incompleta.
Maleprinctech	Directores con carrera técnica incompleta.
Femprctech	Directoras con carrera técnica completa.
Maleprctech	Directores con carrera técnica completa.
Femprinchi	Directoras con preparatoria incompleta.
Maleprinchi	Directores con preparatoria incompleta.
Femprchi	Directoras con preparatoria completa.
Maleprchi	Directores con preparatoria completa.
Femprinpresc	Directoras con estudios incompletos en enseñanza en preescolar.
Maleprinpresc	Directores con estudios incompletos en enseñanza en preescolar.
Femprpresc	Directoras con estudios completos en enseñanza en preescolar.
Maleprpresc	Directores con estudios completos en enseñanza en preescolar.
Femprinctea	Directoras con formación docente incompleta.
Maleprinctea	Directores con formación docente incompleta.
Femprctea	Directoras con formación docente completa.
Maleprctea	Directores con formación docente completa.
Femprincot	Directoras con educación superior incompleta en docencia.
Maleprincot	Directores con educación superior incompleta en docencia.
Femprccot	Directoras con educación superior terminada en docencia.
Maleprccot	Directores con educación superior terminada en docencia.
Femprccotd	Directoras con educación superior terminada en docencia, tituladas.
Maleprccotd	Directores con educación superior terminada en docencia, titulados.
Femprincol	Directoras con estudios universitarios incompletos.
Maleprincol	Directores con estudios universitarios incompletos.
Femprccol	Directoras con estudios universitarios completos.
Maleprccol	Directores con estudios universitarios completos.
Femprccold	Directoras con estudios universitarios completos, tituladas.
Maleprccold	Directores con estudios universitarios completos, titulados.
Femprinemas	Directoras con estudios de maestría incompletos.
Maleprinemas	Directores con estudios de maestría incompletos.

Fuente: construcción propia con base en la información presente en la Estadística 911 de la SEP.

Anexo 2B. Variables utilizadas en el modelo de marcador de propensión (Cont.)

Nombre de las variables	Descripción
Femprcmas	Directoras con estudios de maestría completos.
Maleprcmas	Directores con estudios de maestría completos.
Femprincphd	Directoras con estudios de doctorado incompletos.
Maleprincphd	Directores con estudios de doctorado incompletos.
Femprcphd	Directoras con estudios de doctorado completos.
Maleprcphd	Directores con estudios de doctorado completos.
Femteaielem	Profesoras con primaria incompleta.
Maleteaielem	Profesores con primaria incompleta.
Femteacelem	Profesoras con primaria completa.
Maleteacelem	Profesores con primaria completa.
Femteaincsec	Profesoras con secundaria incompleta.
Maleteaincsec	Profesores con secundaria incompleta.
Femteacsec	Profesoras con primaria completa.
Maleteacsec	Profesores con primaria completa.
Femteaintech	Profesoras con carrera técnica incompleta.
Maleteaintech	Profesores con carrera técnica incompleta.
Femteactech	Profesoras con carrera técnica completa.
Maleteactech	Profesores con carrera técnica completa.
Femteainchigh	Profesoras con preparatoria incompleta.
Maleteainchigh	Profesores con preparatoria incompleta.
Femteainchigh	Profesoras con preparatoria completa.
Maleteachigh	Profesores con preparatoria completa.
Femteaincpre	Profesoras con estudios incompletos en enseñanza en preescolar.
Maleteaincpre	Profesores con estudios incompletos en enseñanza en preescolar.
Femteacpre	Profesoras con estudios completos en enseñanza en preescolar.
Maleteacpre	Profesores con estudios completos en enseñanza en preescolar.
Femteainctea	Profesoras con formación docente incompleta.
Maleteainctea	Profesores con formación docente incompleta.
Femteactea	Profesoras con formación docente completa.
Maleteactea	Profesores con formación docente completa.

Fuente: construcción propia con base en la información presente en la Estadística 911 de la SEP.

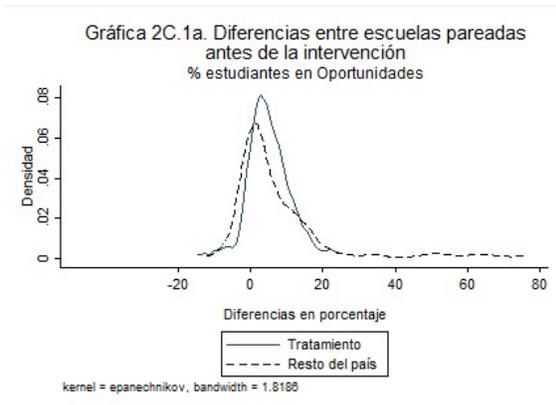
Anexo 2B. Variables utilizadas en el modelo de marcador de propensión (Cont.)

Nombre de las variables	Descripción
Femteainccot	Profesoras con educación superior incompleta en docencia.
Maleteainccot	Profesores con educación superior incompleta en docencia.
Femteaccot	Profesoras con educación superior terminada en docencia.
Maleteaccot	Profesores con educación superior terminada en docencia.
Femteaccotd	Profesoras con educación superior terminada en docencia, tituladas.
Maleteaccotd	Profesores con educación superior terminada en docencia, titulados.
Femteainccol	Profesoras con estudios universitarios incompletos.
Maleteainccol	Profesores con estudios universitarios incompletos.
Femteaccol	Profesoras con estudios universitarios completos.
Maleteaccol	Profesores con estudios universitarios completos.
Femteaccold	Profesoras con estudios universitarios completos, tituladas.
Maleteaccold	Profesores con estudios universitarios completos, titulados.
Femteaincmas	Profesoras con estudios de maestría incompletos.
Maleteaincmas	Profesores con estudios de maestría incompletos.
Femteacmas	Profesoras con estudios de maestría completos.
Maleteacmas	Profesores con estudios de maestría completos.
Femteaincphd	Profesoras con estudios de doctorado incompletos.
Maleteaincphd	Profesores con estudios de doctorado incompletos.
Femteacphd	Profesoras con estudios de doctorado completos.
Maleteacphd	Profesores con estudios de doctorado completos.
Teatot	Número total de profesoes en la escuela.
Secretaries	Número total de secretarias en la escuela.
Admon	Número total de personal administrativo en la escuela.
Janitors	Número total de conserjes en la escuela.
Magister	Número total de profesores en el Programa “Carrera Magisterial”.
Avaroom	Número total de salones de clase disponibles.
Usiroom	Número total de salones de clase en uso.
Matexp	Gasto total anual en material escolar.
Fees	Gasto total anual en cuotas escolares.

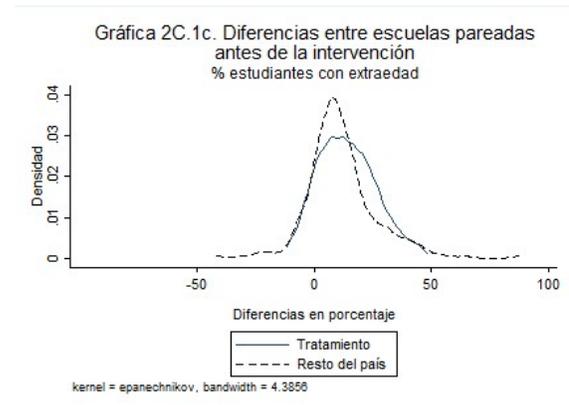
Fuente: construcción propia con base en la información presente en la Estadística 911 de la SEP.

Anexo 2C. Algoritmo alternativo para el apareamiento utilizando tres vecinos más cercanos con imputación de ceros en los valores perdidos

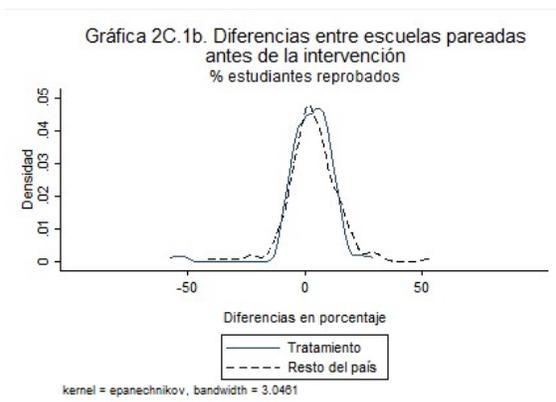
Gráfica 2C.1a-i. Distribución de las diferencias entre escuelas pareadas (NN3); variables de interés. Escuelas primarias.



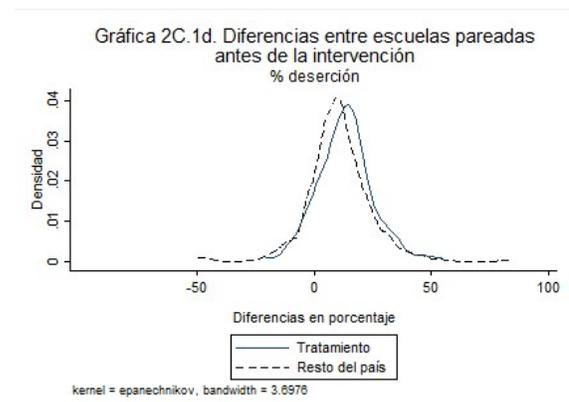
Fuente: cálculos propios.



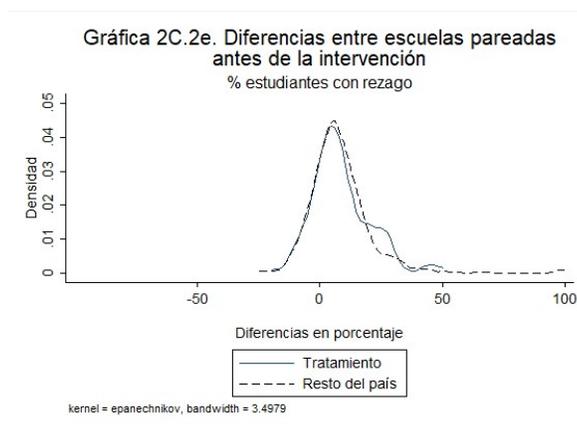
Fuente: cálculos propios.



Fuente: cálculos propios.

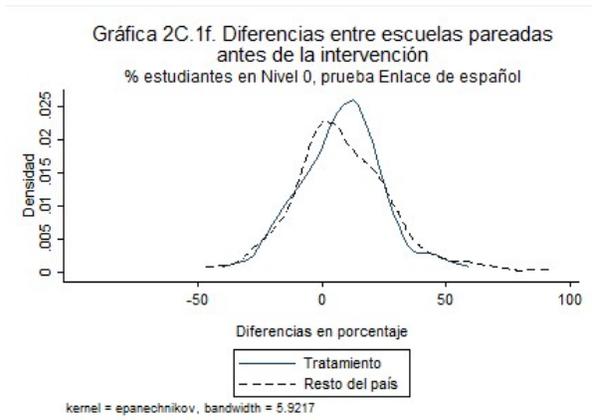


Fuente: cálculos propios.

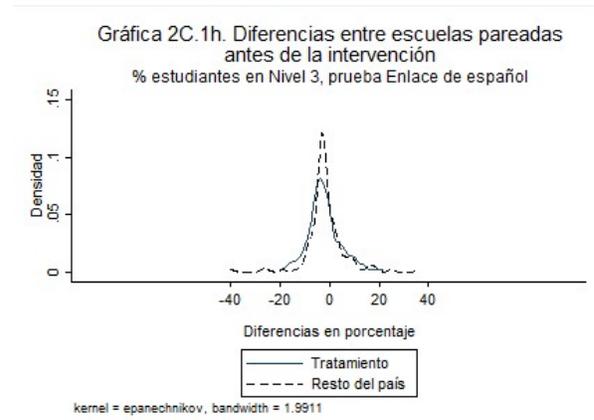


Fuente: cálculos propios.

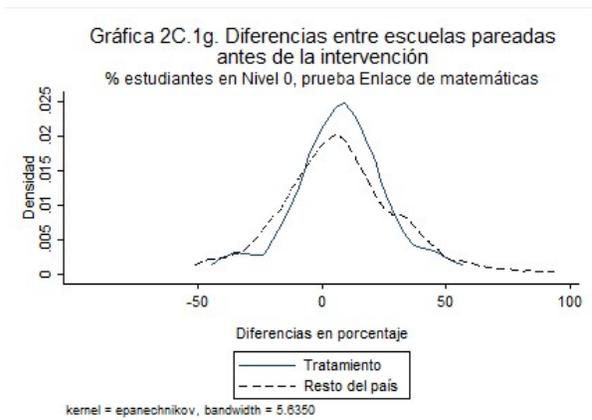
Gráfica 2C.1a-i. Distribución de las diferencias entre turnos; resultados seleccionados
Escuelas primarias (Cont.)



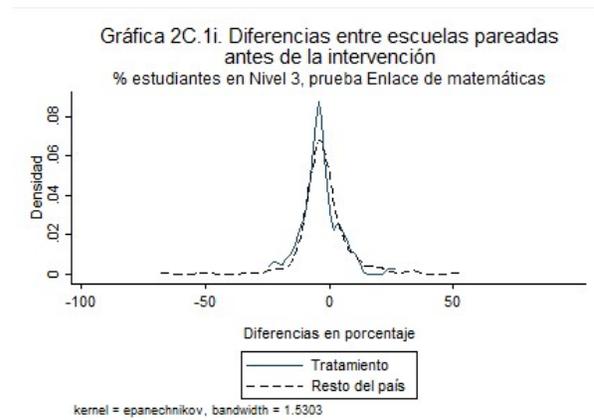
Fuente: cálculos propios.



Fuente: cálculos propios.

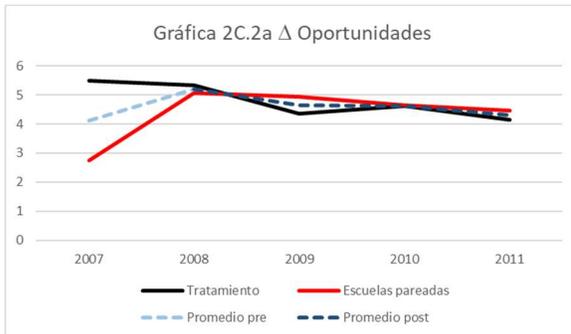


Fuente: cálculos propios.

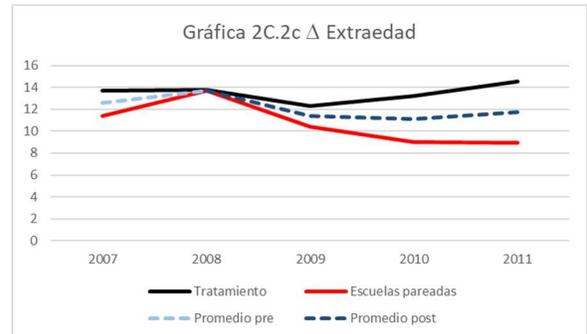


Fuente: cálculos propios.

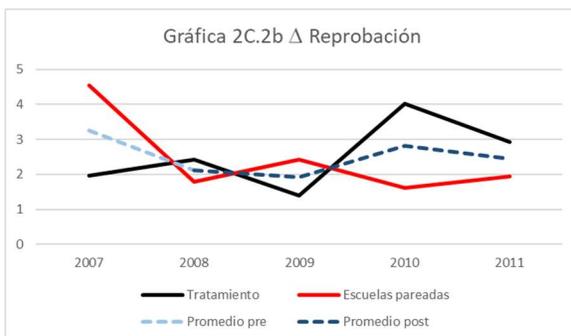
Gráfica 2C.2a-i. Diferencias en las tendencias entre grupos;
cambios en resultados seleccionados (NN3). Escuelas primarias.



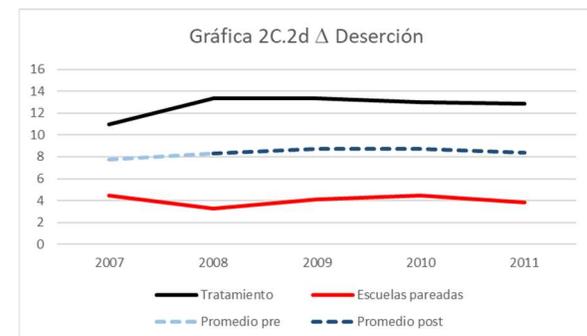
Fuente: cálculos propios.



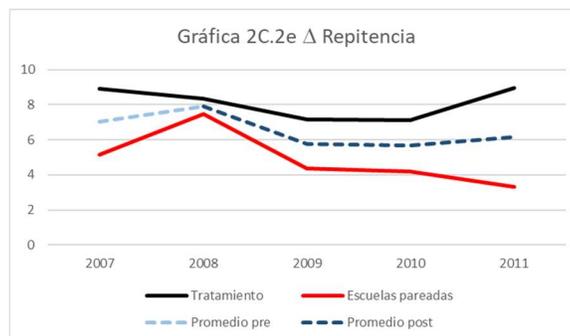
Fuente: cálculos propios.



Fuente: cálculos propios.

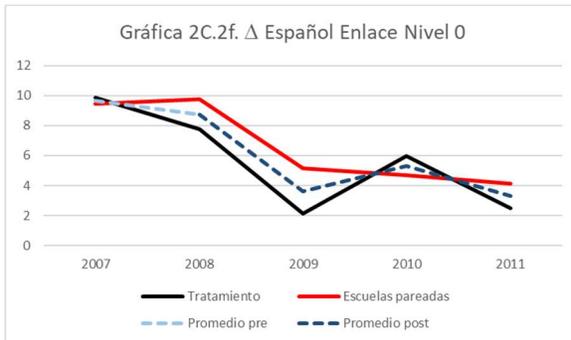


Fuente: cálculos propios.

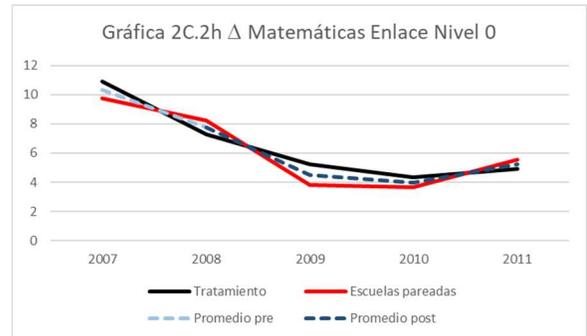


Fuente: cálculos propios.

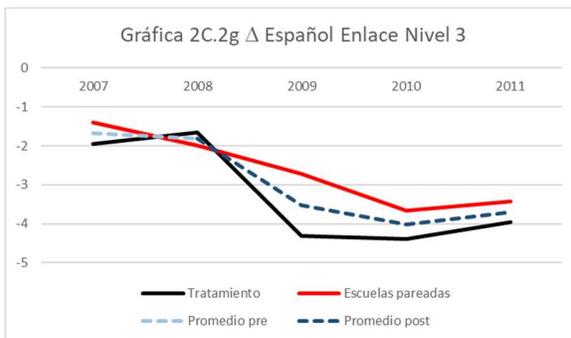
Gráfica 2C.2a-i. Diferencias en las tendencias entre grupos; cambios en resultados seleccionados (NN3). Escuelas primarias (Cont.)



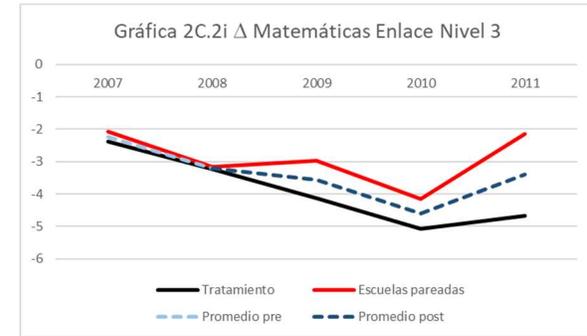
Fuente: cálculos propios.



Fuente: cálculos propios.



Fuente: cálculos propios.



Fuente: cálculos propios.

Tabla 2C.1. Diferencias en las medias entre el grupo de tratamiento y el grupo de control, obtenido mediante el apareamiento, previo al tratamiento

Variable	Tratamiento ¹	NN3		
		Control	Diferencia	Estadístico T
Δ Oportunidades	5.13 (0.44)	6.63 (0.99)	-1.50 (1.09)	-1.38* [0.08]
Δ Reprobación, primer grado	2.45 (0.69)	1.91 (0.38)	0.55 (0.79)	0.69 [0.24]
Δ Extraedad, primer grado	13.67 (0.91)	12.76 (0.87)	0.90 (1.26)	0.72 [0.24]
Δ Deserción, primer grado	-10.40 (2.82)	-7.86 (1.72)	-2.54 (3.30)	-0.77* [0.22]
Δ Repitencia, primer grado	8.15 (0.88)	7.02 (0.71)	1.13 (1.13)	1.00 [0.16]
Δ ENLACE español, nivel 0	7.76 (1.26)	6.13 (1.74)	1.62 (1.74)	0.93 [0.18]
Δ ENLACE español, nivel 3	-1.66 (0.46)	-1.22 (0.63)	-0.44 (0.78)	-0.56 [0.29]
Δ ENLACE matemáticas, nivel 0	7.26 (1.40)	6.91 (1.29)	0.35 (1.90)	0.18 [0.43]
Δ ENLACE matemáticas, nivel 3	-3.23 (0.56)	-3.42 (0.54)	0.19 (0.77)	0.24 [0.41]

Nota: entre paréntesis se presentan las desviaciones estándar, y entre corchetes la significancia de los estadísticos t. Fuente: cálculos propios.

Tabla 2C.2.1. Impactos estimados utilizando diferencias en diferencias (NN3); resultados seleccionados en escuelas primarias pareadas

Variables	2008		2009		2010		2011	
	Δ promedio	Δ promedio	Efecto DID	Δ promedio	Efecto DID	Δ promedio	Efecto DID	
<i>Oportunidades</i>	-1.50* (1.09)	-1.59* (0.87)	-0.09 (0.42)	-0.68 (0.93)	0.82 (1.16)	-0.27 (0.99)	1.22 (1.40)	
Reprobación, primer grado	0.55 (0.79)	-0.87 (0.56)	-1.42 (1.15)	2.05*** (0.67)	1.51 (1.67)	1.73*** (0.64)	1.18*** (0.37)	
Reprobación, segundo grado	0.11 (0.61)			0.83 (0.73)	0.72 (0.62)	0.82 (0.63)	0.70 (1.17)	
Reprobación, tercer grado	-0.59 (0.50)					0.55 (0.59)	1.14*** (0.31)	
Extraedad, primer grado	0.90 (1.26)	3.43** (1.06)	2.52 (1.65)	5.05*** (1.09)	4.14** (1.95)	6.22*** (1.07)	5.32* (3.16)	
Extraedad, segundo grado	4.36*** (1.32)			0.89 (1.13)	-3.46*** (1.31)	6.36*** (1.19)	2.00** (0.84)	
Extraedad, tercer grado	5.68*** (1.38)					3.02*** (1.20)	-2.66* (1.47)	
Deserción, primer grado	-2.54 (3.30)	0.24 (1.66)	2.78 (3.00)	-2.63* (1.92)	-0.09 (2.34)	-3.14* (2.33)	-0.60 (3.51)	
Deserción, segundo grado	1.49 (1.86)			0.44 (1.05)	-1.05 (2.19)	0.74 (1.19)	-0.76 (0.79)	
Deserción, tercer grado	2.72* (1.91)					0.37 (1.39)	-2.35 (1.81)	
Repitencia, primer grado	1.13 (1.13)	3.32*** (0.66)	2.19 (2.42)	4.23*** (0.78)	3.10* (1.87)	5.85*** (0.76)	4.72 (4.13)	
Repitencia, segundo grado	1.85*** (0.74)			-0.42 (0.55)	-2.27* (1.29)	1.40** (0.75)	-0.46 (1.06)	
Repitencia, tercer grado	2.96*** (0.81)					2.78*** (0.55)	-0.18 (0.93)	

Notas: entre paréntesis se presentan los errores estándar.

Códigos de significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01.

Fuente: cálculos propios con base en la Estadística 911 SEP.

Tabla 2C.2.2. Impactos estimados utilizando diferencias en diferencias (NN3); resultados seleccionados en las escuelas primarias pareadas

Variables	2008	2011	
	Δ promedio	Δ promedio	Efecto DID
<i>ENLACE</i> español, nivel 0	1.62 (1.74)	-1.37 (1.22)	-2.99** (1.37)
<i>ENLACE</i> español, nivel 3	-0.44 (0.78)	-1.21 (1.29)	-0.77 (3.83)
<i>ENLACE</i> matemáticas, nivel 0	0.35 (1.90)	0.44 (1.53)	0.09 (3.38)
<i>ENLACE</i> matemáticas, nivel 3	0.19 (0.77)	-2.29** (1.39)	-2.48 (3.14)

Notas: entre paréntesis se presentan los errores estándar.

Códigos de significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01.

Fuente: cálculos propios con base en la Estadística 911 SEP.

CAPÍTULO 3

Una intervención para mitigar la pérdida de habilidades de lectura en verano³⁵

“-Niñas -dijo Meg, gravemente, dirigiendo la mirada desde la cabeza desordenada a su lado hasta las cabecitas en el cuarto próximo-. Mamá desea que empecemos a leer, amar y acordarnos de estos libritos, y tenemos que comenzar inmediatamente. Solíamos hacerlo fielmente, pero desde que papá se marchó y con la pena de esta guerra, hemos descuidado muchas cosas. Pueden hacer lo que gusten, pero yo tendré mi libro aquí sobre la mesita, y todas las mañanas, en cuanto despierte, leeré un poquito, porque sé que me hará mucho bien y me ayudará durante todo el día.”

Louisa May Alcott. *Mujercitas*

3.1. Introducción

En México como en otros países, particularmente de América Latina, abordar el tema de la brecha en habilidades lectoras de niños y jóvenes es necesario para explicar el desempeño escolar, el aprendizaje, e incluso las oportunidades de acceso a niveles educativos superiores (Noguerol, 2002; Freire, 2006). El origen de estas diferencias se halla en factores multidimensionales. Aunque algunos de estos factores están condicionados por las circunstancias de nacimiento (como los genes, la familia, el nivel socioeconómico y el contexto cultural), otros son determinados por el sistema educativo; pertenecen al ámbito de competencia

³⁵ La intervención que se describe en este capítulo se llevó a cabo por el Centro de Investigación y Docencia Económicas (CIDE) en el año 2011. Este capítulo analiza la información proporcionada por el CIDE, y presenta una discusión de los resultados. Se agradece al Dr. Sergio Cárdenas Denham por facilitar las bases de datos que condujeron a este estudio.

de la autoridad de cada país y son propicios para mejorarse desde la esfera gubernamental. Este capítulo presenta una intervención, a partir de una aleatorización, que busca contrarrestar los efectos de lo que se conoce como pérdida de habilidades lectoras en verano (en adelante PHL) (Arlington *et. al.*, 2010). La PHL es un fenómeno que se presenta en el tránsito entre un ciclo escolar y el siguiente; periodo en que el alejamiento del aula tiene repercusiones sobre el desempeño en lectura de los alumnos. Existen dos problemáticas asociadas: la primera es el incremento de las brechas entre niños de diferentes estratos socioeconómicos. En particular, los efectos negativos de la PHL están concentrados entre los niños de entornos socioeconómicos más bajos; y generan rezagos que persisten en el tiempo (Entwisle, Alexander y Olson, 2007). La segunda es la necesidad de dedicar las primeras sesiones de cada ciclo escolar a recuperar estas habilidades.

3.1.1. Definición del problema

Investigaciones que se remontan a más de un siglo han dado cuenta de la caída en puntaje de lectura que presentan los alumnos de primaria antes y después del verano (White, 1906; Entwisle, Alexander y Olson, 2001). En Cooper *et. al.* (1996) se dimensionó la pérdida de habilidades en un mes de clases por cada año escolar, utilizando investigaciones del tema desde los ochenta. La revisión de literatura de este trabajo encontró además que -en el caso de alumnos de bajos ingresos- esta pérdida puede ser del orden de 3 meses y medio (p. 257) durante el verano.

La pérdida de conocimientos deriva de una falta de continuidad en el involucramiento de los menores en actividades pedagógicamente relevantes, y tiene un efecto diferenciado entre alumnos; siendo aquellos con desventajas socioeconómicas o en aprendizaje (o ambas) quienes típicamente tienen una pérdida mayor (Entwisle, Alexander y Olson, 2007, *op. cit.*).

En la literatura, las evaluaciones de impacto a intervenciones para paliar la PHL han tomado un lugar preponderante para dotar de evidencia robusta la efectividad de dichas intervenciones, y nutrir política pública que sea costo-efectiva; encontrando efectos mixtos.

Las evaluaciones sobre la efectividad de medidas para reducir la PHL en un inicio, se concentraron en alumnos en sus primeras etapas como lectores (véase Melosh, 2003). Por una parte, algunos trabajos como los de Kim (2007), Kim y Guryan (2010), o Wilkins *et. al.* (2012) encuentran que dichas medidas no son efectivas. Las explicaciones para la falta de efecto parten de la corta duración de las intervenciones que no logran sustituir a la etapa escolarizada. Adicionalmente, los libros no parecieron ser del nivel adecuado para que su lectura fomentara el aprendizaje, y el seguimiento al avance de los alumnos no permitió involucrarlos más; en especial cuando no hubo acompañamiento por parte de los padres de familia. En el caso de México, por ejemplo, el trabajo pionero de Ortega-Hesles (2012) mostró que la repercusión de la PHL sobre el desempeño educativo no ha sido suficientemente estudiada; con una consecuente carencia de estrategias específicas a nuestro contexto. Utilizando un diseño experimental para evaluar la entrega de libros a un grupo de niños de tercer y cuarto grado en el nivel de primaria, este trabajo generó dos grupos de tratamiento: uno que solo recibió los libros, y otro que adicionalmente recibió sesiones de capacitación para los padres de familia. Ninguno de éstos logró tener efectos estadísticamente significativos sobre comprensión lectora. Las posibles explicaciones incluyeron un tamaño de muestra insuficiente y un bajo periodo de exposición del tratamiento.

Por otro lado, los trabajos de Kim y White (2011), Kim (2006) y Kim y White (2008) encuentran efectos significativos de las intervenciones estudiadas. Estos estudios parten del hecho que, al menos en el caso de países como Estados Unidos y Canadá, el tratamiento remedial a la PHL a través de la entrega de libros resulta inefectiva -aún en el caso en que dichos libros sean elegidos

de acuerdo con el desempeño de cada alumno- sin acompañamiento adicional³⁶. Allington y McGill-Franzen (2010), por su parte, utilizó tratamientos recurrentes durante tres veranos y lo focalizó a poblaciones en desventaja, obteniendo un impacto significativo sobre el desempeño en comprensión lectora de los menores. Estas intervenciones mostraron tener un impacto positivo y mayor sobre el desempeño lector de niños en estratos de ingreso bajo (White *et. al.*, 2015; Guryan *et. al.*, 2015); cuyo impacto es más fácil de alcanzar cuando las intervenciones cuentan con un antecedente de efectividad. No obstante, sus resultados destacan la variabilidad en los efectos del tratamiento entre diferentes contextos escolares y familiares.

Las recomendaciones derivadas de estas investigaciones se centran en generar una discriminación positiva al diseñar estrategias para acercar a los niños a recursos con un contenido pedagógico importante fuera de las escuelas, así como incentivos a las familias para fomentar estos hábitos, particularmente entre aquellos con mayores dificultades socioeconómicas.

3.1.2. Una intervención en verano para mejorar el desempeño en lectura

La intervención que aquí se describe estuvo destinado a mitigar la PHL; con el objetivo de ayudar a orientar política pública e incidir positivamente sobre el aprendizaje de los alumnos que presentan mayores rezagos educativos. Para explorar cómo incidir en las habilidades de los niños, se condujo una intervención en un municipio del estado de Aguascalientes, en México. Dicha intervención estuvo orientada a mejorar el desempeño lector en los estudiantes de primaria en nuestro país. La hipótesis de trabajo es que aminorar la pérdida de contacto con los libros y actividades escolares durante el verano, y concientizar a los padres sobre los beneficios

³⁶ Lo que la investigación de Meichenbaum y Biemiller (1998; citada en Kim y White, 2008) conoce como *scaffolding*, y define como “la práctica de proveer suficiente apoyo (no demasiado ni muy poco) para ayudar a que los estudiantes tengan éxito” (P. 3; traducción propia). Para un estudio más profundo de las teorías sobre acompañamiento véase Wood, Bruner y Ross (1976).

de promover la lectura entre sus hijos mitigaría la PHL. Recopilando información en 15 escuelas del municipio de Jesús María, con la colaboración de la Secretaría de Educación de Aguascalientes, se obtuvieron las direcciones de 1075 alumnos de quinto grado identificados en dos turnos de los 15 planteles seleccionados, asignándose aleatoriamente 50% al grupo de tratamiento y 50% al grupo de control, llevando a cabo la aleatorización a nivel de alumno, a fin de minimizar las diferencias observables y no observables. Se hizo llegar una solicitud para incorporarse al grupo de tratamiento acudiendo al domicilio de cada uno de los alumnos sorteados, a fin de no contaminar el grupo de control.

La muestra se seleccionó a finales del ciclo escolar 2010-2011. Los hogares se visitaron el mes anterior al cierre del ciclo y la intervención se realizó durante los meses de verano de 2011. Los cuestionarios de seguimiento fueron aplicados durante el mes de octubre.

A los alumnos en el grupo de tratamiento se les entregaron tres libros y, de forma adicional, se les proveyeron sesiones de capacitación y un folleto informativo a los padres de familia de los menores en el grupo de tratamiento. El folleto resaltó la importancia de fomentar el hábito de la lectura entre los menores. El impacto sobre desempeño pre y post se hizo sobre la prueba ENLACE de español, cuya principal característica es la precisión de su medición.

Del total, la tasa de atrición fue del 25.04%. La muestra de trabajo consistió en 805 alumnos, que formaron parte de la línea basal. La inversión tuvo un costo total de \$150,000 pesos, más el pago al personal del CIDE y sin contar el valor del trabajo de profesores voluntarios de Aguascalientes que ayudaron en la intervención.³⁷ Los impactos -aunque con el sentido esperado (0.72 puntos porcentuales)- no resultaron estadísticamente significativos. Entre las posibles

³⁷ A quienes se hace extensivo el agradecimiento, en reconocimiento al esfuerzo que realizaron adicional a su noble labor.

explicaciones se encuentra un tamaño de muestra insuficiente para identificar con precisión los efectos o un periodo de exposición que no fue de bastante duración.

El presente capítulo se organiza de la siguiente manera: la sección 3.2 presenta el marco teórico en el que se circunscribe la intervención evaluada, a partir de la teoría sobre la pérdida de las habilidades lectoras en verano. En la sección 3.3 se presenta la justificación para llevar a cabo un estudio de esta naturaleza, y se revisan las investigaciones -realizadas mediante el uso de métodos experimentales y cuasiexperimentales- más recientes que han buscado contrarrestar los efectos de la PHL. La sección 3.4, a su vez, presenta la metodología seguida para llevar a cabo la intervención motivo de este estudio, incluyendo: participantes, diseño y línea basal, instrumentos de medición, implementación del experimento y el análisis de diferencias-en-diferencias llevado a cabo. La sección 3.5 presenta los resultados, y la sección 3.6 las conclusiones y recomendaciones de política pública.

3.2. Revisión de literatura

3.2.1. La pérdida de habilidades lectoras

El fenómeno de la PHL tiene más de cuatro décadas de ser analizado. Las primeras investigaciones (Hayes y Grether, 1969) comenzaron a estudiar las diferencias de desempeño entre escuelas derivadas de los periodos vacacionales. No obstante, fue la investigación pionera de Heyns (1978) la que acuñó el término *Pérdida de habilidades lectoras en verano* (Allington y McGill-Franzen, 2003) para describir el declive que los alumnos presentan en sus habilidades de lectura durante el tránsito entre un ciclo escolar y el siguiente. Este trabajo argumentó que la lectura era la actividad más fuertemente relacionada con el aprendizaje durante el periodo de verano, impactando sobre el desempeño de los alumnos y las habilidades de los menores en dicho periodo. Heyns (1978, *op. cit.*) halló que las brechas entre alumnos con rezagos en aprendizaje y alumnos aventajados crecía sustantivamente durante las vacaciones entre ciclos

escolares, cuando los factores extraescolares ejercían mayor influencia que los propios del sistema educativo. La mayor aportación de Heyns (*op. cit.*, 1978 y 1987) es presentar una conceptualización de “parámetros educativos” (Bro, Downey y von Hippel, 2004) para describir el efecto del periodo no escolarizado sobre el desempeño en lectura de los menores. El efecto acumulativo de esta pérdida se estimó en el equivalente a 2.8 años menos de aprendizaje escolarizado al final de sexto grado (Hayes y Grether, *op. cit.*), generando un debate en la academia sobre el sentido y la magnitud de esta diferencia.

El efecto sobre aprendizaje en los meses de verano fue estudiado a mayor profundidad en Entwisle, Alexander y Olson (2001, *op. cit.*), trabajo en el que se postuló la llamada “teoría del grifo” (*Faucet Theory*) para describir la forma en que el flujo de recursos pedagógicos fluye a lo largo del periodo escolarizado, pero disminuye significativamente durante el periodo vacacional, siendo el factor de mayor relevancia para explicar el rezago de unos alumnos frente a otros. En particular, este documento señala que el nivel socioeconómico de los hogares es determinante para compensar los periodos cuando el flujo de recursos del “grifo escolar” permanece cerrado (*ibid.*).

Los patrones de rezago educativo, y en particular lector, se extienden -para el caso de Estados Unidos- a través de diferencias culturales, lingüísticas y raciales (Neuman y Celano, 2001). No obstante, ninguna distinción analítica tiene un impacto mayor en las diferencias entre grupos que el nivel socioeconómico. Una primera explicación de estas diferencias se encuentra en el acceso de los niños a material escrito propio para su edad. En Neuman y Celano (2006 y 2012) se reconoce la insuficiencia de la disponibilidad; ya que las diferencias socioculturales entre hogares condicionan la forma en que se aprovechan los recursos lectores entre los menores durante los periodos fuera de la escuela. De hecho, la investigación de estas autoras encuentra que los nuevos recursos pedagógicos son utilizados con una mayor orientación al aprendizaje

entre los alumnos más favorecidos, incrementando aún más las brechas de aprendizaje entre unos alumnos y otros (Neuman y Celano, 2006, *op. cit.* pp. 189-190).

La investigación de Entwisle, Alexander y Olson (2007), a su vez, recopila datos sobre desempeño de una cohorte de niños entre primero de primaria y tercero de secundaria. Este trabajo encontró que las ganancias en desempeño académico se deben principalmente al aprendizaje durante el año escolar. Sin embargo, las brechas en dicho desempeño resultan también de las diferencias generadas durante las vacaciones de verano.

La evidencia para Estados Unidos apunta a que el ámbito donde se genera la mayor brecha en el desempeño entre alumnos es el hogar, y que el verano es el periodo de mayor acentuación de esta desigualdad. Como señala el estudio de Allington y McGill-Franzen (2013) “La evidencia disponible señala a la pérdida de habilidades de lectura en el verano como la explicación más poderosa del aumento en la brecha de desempeño en lectura entre niños pobres y ricos durante primaria y secundaria” (P. 7; traducción propia). Evidencia para el caso de países como el nuestro, hispanohablante, apunta directamente al papel fundamental que la edad tiene sobre el desarrollo lector; que tiende a lograr una madurez en la lectura de pseudopalabras entre los 8 y 9 años, con una velocidad que está directamente relacionada a la capacidad de concentración, atención y conocimiento del vocabulario (Rosselli, Matute y Ardilac, 2006). Para fortalecer estas habilidades, el trabajo de Gómez y Silas (2012) apunta a los ejercicios de comprensión lectora como intervenciones de alta efectividad en el caso específico de México.

3.2.1.1. La pérdida de habilidades lectoras en México

Uno de los principales determinantes del hábito de leer entre la población mexicana es, de acuerdo con datos de la Encuesta Nacional de Lectura (Conaculta, 2018), “disfrutar mucho” la lectura. Este factor incide en que la población lea 80% más libros frente al promedio nacional (p. 6). Si bien esta actividad está fuertemente marcada -entre los niños y jóvenes- por las

diferencias económicas³⁸, el acceso a internet y a materiales impresos³⁹, el efecto que tienen los hábitos familiares y las motivaciones en torno a esta actividad son muy poderosos.

El hogar es un ámbito en el que la lectura resulta indispensable para fortalecer las habilidades de los menores. Investigaciones internacionales asocian fuertemente el hábito de los padres de leer a sus hijos en voz alta -y la lectura que se lleva a cabo al menos tres veces por semana a los menores en el hogar- con un incremento del doble de probabilidad para los niños de obtener un puntaje en el 25% superior de la distribución de calificaciones en su cohorte (Denton, Kristen y West, 2002).

Tener acceso a una amplia gama de materiales de lectura en el hogar ha probado ser una variable crítica que afecta la adquisición de habilidades (McQuillian, 1998). En ese sentido, la falta de materiales adecuados para fomentar las habilidades lectoras de niños y jóvenes se convierte en un obstáculo más; relacionado a los ingresos del hogar y el entorno donde los menores se desarrollan (Evans *et. al.*, 2010). A estas carencias se suman las propias del acceso a las nuevas tecnologías de la información; y por ende a los materiales escritos disponibles en éstas (Horrigan, 2015).

De ello resulta indispensable poner en contacto a los lectores con textos que les resulten atractivos y sean adecuados a su nivel de lectura. Este aspecto, de acuerdo con diversas investigaciones (Durkin, 1993; Carver y Leibert, 1995; O'Connor *et. al.*, 2005), es crucial para mejorar las habilidades. Por un lado, los textos deben ofrecer un nivel suficiente para que su

³⁸ La Encuesta Nacional de Lectura 2015 reporta que el 15.6% de la población con ingresos familiares menores o iguales a \$1,199 mensuales lee libros; comparado con el 40.6% de la población con ingresos iguales o superiores a \$11,600 o más. De acuerdo con información de la ENIGH, los primeros formarían parte del primer y segundo decil de ingresos (antes de transferencias), mientras que los segundos corresponderían al segmento superior del decil 8, y los deciles 9 y 10 (nuevamente, antes de transferencias) (ENA; cálculos propios).

³⁹ Los encuestados pertenecientes al estrato más bajo establecido en la ENA 2015 reportan tener 19.3 libros, en promedio, en el hogar (distintos a libros de texto), mientras que aquellos en el grupo de ingresos más alto reportan tener, en promedio, 168.5 libros. De igual manera, para el primer grupo el número de libros en formato digital es nulo, mientras que para el grupo de mayores ingresos es del orden de 61.9, en promedio.

consumo resulte provechoso; contribuyendo a enriquecer el vocabulario y el procesamiento de información en el joven lector. Pero, por otro lado, deben evitar ser demasiado áridos o complejos para el nivel de lectura de quien los lee, de tal suerte que no generen una pérdida de motivación hacia el texto y, en última instancia, hacia la lectura en su conjunto (Hunt, 1996; citado en Melosh, *op. cit.*).

3.2.2. Experimentos para mitigar la pérdida de habilidades lectoras

Debido a su capacidad para aislar los efectos de las intervenciones del resto de factores que pudieron influir en los resultados, la evidencia de los diseños de evaluación de impacto ha ido creciendo en relevancia para estimar el efecto de intervenciones para mitigar la PHL. En Melosh (2003) se estudia el efecto de la entrega de libros durante el verano sobre alumnos de familias con ingresos bajos; pertenecientes al segundo grado en el nivel de primaria. Se les evaluó a estos alumnos en medidas de fluidez (tasa, precisión y automaticidad), comprensión y motivación hacia la lectura. No hubo diferencias estadísticamente significativas para la fluidez lectora; pero sí para el autoconcepto de los alumnos como lectores. Los resultados aproximaron significancia estadística para comprensión lectora en el grupo que recibió al menos el 70% de todas las entregas de libros durante el periodo de la intervención. El bajo tamaño de la muestra (N=46) comprometió la validez estadística de los resultados.

La investigación de Kim (2006), por su parte, estudia los efectos de una intervención de verano a niños de cuarto grado en el nivel de primaria en Estados Unidos. Utilizando una muestra de 552 alumnos en 10 escuelas, y un solo grado escolar, el estudio encontró efectos positivos y significativos -sobre el puntaje de un examen estandarizado sobre desempeño lector⁴⁰- de la entrega de libros y motivación a la lectura sobre grupos minoritarios que se encuentran en

⁴⁰ El *Iowa Test of Basic Skills* es un examen estandarizado sobre habilidad en la lectura que se normó a nivel nacional en Estados Unidos a partir del año 2000.

desventaja. Además, halló un incremento en el tiempo que los niños leen con su familia, a partir del experimento. La conclusión de esta investigación es que una intervención de lectura en verano puede ser una política escalable y costo-efectiva para mejorar el desempeño lector entre alumnos de bajo desempeño.

Con el objetivo de contrastar la hipótesis sobre dónde concentrar los esfuerzos para lograr intervenciones más efectivas -que mejoren el desempeño lector de los niños- en Kim (2007; citado en Kim y Gurian, 2010, p. 7) se estudia, a través de un diseño experimental, el efecto sobre comprensión lectora que tiene entregar libros a menores desde primero hasta quinto grado en el nivel de primaria. A pesar de no contar con una muestra suficientemente grande para obtener poder estadístico, los resultados dieron cuenta de que el mayor efecto se concentra entre los niños que cursan los grados superiores (pp. 18-19). Kim (2007) apoya la hipótesis de que una intervención de lectura en verano puede favorecer el desempeño lector de alumnos que se encuentran en los últimos grados de primaria. Este trabajo utiliza una prueba estandarizada de desempeño lector, una encuesta de actitudes hacia la lectura y una encuesta de preferencias de lectura. Aunque el grupo de tratamiento reportó leer más libros y participar en más actividades de lectura que el grupo de control, no hubo una diferencia estadísticamente significativa en desempeño en lectura.

En el trabajo de Kim y White (2008) se lleva a cabo un estudio con la participación de 400 alumnos y 24 profesores de tercero a quinto grado en el nivel de primaria, asignados aleatoriamente a una de las siguientes cuatro condiciones experimentales: 1) control, 2) recepción de libros, 3) libros y acompañamiento en lectura en voz alta, y 4) libros con acompañamiento en lectura en voz alta y comprensión. Los alumnos se evaluaron en medidas pre y post de fluidez en lectura en voz alta y lectura en voz baja. Los resultados mostraron que los alumnos que recibieron libros acompañados de apoyo para lectura en voz alta y comprensión

tuvieron un puntaje significativamente mayor tanto a los niños en el grupo de control como a los niños que únicamente recibieron los libros (pp. 16-17). Asimismo, los alumnos en los dos grupos de acompañamiento combinados tuvieron un puntaje significativamente mayor al final de la intervención que los alumnos que estuvieron tanto en el grupo de control como en el que únicamente recibió los libros.

En Seward (2009) la intervención de verano consistió en un campamento de verano para niños -con duración de dos semanas- y dos sesiones de capacitación para padres. La muestra incluye alumnos en el grupo de control, alumnos que solo participaron en el campamento, alumnos que participaron en el campamento y cuyos padres estuvieron en las sesiones de capacitación y alumnos que no estuvieron en el campamento, pero cuyos padres sí estuvieron en las sesiones de capacitación. Todos los alumnos fueron identificados como población con dificultades en atención fonológica y habilidades lectoras al final de educación preescolar. Adicionalmente, se eligió otro grupo de control consistente en alumnos de la misma cohorte que fueron identificados como de desempeño normal en habilidades lectoras. Las intervenciones buscaron mejorar las habilidades de lectura proveyendo enseñanza explícita en atención fonológica, así como oportunidades para practicar otras habilidades de lectura emergentes. Los alumnos en los tres grupos intervenidos mostraron un cambio positivo en atención fonológica y lectura de palabras comparado con el grupo de control con las mismas habilidades iniciales. Los alumnos que recibieron ambos componentes de la intervención mostraron la mayor mejoría en los exámenes de atención fonológica y habilidad de lectura de palabras, de acuerdo con la evaluación de seguimiento que se realizó al final de primer grado.

En Allington y McGill-Franzen (2010) se utilizó una muestra de 852 estudiantes de 17 escuelas en zonas de alta marginación y se otorgaron libros que los menores eligieron al final del ciclo escolar durante tres periodos consecutivos. Se compararon con 478 estudiantes de las mismas

escuelas que no recibieron libros y sirvieron como grupo de control. Los resultados de la evaluación en lectura indican un efecto estadísticamente significativo de proveer acceso a libros para su lectura en verano al igual que un tamaño del efecto significativo. Se encontraron hallazgos más grandes cuando se comparó el desempeño de los alumnos en mayor desventaja económica tanto en el grupo de tratamiento como en el de control.

El trabajo de Butler (2010) mide la efectividad de tres programas de verano sobre el desempeño en lectura de aprendices del idioma inglés (*english language learners*) y alumnos en desventaja académica que hablaban inglés entre segundo y cuarto grado en el nivel de primaria. Los grupos experimentales incluyeron 1) estudiantes que recibieron una visita semanal a sus hogares y libros de parte del personal de la escuela, 2) alumnos que llevaron diez libros al hogar durante el verano y 3) un grupo de control. Los resultados muestran que ambos grupos de tratamiento tuvieron ganancias estadísticamente significativas frente al grupo de control tanto en fluidez oral de lectura como en repeticiones verbales. No se encontraron diferencias significativas entre aprendices del idioma inglés y alumnos con desventaja académica, lo que sugiere que el inglés como primer idioma no fue un factor en el éxito del programa.

El trabajo de Kim y Guryan (2010, *op. cit.*) realiza una intervención de verano enfocada en alumnos que transitaban entre el cuarto y el quinto grado en el nivel de primaria. La investigación consistió en una línea base de 370 niños que fueron evaluados a través de una prueba de comprensión lectora y vocabulario y asignados aleatoriamente a 1) un grupo de tratamiento que recibió 10 libros autoseleccionados durante las vacaciones de verano, 2) un grupo de aprendizaje lector en familia que recibió 10 libros autoseleccionados durante las vacaciones de verano y cuyos padres fueron invitados a atender 3 eventos, de 2 horas cada uno, relacionados con el aprendizaje lector y 3) un grupo de control. Es importante destacar que este componente de capacitación a padres, para evaluar mejor la capacidad lectora de sus hijos, se

llevó a cabo en un contexto donde casi todas las familias estudiadas no tenían el inglés (idioma evaluado) como lengua principal en sus hogares. A pesar de que los alumnos -en el grupo de tratamiento que recibió las sesiones- reportaron leer más libros que aquellos en el grupo de control, no hubo un efecto significativo sobre comprensión lectora o vocabulario. Incrementó las experiencias lectoras de los menores, pero las sesiones no resultaron de impacto sobre los indicadores (P. 28).

En Pagan (2010) se prueba si una intervención de lectura en verano mejoraría las medidas de desempeño lector de una muestra de 57 alumnos de bajo desempeño entre tercero y quinto de primaria. Los alumnos fueron aleatoriamente distribuidos entre un grupo de lectura de libros y un grupo de control. Durante 8 semanas los niños en el grupo de tratamiento recibieron un libro que estuviera empatado con sus intereses y nivel de lectura. Adicionalmente, los padres de familia actuaron como acompañamiento y recibieron entrenamiento sobre cómo usar estrategias efectivas de lectura. De esta manera, los alumnos en el grupo que recibió libros mostraron ganancias significativas en desempeño lector, comparados con los niños en el grupo de control. No se hallaron diferencias para habilidades verbales ni para motivación al leer.

El trabajo de Van Andel (2011) estudia -a partir de una muestra de alumnos de segundo a quinto grado, entre 7 y 11 años- los efectos de un programa de lectura en biblioteca durante los meses de julio y agosto. El diseño cuasiexperimental consistió en 16 alumnos que participaron en el programa, comparados con un grupo de 53 alumnos que no participó. Los resultados muestran que aquellos que participaron en el programa tuvieron puntajes más altos que el grupo de control en comprensión lectora, mas no en otros componentes de lectura, y que la lectura de alta frecuencia durante el verano estuvo asociada con un incremento en sus habilidades.

En Wilkins *et. al.* (2012) se investigan los efectos de proveer libros a los estudiantes, aunado a enviarles postales de recordatorio durante el verano, pero sin otro tipo de componente de

motivación o instrucción docente. Se utilizó un diseño experimental para probar la efectividad de un programa de lectura en verano para mejorar la comprensión lectora de estudiantes en tercer grado, en el nivel de primaria, en precariedad económica cuyo nivel de lectura estuvo debajo del percentil 50 en pruebas estandarizadas a nivel nacional. Cada estudiante en el grupo de tratamiento recibió una única entrega de ocho libros de acuerdo con su nivel de lectura y área de interés al principio del verano, seguido de una postal cada semana por seis semanas. Los libros siguieron una metodología econométrica para asignarse a cada menor de acuerdo con su nivel de desempeño lector. El análisis confirmatorio halló que el programa de verano no tuvo un impacto estadísticamente significativo sobre la comprensión lectora de los menores.

En México, el trabajo pionero de Ortega-Hesles (2012) da cuenta que la repercusión de la PHL sobre el desempeño educativo no ha sido suficientemente estudiada; con una consecuente carencia de estrategias específicas a nuestro contexto. Utilizando un diseño experimental para evaluar la entrega de libros a un grupo de niños de tercer y cuarto grado en el nivel de primaria, este trabajo generó dos grupos de tratamiento: uno que solo recibió los libros, y otro que adicionalmente recibió sesiones de capacitación para los padres de familia. Ninguno de éstos logró tener efectos estadísticamente significativos; aunque sí con el sentido esperado.

Los últimos trabajos revisados en el presente estudio fueron los de Guryan *et. al.* (2015) White *et. al.* (2015) y Kim *et. al.* (2016), que revisan y dan seguimiento a las poblaciones producto de un diseño experimental de alumnos de tercer grado en el nivel de primaria, intervenidos con un programa para mitigar la PHL. La característica distintiva de este experimento en sus diferentes etapas es que proveyó libros adecuados al nivel de lectura de cada menor y que además les resultaban atractivos; asimismo, brindó acompañamiento a los profesores a través de una capacitación al final del ciclo escolar. Se generó incluso un subgrupo de tratamiento que ofreció llamadas de seguimiento a los padres de familia para conocer el avance de los menores.

Adicionalmente a la presente revisión, y siguiendo el ejemplo de Kim y Quinn (2014), se llevó a cabo un análisis de sensibilidad de los diseños de evaluación de impacto que componen este apartado. Debido a la heterogeneidad de variables que miden el tamaño del efecto, se optó por computar una medida estándar del tamaño del efecto a través de la d de Cohen, siguiendo el ejemplo del trabajo antes citado. Los resultados se presentan como un ejercicio adicional en el Anexo 3A.

3.3. Metodología

3.3.1. Diseño del estudio

3.3.1.1. Aleatorización y recolección de los datos

Diseñar experimentos en ciencias sociales desde la academia presenta un conjunto de retos de origen. En principio, se necesita contar con el apoyo y venia de las autoridades civiles que se vean involucradas en el ámbito de sus competencias y/o funciones; a fin de no transgredirlas. Además, debe contarse con recursos suficientes para aplicar el tratamiento a un grupo numeroso; lo que -para diseños experimentales con suficiente poder estadístico- implica recurrir a gastos cuantiosos.

Por estas y otras limitaciones, es que diversos autores (p. ej. Baker, 2000; Duflo, 2007; Gertler *et. al.*, 2011; Shadish *et. al.*, 2002) recomiendan implementar diseños experimentales con una base teórica sólida; cuya efectividad ya haya sido probada en contextos similares o, al menos, parta de una teoría de causalidad robusta.

La hipótesis de trabajo es que la PHL en México se debe, al menos en parte, a la pérdida de contacto con los libros y actividades escolares. A partir de esta premisa se diseñó la intervención que se describe a continuación.

En principio, se eligió el quinto grado en el nivel de primaria; permitiendo que los alumnos hayan transitado ya por todas las etapas básicas de aprendizaje de la lengua, y se encuentren en

el periodo de reforzamiento de sus habilidades. De ahí que se espere que la exposición a los textos efectivamente mejore su desempeño lector.

Para la selección a los grupos de tratamiento y control se recurrió al generador de números aleatorios de RANDOM.ORG[®], aplicado a una base de datos con el total de niños en la muestra. Por tanto, los impactos estimados representan la intención a tratar del programa (*intent-to-treat*) pues, aunque un alumno haya sido asignado al tratamiento, la familia o el alumno pueden optar por no participar. En los siguientes tres apartados se describen, respectivamente, los participantes, los instrumentos, y el diseño de esta intervención.

3.3.1.2. Participantes

Autoridades educativas

Con el propósito de realizar la intervención se colaboró con la agencia municipal de Jesús María, en Aguascalientes. El primer paso fue pedir a la Secretaría de Educación de Aguascalientes facilitar la información referida a los alumnos de quinto grado de las primarias de uno de sus municipios. Inicialmente, dicha información se solicitó a través del portal de transparencia del estado. Tras obtener una negativa -por tratarse de datos personales de carácter confidencial- personal del CIDE acudió personalmente a obtener la venia de las autoridades, explicando la naturaleza de la investigación, así como el manejo confidencial y sin difusión que se haría de todos los datos personales. También se especificó que no se revelaría ninguna cifra por alumno, sino solo agregados estadísticos. Finalmente, se obtuvo permiso para llevar a cabo el estudio, facilitando la información de la prueba ENLACE y canalizando al equipo de trabajo con las autoridades educativas de cada escuela.

A partir de ese momento, se procedió a recopilar información de los menores seleccionados. En el caso de aquellos que formaron parte del grupo de tratamiento, se les solicitó su teléfono y

dirección acudiendo físicamente a cada plantel, y entregando la carta solicitud que se muestra en el Anexo 3B.

Primero se obtuvieron los datos de los alumnos pertenecientes a 15 planteles, y posteriormente se llevó a cabo la aleatorización de estos alumnos a los grupos de tratamiento y control a partir de una base de datos construida con su nombre y puntaje de la prueba ENLACE a nivel de salón de clases.

Padres de familia

En un principio, la intervención se dio a conocer únicamente a los padres de los alumnos en el grupo de tratamiento; haciéndoles llegar a su domicilio la solicitud para participar. Realizada la intervención, a todos los padres de familia se les envió un cuestionario, por conducto de los alumnos, para recopilar información sobre los hábitos de lectura de sus hijos en el verano. Adicionalmente, a la familia de los alumnos en el grupo de tratamiento se les hizo llegar un cuestionario para estudiar las percepciones y grado de cumplimiento de las sesiones de lectura durante las vacaciones.

Al finalizar la investigación, se entregó de forma gratuita el paquete de libros utilizado también a los niños en el grupo de control. Se informó a los padres de familia de los motivos de la entrega, y se les dio libertad tanto para recibir los libros, como para excluir los resultados de sus hijos del estudio (aclarando la independencia de estos dos eventos).

Alumnos

Los menores fueron asignados de forma aleatoria al grupo de tratamiento (308) y al de control (497). Del total de la muestra el 51.43% son niñas. La información fue levantada entre agosto y octubre de 2011.

3.3.1.3. Instrumentos

A fin de evaluar la implementación y los efectos de la intervención, se recurrió a una batería de instrumentos tanto para los alumnos como para los padres. A continuación, se detalla cada uno de los materiales que constituyeron la aplicación y medición del tratamiento.

a) *Prueba de lectura ENLACE*. Es un examen nacional, efectuado en el mes de abril en la llamada *Semana Nacional de Evaluación*. Comprende del tercer al sexto grado en el nivel de primaria; y de primero a tercer grado en el nivel de secundaria. Evalúa los contenidos curriculares de español y matemáticas, además de una materia adicional que varía año con año, pero se repite cíclicamente. Para establecer la línea basal, se utilizó la batería de preguntas del examen correspondiente a la asignatura de español; mientras que los efectos del programa fueron capturados aplicando un examen de español de quinto grado de ciclos escolares previos. Ambos exámenes fueron estandarizados a fin de que los puntajes pudieran ser comparables.

b) *Folleto informativo sobre la importancia de la lectura*. Contiene recomendaciones para acompañar e interactuar con los niños en el periodo de lectura. Se entregó a los padres de familia en el grupo de tratamiento durante la visita a sus hogares.

c) *Cuestionario de hábitos de lectura de alumnos*. Es un instrumento diseñado para la investigación. Consiste en una batería de preguntas sobre hábitos del menor, enfocados hacia la lectura durante el verano.

d) *Cuestionario de padres sobre hábitos de lectura de sus hijos*. Incluye, además de los hábitos percibidos del menor en el hogar y el apoyo de los padres a la lectura, una sección con variables de control sobre nivel socioeconómico.

Financiamiento

La intervención que aquí se describe fue financiada por el Centro de Investigación y Docencia Económicas (CIDE). Consistió en la entrega de 3 libros⁴¹ a los alumnos para su lectura durante las vacaciones de verano, acompañada de sesiones de orientación a padres de familia. Las lecturas fueron seleccionadas del catálogo del *Programa Nacional de Lectura*⁴², y acordes a la edad de los menores. El costo por libro fue de aproximadamente 40 pesos, representando una inversión por alumno de alrededor de \$150 pesos, incluyendo el material de las sesiones de capacitación.

Para la aplicación de los cuestionarios y de los exámenes se contó con profesores voluntarios del estado de Aguascalientes, así como personal del CIDE.

3.3.2. Metodología de evaluación y línea basal

La evaluación de impacto partió de la base de un diseño experimental. La principal ventaja de utilizar este diseño es que resulta el más efectivo para minimizar el sesgo de selección. La Figura 3.1 muestra la línea de tiempo de esta intervención.

Más allá de las dificultades encontradas en la implementación de diseños experimentales (Murnane y Willett, 2011) se optó por llevar a cabo una intervención de pequeña escala a fin de estudiar el impacto potencial.

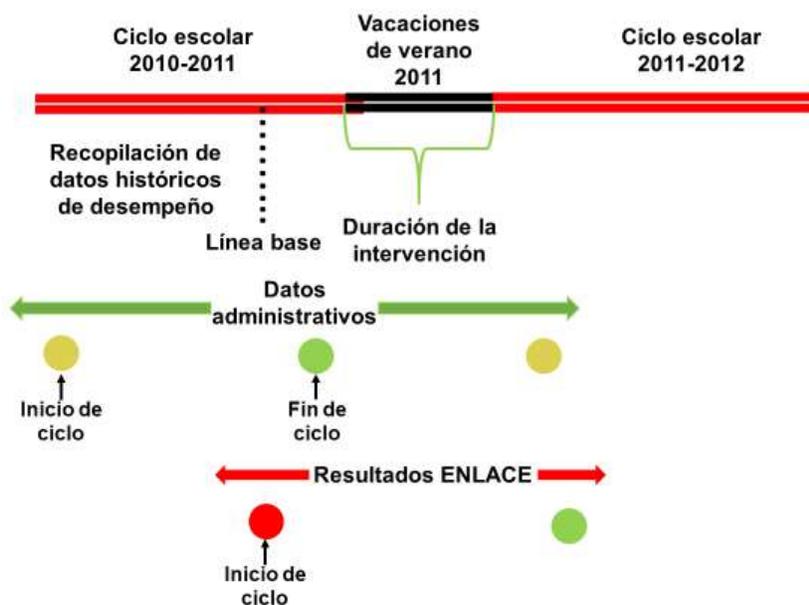
En julio de 2011, con el apoyo de las autoridades del estado de Aguascalientes, se seleccionaron aleatoriamente 15 escuelas de doble turno de uno de los municipios de la entidad, con la venia de los directores de dichos planteles; eligiendo también de forma aleatoria a los alumnos dentro de cada salón de clases de quinto grado en el nivel de primaria.

⁴¹ Los libros fueron *Crispín el tuerto*; *El enmascarado de lata* y *El curro*.

⁴² Como se describe en Hesles (2012, *op. cit.*, p. 11) el *Programa Nacional de Lectura* “provee de librerías escolares a todas las escuelas públicas de México”.

A fin de no contaminar el experimento, los alumnos que resultaron sorteados para pertenecer al grupo de tratamiento fueron contactados directamente en sus hogares -nuevamente, previa autorización de la Secretaría de Educación del estado- y con una total secrecía respecto a sus datos de contacto.

Figura 3.1. Línea de tiempo del experimento



Fuente: elaboración propia

Para no afectar el comportamiento de los niños en el grupo de control, la intervención se realizó llevando los libros -para los alumnos que resultaron sorteados- hasta sus hogares. Asimismo, la invitación a participar tanto a ellos como a sus padres se les hizo llegar a su domicilio. Toda la información personal de los menores fue obtenida con el consentimiento de la autoridad educativa del estado; manteniéndose confidencial y comprometiendo su destrucción al terminar el estudio.

Dada la evidencia presente en la literatura, se optó por utilizar la estrategia más efectiva encontrada en Hesles (*Ibidem*, p. 18). Para ello, se acompañó la entrega de libros de un folleto

dirigido a los padres; donde se explicaba la importancia de acompañar a sus hijos en la lectura, proveer apoyo, etc. Además, el equipo encargado de la entrega de libros ofreció una breve capacitación en las visitas a domicilio, como se señalaba en la hoja de registro que dichos capacitadores llenaron, mostrada en el Anexo 3C.

Para la línea basal, se obtuvieron los puntajes de la prueba ENLACE de español del mes de abril de 2011, a fin de mostrar que ambos grupos eran similares antes de la intervención. La entrega de libros e información en los hogares de los alumnos tratados se llevó a cabo entre los meses de julio y agosto. Finalmente, para medir la efectividad del tratamiento, se aplicaron a principios del mes de octubre cuestionarios a padres y alumnos; así como un nuevo examen.

Entre el 24 y el 26 de octubre de 2011, se aplicó en las aulas la segunda prueba ENLACE de español y el cuestionario de contexto de alumnos, ahora en sexto grado; además se hizo entrega del cuestionario de padres para que fuera contestado y devuelto al día siguiente. Estos indicadores son usados para medir el impacto de la intervención.

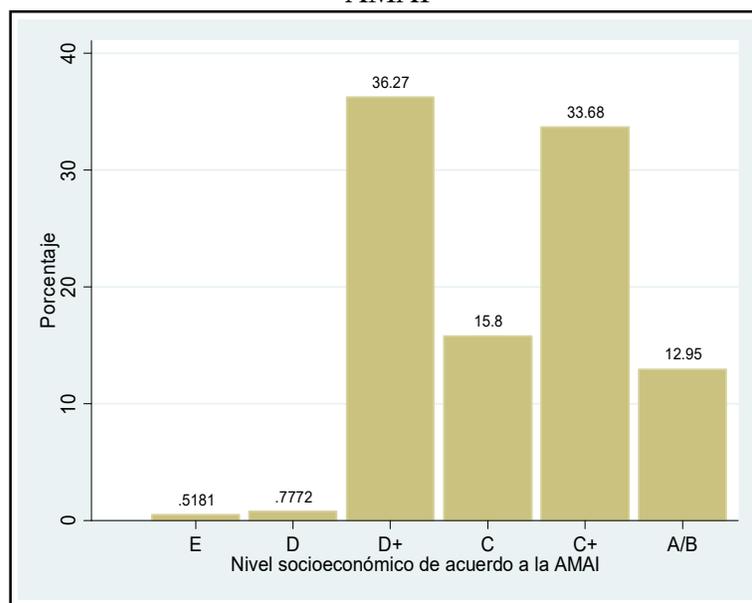
Como se mencionó anteriormente, la evidencia recopilada apunta a que, en México, el principal motor para fomentar la lectura entre los más jóvenes es su componente recreativo; y está marcado por el acceso a materiales de lectura en el hogar y la escuela. En el caso de la población intervenida, la información específica -correspondiente a sus hábitos y contexto- fue capturada a través de las dos entrevistas levantadas al final del experimento. Del *Cuestionario de hábitos de lectura de alumnos* se obtuvo información relevante respecto a la aproximación que tienen los menores hacia la lectura y el contexto en que la desarrollan; en particular durante las vacaciones. Consta de 980 observaciones, tanto de alumnos en el grupo de tratamiento como de alumnos en el grupo de control.

A su vez, el *Cuestionario de padres sobre hábitos de lectura de sus hijos* dio luz acerca del contexto socioeconómico de los menores participantes en el estudio; al igual que sobre el

acompañamiento en la lectura que los participantes en el estudio recibieron por parte de sus familiares. Este cuestionario fue respondido por 739 padres de familia. Cabe resaltar que ambos cuestionarios se entregaron en las aulas, el primero para ser llenado en el salón de clases y el segundo para responderse en casa. Por ello, dicha información es representativa de la población objetivo, mas no de los efectivamente tratados.

Para tener una mejor idea de cómo es la población incluida en el estudio, la Gráfica 3.1 presenta la distribución de los hogares a los distintos niveles socioeconómicos según la clasificación de la Asociación Mexicana de Inteligencia de Mercado y Opinión Pública (AMAI). Para su construcción se empleó una batería de preguntas acerca de distintos bienes y características de los hogares, que se detalla en el Anexo 3D.

Gráfica 3.1. Distribución de los encuestados en los distintos niveles socioeconómicos de la AMAI



Fuente: cálculos propios con base en la información recopilada por el CIDE.

La información sociodemográfica presente en el *Cuestionario de padres sobre hábitos de lectura de sus hijos* reconoce que, en promedio, seis personas viven en el hogar de los menores entrevistados. La ocupación más común de la madre es ser ama de casa; mientras que entre los

padres predominan los trabajadores de labores manuales, artes y oficios. 198 familias dicen ser beneficiarias del programa *Oportunidades*; 26.8% del total entrevistado.

De acuerdo con la clasificación de la AMAI, más del 60% de los hogares que participaron en la intervención no sufren de carencias en los bienes y servicios al interior del hogar; mientras que prácticamente el resto sufre únicamente una o dos carencias; siendo la disponibilidad de agua al interior de la vivienda la principal.⁴³

3.3.3. Balance del tratamiento

La Tabla 3.1 presenta características de los grupos de tratamiento y control antes de la intervención y se lleva a cabo una prueba de balance de dichas características entre los alumnos seleccionados para formar parte del grupo de tratamiento y los alumnos en el grupo de control; regresionando el valor de cada una de las características utilizando como variable explicativa la variable dicotómica de tratamiento. No se hallaron diferencias estadísticamente significativas en ninguna de estas características.

Tabla 3.1. Prueba de balance de las características entre grupos

Variable	Media	Dummy de tratamiento	Número de observaciones
Edad	10.91 (0.03)	-0.02 (0.05)	707
Horas al día que ve televisión	2.76 (0.12)	-0.07 (0.19)	707
Horas que pasa en internet	2.16 (0.16)	-0.13 (0.27)	707
Cuenta con amigos en la colonia (1 = sí, 0 = no)	0.88 (0.01)	0.03 (0.02)	707
Hora a la que se duerme entre semana (en formato de 24 hrs.) ⁴⁴	21.83 (0.14)	-0.01 (0.23)	707
Asistió a una biblioteca durante las vacaciones (1 = sí, 0 = no)	0.29 (0.02)	0.00 (0.04)	707

Nota: errores estándar en paréntesis. Significancia: *Pr (T<t) < 0.1, ** Pr (T<t) < 0.05, *** Pr (T<t) < 0.01.
Fuente: cálculos propios.

⁴³ Fuente: <http://nse.amai.org/nseamai2/>

⁴⁴ Para efectos estadísticos, las 0 hrs se recodificaron como “24”, la 1 a.m. se recodificó como “25” y las 2 a.m. como “26”.

Para contar con equivalencia entre las pruebas, se solicitó a la Secretaría de Educación Pública la estandarización de los puntajes entre una prueba y otra. Asimismo, para evaluar el efecto de la intervención, se compararon los resultados de la prueba ENLACE que los alumnos presentaron al final de quinto año con los resultados de una batería de preguntas de esa misma prueba; una vez que los menores ingresaron a sexto año.

La Tabla 3.2 presenta la información de la línea basal, desagregada por género, incluyendo la tasa de atrición, así como el resultado de la prueba ENLACE realizada antes del tratamiento.

Tabla 3.2. Características y puntajes promedio antes del tratamiento

	Niñas			Niños		
	Grupo de tratamiento	Grupo de control	Diferencia	Grupo de tratamiento	Grupo de control	Diferencia
Población objetivo (N)	209	376	167	211	337	126
Línea basal (n)	155	259	104	153	238	85
Tasa de no cumplimiento	25.84%	31.12%	5.28%* [3.86]	27.49%	29.38%	1.89% [3.95]
Puntaje pre-ENLACE	514.39 (90.55)	497.50 (84.64)	16.89** [8.83]	473.27 (96.52)	481.64 (91.54)	8.36 [9.80]

Nota: entre paréntesis se presentan las desviaciones estándar, y entre corchetes los errores estándar. Niveles de significancia: *P<0.1, **P<0.05, ***P<0.01. Fuente: cálculos propios con base en la información recopilada por el CIDE.

No existen diferencias estadísticamente significativas en la decisión de participar entre los grupos de tratamiento y de control en el caso de los niños; pero sí entre el grupo de tratamiento y de control en el caso de las niñas. Además, aquellas que se encuentran en el grupo de tratamiento exhiben un mayor puntaje en la prueba. Para controlar por las diferencias entre ambos grupos se estimó la diferencia de las diferencias (DID) tras la intervención para determinar el efecto de la política.

En principio, sea t' el periodo previo a las vacaciones de verano, y t el periodo posterior. El estimador DID está dado por

$$DID = E(Y_{1t} - Y_{1t'} | D = 1) - E(Y_{0t} - Y_{0t'} | D = 0) \quad (IV.1)$$

donde D denota la exposición al tratamiento, siendo 1 para los que recibieron la intervención y 0 para aquellos que no la recibieron, y E es el promedio de cada diferencia. Para que el estimador DID sea un estimador insesgado del impacto del programa, el supuesto necesario es que los alumnos que no participaron en el programa se hubieran comportado como los alumnos que sí participaron, en ausencia del tratamiento. Una forma de expresar esto es:

$$E(Y_{0t} - Y_{0t'} | D = 1) = E(Y_{0t} - Y_{0t'} | D = 0) \quad (\text{IV.2})$$

Este supuesto, pese a ser más restrictivo que en el caso de que la aleatorización hubiera sido exitosa, resulta más laxo que el suponer que las diferencias pre-intervención no alterarían los resultados después del verano. Esto se debe a que el estimador de diferencias en diferencias elimina el sesgo generado por diferencias antes del tratamiento; siempre y cuando las variaciones en el tiempo de estas variables no sean distintas entre ambos grupos (Duflo, 2000).

3.4. Resultados

Las actividades que los menores realizaron durante las vacaciones al interior de cada hogar, pese a ser muy variadas, quedaron parcialmente capturadas en el *Cuestionario de hábitos de lectura de alumnos*. Por ejemplo, 36.6% de los alumnos dijo quedarse en casa, mientras 30.4% dijo ir a un curso de verano, 25% salir de viaje más de una semana, 50% ayudar a sus papás en casa, 14.3% trabajar y 45% jugar solo o con amigos.

Respecto a la lectura, 29.2% de los alumnos identificados en el grupo de control reportó haber ido a una biblioteca en las vacaciones, comparado con 30.8% en el grupo de tratamiento; la diferencia entre ambos grupos no resultó estadísticamente significativa ($t(530) = 0.44$, $p = 0.33$).

La Tabla 3.3 presenta el grado de avance en la lectura que los alumnos -en el grupo de tratamiento- reportó para cada uno de los libros en la intervención.

Tabla 3.3. Auto-reporte de lectura de la población en el grupo de tratamiento

Libro	El Curro (%)	Crispín el tuerto (%)	El enmascarado de lata (%)
	[900L]	[1060L]	[1160L]
No leyó nada	23.47	27.42	17.36
Leyó un poco	9.32	10.97	13.18
Leyó la mitad	6.43	5.48	6.43
Leyó casi todo	12.54	12.58	11.90
Leyó todo	48.23	43.55	51.13
Total	100.00	100.00	100.00

Fuente: cálculos propios con base en la información recopilada por el CIDE. Entre corchetes se presenta el grado de dificultad de cada libro, medido a través del Spanish Lexile Analyzer⁴⁵.

Como se puede apreciar, en el caso de los primeros dos libros la tendencia es la esperada: conforme aumenta el grado de dificultad del libro, los alumnos tienden a completar un menor porcentaje de la lectura. Sin embargo, éste no es el caso del tercer libro, que reportan haber leído completo el mayor porcentaje de alumnos. De hecho, éste es el que más reportan como su lectura favorita en vacaciones (21.86% del total de los alumnos en el grupo de tratamiento).

3.4.1. Efectos capturados por condición de tratamiento

El efecto del programa se determinó por las diferencias-en-diferencias de las pruebas pre y post intervención; utilizando una variable dicotómica para identificar el grupo de tratamiento. La Tabla 3.4 presenta estimaciones separadas para las diferencias entre los grupos de tratamiento y control antes y después del programa. Estadísticamente, no es posible identificar un efecto en pérdida de habilidades de lectura; cuando utilizamos para ello los resultados obtenidos por el grupo de control. La evidencia sugiere que, si bien no hubo ganancia en habilidades, éstas

⁴⁵ Disponible en <https://www.lexile.com/analyzer/>

tampoco disminuyeron significativamente. Tampoco obtenemos ganancias estadísticamente significativas de la intervención; al comparar las diferencias obtenidas entre ambos grupos.

Tabla 3.4. Diferencias entre los puntajes de las pruebas ENLACE de español antes y después del verano

	Participantes con puntajes en ambos exámenes (n =686)			
	Pre ENLACE	Post ENLACE	Diferencia	Prueba T pareada
Grupo de Control (n =418)	489.95 (88.25)	490.41 (92.44)	0.46 [63.87]	0.15 [0.44]
Grupo de Tratamiento (n =268)	493.68 (95.68)	497.93 (95.63)	4.25 [63.77]	1.09 [0.13]
Diferencia (tratamiento – control)	3.73 (92.03)	7.52 (94.09)	3.79 [93.05]	0.75 [0.23]

Nota: entre paréntesis se presentan las desviaciones estándar, y entre corchetes el valor p asociado a los estadísticos t. Niveles de significancia: *P<0.1, **P<0.05, ***P<0.01. Fuente: cálculos propios con base en la información recopilada por el CIDE.

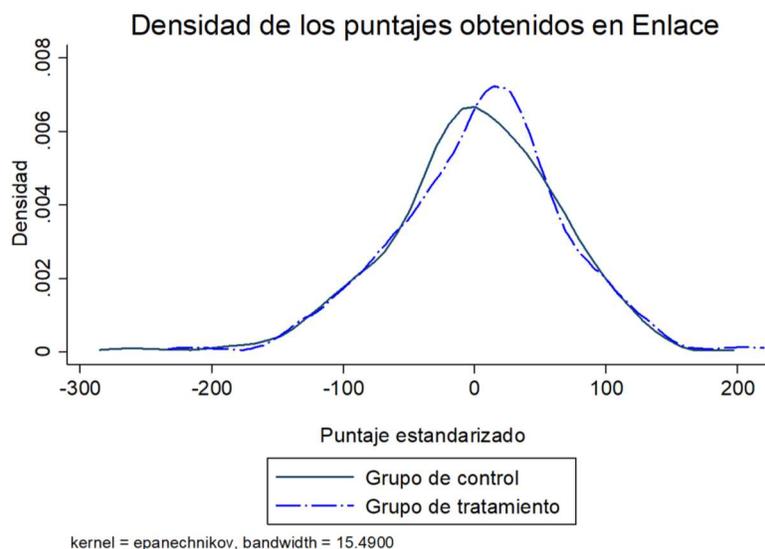
Una inspección minuciosa a la distribución de los puntajes entre ambos grupos después de la intervención se presenta en la Gráfica 3.2. Como se puede observar, no hay diferencias observables en ambas colas, indicando que existe, en promedio, una proporción similar de alumnos que tuvieron ganancias o pérdidas extraordinarias en lectura durante el verano.

Sin embargo, en la forma de ambas distribuciones alrededor de la media indica que sí existen diferencias en el sentido esperado (más alumnos por encima de la media), aunque estas diferencias no resultan estadísticamente significativas con la prueba formal presentada anteriormente.

En una etapa adicional, cuyos alcances y limitaciones se discuten en el siguiente apartado, se generaron tres puntos de corte alrededor del origen de la distribución de las diferencias de puntajes antes y después de la intervención; eliminando las observaciones más alejadas de la media de las distribuciones (*outliers*).⁴⁶

⁴⁶ Se buscaron, en tres ejercicios distintos, capturar los efectos de eliminar las observaciones alejadas a 1.5, 2 y 2.5 desviaciones estándar de la media. El Anexo 3E presenta estos resultados.

Gráfica 3.2. Distribución de la diferencia en puntajes entre alumnos en los grupos de tratamiento y control



Fuente: cálculos propios con base en la información recopilada por el CIDE.

3.4.2. Discusión de resultados

La intervención aquí descrita, a pesar de contar con un diseño robusto y un cuidado minucioso de la implementación, no obtuvo impactos estadísticamente significativos. Dos posibilidades emergen: la primera, que la muestra no haya sido lo suficientemente grande para alcanzar el poder estadístico necesario para identificar los efectos. La segunda, que el impacto del programa no haya sido distinto de cero; lo que indicaría que la intervención no tuvo impactos positivos. A continuación, se elabora sobre estas dos posibilidades.

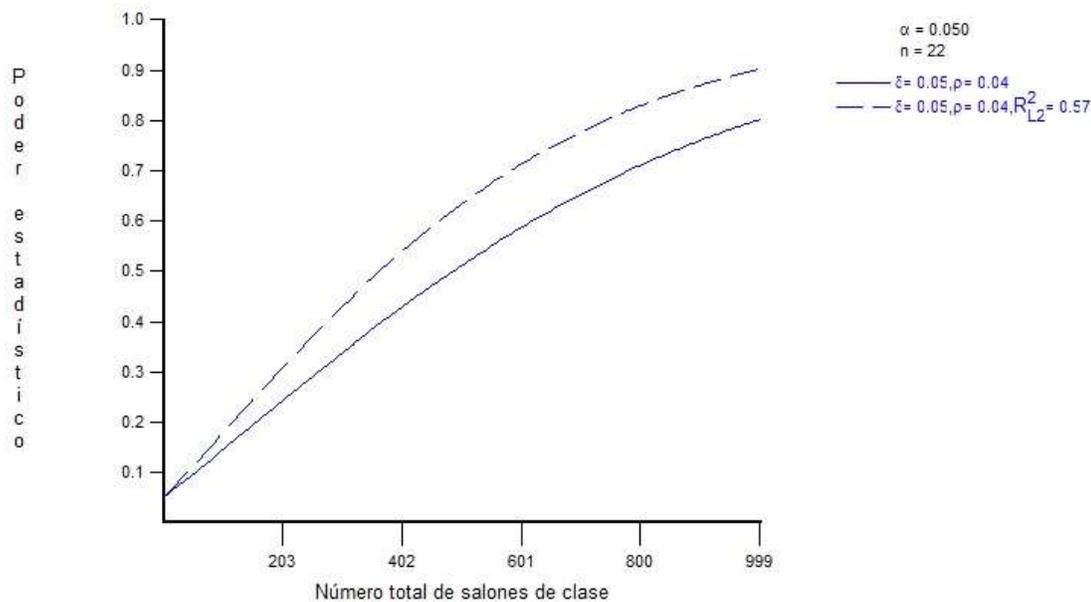
Primera posibilidad. La magnitud del impacto requiere un mayor tamaño de muestra

La primera posibilidad recae en que el diseño sí haya tenido un impacto positivo, pero la significancia estadística se haya visto comprometida por el tamaño de la muestra. La figura 3.2 presenta una estimación del total de escuelas requerido para alcanzar significancia estadística al 5%, dada la magnitud del impacto; asumiendo *clusters* de 22 niños por salón de clases. Como

se puede apreciar, el número requerido, cercano a los mil salones de clase para alcanzar un poder de 90%, es muy superior al utilizado en el estudio.

Pese a que la existencia de grandes diferencias en el aprendizaje antes y después del examen son teóricamente posibles bajo una distribución normal, estas son poco probables. Además, la diferencia en tiempo entre un examen y otro fue bastante breve; por lo que cambios dramáticos en el desempeño -tanto positivos como negativos- son poco creíbles. Recordemos que, aunque la prueba ENLACE oficial puede tener aparejados incentivos para los alumnos a fin de maximizar su esfuerzo en responder correctamente, durante la aplicación de la segunda prueba se hizo explícito que no habría ninguna consecuencia derivada del resultado.

Gráfica 3.3. Tamaño de muestra requerido para obtener efectos significativos
 Número de escuelas con nivel 2 de tratamiento y resultados a nivel de alumnos. Se asume $n=22$, $\alpha=0.05$, $\beta=0.80$, $\delta=0.05$, $\rho=0.035$



Fuente: cálculos propios utilizando el programa *Optimal Design*.

Segunda posibilidad: el programa tuvo un impacto nulo

En Shadish, Cook y Campbell (2002) se advierten las dificultades presentes en estudios basados en experimentos sociales. Los autores enuncian 6 eventos posibles que minan la validez de este tipo de diseños, y que pueden conducir a resultados sin efectividad (pp. 488-497):

Experimentos que no pueden ser plenamente implementados, debido a deficiencias operativas.

En estos casos, no existe confianza suficiente para afirmar que, en la práctica, la información proporcionada por este diseño sea superior a otra.

Una base insuficiente de teoría para justificar el experimento, que puede degenerar en un ejercicio vacío y sin contribución.

Decisiones de diseño que no valen la pena; por ejemplo, supeditar el estudio al diseño de la intervención, y no al resultado más estrechamente vinculado con el fenómeno estudiado; lo que puede conducir a un documento científico riguroso, pero cuya relevancia sea nula. O utilizar criterios muy conservadores para evitar errores de tipo I, pero dejando un margen amplio para errores de tipo II. Los críticos también objetan que los experimentos priorizan causalidades descriptivas por encima de las explicativas.

Asumir un modelo inválido de uso de investigación, a veces basado en tendencias o valores ampliamente aceptados, pero que no contienen una base teórica sólida. Incluso, el fenómeno de resultados contradictorios puede emerger como resultado de este problema, generando información que no es utilizable por un tomador de decisión.

Condiciones de experimentación que difieren de las condiciones de implementación de la política. Por ejemplo, cuando aquellos elegidos para formar parte del grupo de tratamiento difieren de aquellos a quienes la política va realmente dirigida; o cuando la aleatorización cambia el comportamiento de una población comparado con el que tendría alguien que se auto seleccionara para participar de un programa. Incluso, el hecho natural de que la implementación

de una política sea un evento de gran escala, comparado con los experimentos focalizados y casuísticos, es un elemento de interrupción que distorsiona los efectos finales de los pronosticados.

Imponer un tratamiento; lo cual sesga la dirección de las soluciones, que pueden ser diametralmente opuestas a las que se elegirían localmente si se diera la oportunidad.

Este problema se agudiza cuando no es posible contrastar a través de métodos de investigación la efectividad de dos posibles soluciones a un problema público. En particular, la solución local se va construyendo con actores determinados en un contexto específico; por lo que no es posible reproducirla a través de ejercicios académicos de pequeña escala.

En primer lugar, el impacto estadístico de intervenciones sociales, a través de diseños de tipo experimental, requiere generalmente de una exposición prolongada para ser significativo; en particular cuando se trata de incidir en el aprendizaje. Las intervenciones de política pública usualmente necesitan implementarse por periodos de tiempo que suelen ser de larga duración (seis meses o más) para generar un impacto estadísticamente significativo sobre las poblaciones tratadas.

Aunado al problema anterior, se encuentra el hecho de que, por una cuestión de recursos y logística, la población de alumnos bajo la intervención no pudo tener un seguimiento, por lo que los efectos de una exposición repetida al tratamiento no son conocidos.

Otro aspecto que pudo minar el efecto de la intervención se encuentra en las características de la población tratada. Kim (2006) advierte menores impactos esperados conforme el nivel socioeconómico de los estudiantes se incrementa⁴⁷; lo que haría -a la muestra de alumnos aquí evaluada- poco susceptible de mejora bajo esta intervención; de acuerdo con las características

⁴⁷ En condiciones donde el nivel socioeconómico no esté asociado a una segunda lengua; distinta a la que se quiere enseñar (Para una discusión más detallada cuando éste es el caso, véase Kim y Gurian, 2010).

sociodemográficas reportadas en el *Cuestionario de padres sobre hábitos de lectura de sus hijos*. Idealmente, el diseño aquí presentado habría aleatorizado en grupos sociodemográficos para probar esta hipótesis; pero desafortunadamente no se hizo.

Adicionalmente, la literatura internacional reconoce la importancia de empatar los libros entregados a cada niño durante el verano con las habilidades lectoras de cada uno; a fin de que no les resulten ni tan complicados (de forma que terminen renunciando a su lectura) ni tan sencillos (de forma que no aporten al desarrollo de sus habilidades lectoras). La presente intervención solo tuvo el alcance para otorgar los mismos libros a todos los estudiantes, atendiendo un criterio promedio de las competencias de éstos; de forma tal que pudo no resultar efectiva ni para los lectores más aventajados, ni para aquellos con mayor rezago.

Por último, a un nivel de diseño, en Kim y Quinn (2013) se mostró que la supervisión por parte de los padres para generar incrementos estadísticamente significativos sobre vocabulario y comprensión lectora resultan insuficientes (pp. 400-410). Adicionalmente, se advierte que dicha supervisión requiere de capacitación que, se avizora, necesita un involucramiento potencialmente intensivo en tiempo por parte de los padres.

Aunque la presente intervención sí incluye la figura de un capacitador y un folleto para padres, no existió ningún tipo de verificación externa de que éstos hayan implementado correcta y regularmente las técnicas sugeridas con sus hijos. El proxy para obtener esta información fueron auto-reportes, sujetos a posibles sesgos en las respuestas de los encuestados.

3.5. Conclusión y recomendaciones de política pública

Este trabajo buscó presentar una intervención que -aún constreñida por su dimensión y recursos- resulta un esfuerzo necesario, dada la escasez de este tipo de estudios en el país. El objetivo fue el de extender las intervenciones de verano para contrarrestar la PHL en México.

En un primer momento, se buscó aportar al estudio de Ortega-Hesles (2012, *op. cit.*), atendiendo sus recomendaciones sobre tamaño de muestra, tipo de tratamiento y grupos contruados (pp. 16-33). En particular, la elección de un mayor número de alumnos, un solo tipo de tratamiento, y una aleatorización a nivel de alumno dentro del salón de clases buscaron contrarrestar el bajo poder estadístico de los efectos esperados, a fin de incrementar el impacto.

No obstante, como se observó, aún estas correcciones fueron insuficientes para dirimir la controversia entre una falta de significancia estadística debida a 1) la falta de una muestra suficientemente grande o 2) impactos nulos del programa.

Aún cuando el objetivo de lograr impactos estadísticamente significativos con esta nueva intervención no se alcanzó, los resultados abonan a la comprensión sobre los alcances de estas intervenciones en nuestro contexto; así como los retos en la implementación de éstas y los potenciales sesgos que enfrentan.

Ortega-Hesles (2012, *op. cit.*) era la única evidencia de los efectos de intervenciones de verano para paliar la PHL en nuestro país. Explorar nuevos contextos para dichas intervenciones, así como diseñar nuevas formas de incrementar su efectividad, eran líneas de investigación abiertas por este trabajo.

En particular, con esta intervención se postula que -en el contexto en el que fue aplicada- caben esperar impactos muy modestos, tanto por las características socioeconómicas relativamente favorecidas de la población tratada como por el problema de incentivos, que se puede traducir en sesgos en la estimación de los impactos. Y que sería importante contar con evidencia del impacto de intervenciones de mediano y largo plazo; por ejemplo, que incluyeran el periodo lectivo del año escolar -como bibliotecas virtuales disponibles para los alumnos en todo momento- a fin de incrementar el tiempo de exposición al tratamiento.

Pese a que un análisis más detallado de los resultados -construyendo las muestras a las que tentativamente tienen mayor validez y confiabilidad- pareciera sustentar la hipótesis de la falta de poder estadístico⁴⁸; la evidencia aquí presentada no es concluyente. Una limitación de origen que pudo influir fue la diferencia entre las niñas en los grupos de tratamiento y control en la línea base, aún cuando el cálculo desagregado por género no mostró significancia estadística ni para niños, ni para niñas. Los resultados, no obstante, muestran consistencia. Como se observó, el nivel socioeconómico de los alumnos, de acuerdo con la tendencia internacional, permitía esperar impactos limitados de la intervención (que se beneficia menos que la población en mayor desventaja socioeconómica). Los estudios de Guryan *et. al.* (2015, *op. cit.*) y White *et. al.* (2014) muestran que los impactos de intervenciones similares sí resultan significativos para alumnos en un contexto de pobreza, en el caso de Estados Unidos. Del mismo modo, la falta de una entrega de libros personalizada al nivel de lectura de cada menor disminuyó el efecto esperado. En un último aspecto, este experimento pudo haberse nutrido de una supervisión más estrecha hacia los ejercicios de lectura que los menores llevaron a cabo durante las vacaciones. La falta de seguimiento (y en última instancia, de aseguramiento de su ejecución) de la lectura de los libros entregados, dificultó su cumplimiento efectivo; lo que pudo, una vez más, mermar la efectividad.

Estas intervenciones son ilustrativas del alcance que pueden tener las evaluaciones de impacto en el contexto mexicano; de las políticas -en particular educativas- que se pueden impulsar, y dotar de evidencia la efectividad de estas políticas; así como de los retos de diseño, operativos y de recursos que enfrentan.

⁴⁸ Como principal causa de los impactos nulos del programa.

La intervención descrita hasta ahora enfrentó varios retos. En cuanto al diseño de investigación, el primer reto consistió en tener un tamaño de muestra suficiente que permitiese resultados robustos en términos de su significancia estadística; dada la magnitud esperada. La principal limitación para obtener el tamaño de muestra deseado estribó en el financiamiento (tanto para comprar los libros como para capacitar a los padres); aunque existieron otros aspectos como el monitoreo (para evaluar cuántas familias llevaron a cabo efectivamente la intervención) o minimizar la tasa de atrición.

También ocurrieron otros imprevistos, como los factores externos durante la implementación que dificultaron el levantamiento de la información; y que se buscaron solucionar con planes contingentes y el aplazamiento del calendario por casi tres meses. El abordaje de las familias en el grupo de tratamiento se realizó de forma individual, producto del trabajo de campo.

Recomendaciones de política pública

Intervenciones de esta naturaleza pueden jugar un papel importante para mejorar las oportunidades de aprendizaje en los niños; en especial sus habilidades lectoras. Esto puede ser de particular interés si se enfoca en grupos desfavorecidos; por ejemplo, aquellos que se encuentran en situaciones de mayor desventaja al no contar en sus hogares con suficiente acceso a materiales impresos, o internet. El presente ejercicio arroja luz sobre el futuro camino a tomar a fin de dimensionar el potencial de este tipo de intervenciones sobre el desempeño lector de los niños mexicanos.

Es importante recordar que la magnitud que alcanza la PHL se da en contextos que se repiten año con año a lo largo de seis, siete o más ciclos escolares de los alumnos. El efecto acumulativo de la PHL se da en contextos de rezago progresivo (Entwisle, Alexander y Olson, 2007, *op. cit.*), por lo que intervenciones de dos o más veranos consecutivos podrían ser más efectivas y alcanzar impactos estadísticamente significativos.

Programas que incrementen el acceso a experiencias educativas enriquecedoras -para estudiantes durante las vacaciones de verano- deberían repetirse periódicamente a lo largo de la educación básica. Una forma sería crear políticas que incrementen el acceso a materiales de lectura apropiados, instrucción de alta calidad, y experiencias pedagógicas; particularmente para alumnos en condiciones menos favorables y/o con dificultades en el aprendizaje.

Respecto a las limitaciones, los resultados obtenidos en este trabajo se suman al cúmulo de evidencia, principalmente para el caso de Estados Unidos, respecto a los modestos impactos que, sobre lectura, caben esperar en el reducido tiempo de unas vacaciones de verano (Kim y Quinn, 2013, *op. cit.*).

Una posibilidad adicional es la de generar intervenciones poco intensivas durante las vacaciones de verano, pero cuya duración se prolongue durante el ciclo escolar. Un ejemplo es la propuesta de leer 20 minutos al día, como hábito, impulsado por la asociación civil Consejo de la Comunicación, en México⁴⁹.

En ese sentido, el desarrollo de la lectura podría convertirse en una intervención transversal para las autoridades educativas, incorporándola a políticas ya existentes; como la entrega de tabletas electrónicas con bibliotecas virtuales o la provisión de bibliotecas escolares, incluyendo su uso en el plan escolar; y estudiando su efecto. La evidencia encontrada para México y Estados Unidos parece indicar que el acceso a materiales de lectura apropiados es el factor más correlacionado con el desempeño lector (Kim y Guryan, 2010; Hesles, 2012).

3.6. Referencias

Allington, Richard L. y Anne McGill-Franzen (Eds.) (2013) *Summer reading: Closing the rich/poor reading achievement gap*. Teachers College Press.

49 <http://leermx.com/tag/lee-20-minutos-al-dia/>

- (2017) “Summer Reading Loss Is the Basis of Almost All the Rich/Poor Reading Gap”, en Rosalind Horowitz y S. Jay Samuels (eds.) *The Achievement Gap in Reading. Complex Causes, Persistent Issues, Possible Solutions*. Routledge, New York.
- Allington, Richard L.; Anne McGill-Franzen, Gregory Camilli, Lunetta Williams, Jennifer Graff, Jacqueline Zeig, Courtney Zmach y Rhonda Nowak (2010) “Addressing summer reading setback among economically disadvantaged elementary students”, en *Reading Psychology*, Vol. 31, No. 5, pp. 411-427.
- Angrist, Joshua D. y Jörn-Steffen Pischke (2008) *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- (2014) *Mastering Metrics: The path from cause to effect*. Princeton University Press.
- Blanton, Morgan V. (2015) “Keys to Reducing Summer Regression: The Reader, Routine and Relationship”, en *Journal of Organizational & Educational Leadership*, Vol. 1, No. 1, Art. 1.
- Borman, Geoffrey D. y Jerome V. D’Agostino (1996) “Title I and Student Achievement: A Meta-Analysis of Federal Evaluation Results”, en *Educational Evaluation and Policy Analysis*, Vol. 18, No. 4, pp. 309-326.
- Butler, Tammy Lynn (2010) “The effects of summer reading programs on low-income English language learners’ and low-income English L1s’ reading achievement in grades 2-4”, *Doctoral dissertation UMI 3407348*. University of Arkansas, ProQuest Dissertations Publishing, 2010.
- Burns, M. Susan y Catherine E. Snow (2002) *Un buen comienzo: guía para promover la lectura en la infancia*. Secretaría de Educación Pública.
- Bruns, Barbara; Deon Filmer y Harry Anthony Patrinos (2011) *Making schools work: New evidence on accountability Reforms*. The World Bank, Washington D.C.

- Campbell, Donald T. (1970) “Considering the case against experimental evaluations of social innovations.”, en *Administrative Science Quarterly*, Vol. 15, No. 1, pp. 110-113.
- Carver, Ronald P. y Robert E. Leibert (1995) “The Effect of Reading Library Books at Different Levels of Difficulty upon Gain in Reading Ability”, en *Reading Research Quarterly*, Vol. 30, No. 1, pp. 26-48.
- Castro-Caldas, A., P. Cavaleiro Miranda, I. Carmo, A. Reis, F. Leote, C. Ribeiro y E. Ducla-Soares (1999) “Influence of learning to read and write on the morphology of the corpus callosum”, en *European Journal of Neurology*, Vol. 6, No. 1, pp. 23-28.
- Chamblee, Nikki A. (2014) *The impact of a summer library program on the reading attitude and reading comprehension skills of a diverse group of beginning third graders*. UMI Dissertation No. 3663074, New York, EUA.
- Chin, Tiffany y Meredith Phillips (2004) “Social Reproduction and Child-rearing Practices: Social Class, Children’s Agency, and the Summer Activity Gap”, en *Sociology of Education*, Vol. 77, pp. 185-210.
- CONEVAL (2015) *La política de evaluación en México: 10 años del CONEVAL*. México.
- Consejo Nacional para la Cultura y las Artes (2018) *Encuesta Nacional de Lectura y Escritura 2015-2018*. México.
- Cooper, Harris; Barbara Nye, Kelly Charlton, James Lindsay y Scott Greathouse (1996) “The effects of summer vacation on achievement test scores: A narrative and meta-analytic review”, en *Review of Educational Research*, Vol. 66, No. 3, pp. 227-268.
- DeCasper, Anthony J. y Spence, Melanie J. (1986) “Prenatal maternal speech influences newborn’s perception of speech sounds”, en *Infant Behavior and Development*, Vol. 9, pp. 133–150.

- Dehaene, Stanislas (2009) *Reading in the Brain, The New Science of How We Read*. Penguin Books.
- Dehaene, Stansilas; Felipe Pegado, Lucia W. Braga, Paulo Ventura, Gilberto Nunes Filho, Antoinette Jobert, Ghislaine Dehaene-Lambertz, Régine Kolinsky, José Morais y Laurent Cohen (2010) “How learning to read changes the cortical networks for vision and language”, en *Science*, Vol. 330, No. 6009, pp. 1359-1364.
- Downey, Douglas B, Paul T. Von Hippel y Beckett A. Broh (2004) “Are Schools the Great Equalizer? Cognitive Inequality during the Summer Months and the School Year”, en *American Sociological Review*, Vol. 69, No. 5, pp. 613-635.
- Duflo, Esther (2000) *Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment* (No. w7860). National Bureau of Economic Research.
- Durkin, Dolores (1993) *Teaching them to read* (6th ed.) Allyn & Bacon. Boston.
- Entwisle, Doris R., Karl L. Alexander y Linda Steffel Olson (2001) “Keep the faucet flowing: Summer learning and home environment.”, en *American Educator*, Vol. 25, No. 3, pp. 10-15.
- (2007) “Lasting Consequences of the Summer Learning Gap”, en *American Sociological Review*, Vol. 72, No. 2, pp. 167-180.
- Fondo de Cultura Económica/Secretaría de Educación Pública (2000) *Un buen comienzo. Guía para promover la lectura en la infancia*.
- Freire, Paulo (2006) *La importancia del acto de leer*. Laboratorio Educativo.
- Gershenson, Seth (2013) “Do Summer Time-Use Gaps Vary by Socioeconomic Status?”, en *American Educational Research Journal*, Vol. 50, No. 6, pp. 1219–1248.

- Gómez López, Luis Felipe; y Juan Carlos Silas Casillas (2012) “Impacto de un programa de comprensión lectora”, en *Revista latinoamericana de estudios educativos*, Vol. 42, No. 3, pp. 35-63.
- Green, Donald P.; Shang E. Ha y John G. Bullock (2010) “Enough already about “black box” experiments: Studying mediation is more difficult than most scholars suppose”, en *The Annals of the American Academy of Political and Social Science*, Vol. 628, No. 1, pp. 200-208.
- Guryan, Jonathan; James S. Kim, Lauren Capotosto, David M. Quinn, Helen Chen Kingston, Lisa Foster y North Cooc (2015) “Can a Scaffolded Summer Reading Intervention Reduce Socioeconomic Gaps in Children’s Reading Comprehension Ability and Home Book Access? Results from a Randomized Experiment”, *Institute for Policy Research*, Northwestern University, WP-15-15.
- Hayes, Donald P. y Judith Grether (1969) “The School Year and Vacations: When Do Students Learn?”, *Paper presented at the Eastern Sociological Association Convention*, New York.
- Heyns, Barbara (1978). *Summer Learning and the Effects of Schooling*. Academic Press.
- (1987). “Schooling and cognitive development: Is there a season for learning?”, en *Child Development*, Vol. 58, pp. 1151-1160.
- Horowitz, Rosalind y S. Jay Samuels (eds.) (2017) *The Achievement Gap in Reading. Complex Causes, Persistent Issues, Possible Solutions*. Routledge, New York.
- Hunt, Lyman C. (1970/1996) “The effect of self-selection, interest, and motivation upon independent, instructional, and frustrational levels”, en *The Reading Teacher*, Vol. 24, No. 2, pp. 278- 282.
- Imai, Kosuke; Luke Keele, Dustin Tingley y Teppei Yamamoto (2011) “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and

- Observational Studies”, en *American Political Science Review*, Vol. 105, No. 4, pp. 765-789.
- Ishai, Alomit, Leslie G. Ungerleider, Alex Martin y James V. Haxby (2000) “The Representation of Objects in the Human Occipital and Temporal Cortex”, en *Journal of Cognitive Neuroscience*, No. 12, Suplemento 2, pp. 35-51.
- Kim, James S. (2007) “The effects of a voluntary summer reading intervention on reading activities and reading achievement”, en *Journal of Educational Psychology*, Vol. 99, No. 3, pp. 505–515.
- (2006) “The Effects of a Voluntary Summer Reading Intervention on Reading Achievement: Results from a Randomized Field Trial”, en *Educational Evaluation and Policy Analysis*, Vol. 28, No. 4, pp. 335-355.
- (2004) “Summer reading and the ethnic achievement gap”, en *Journal of Education for Students Placed at Risk*, Vol. 9, No. 2, pp. 169-188.
- Kim, James S. y Jonathan Guryan (2010) “The Efficacy of a Voluntary Summer Book Reading Intervention for Low-Income Latino Children from Language Minority Families”, en *Journal of Educational Psychology*, Vol. 102, No. 1, pp. 20-31.
- Kim, James S. y David M. Quinn (2013) “The effects of summer reading on low-income children’s literacy achievement from Kindergarten to Grade 8: A meta-analysis of classroom and home interventions”, en *Review of Educational Research*, Vol. 83, No. 6, pp. 386-431.
- Kim, James S. y Thomas G. White (2008) “Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study”, en *Scientific Studies of Reading*, Vol. 12, No. 1, pp. 1-23.

- Kim, James S., Jonathan Guryan, Thomas White, David Quinn, Lauren Capotosto y Helen Chen Kingston (2016) “Delayed effects of a low-cost and large-scale summer reading intervention on elementary school children’s reading comprehension”, en *Journal of Research on Educational Effectiveness*, 9(sup1), 1-22.
- Kremer, Michael, Edward Miguel y Rebecca Thornton (2009) “Incentives to Learn”, en *Review of Economics and Statistics*, Vol. 91, No. 3, pp. 437–456.
- Lalive; Rafael y M. Alejandra Cattaneo (2009) “Social Interactions and Schooling Decisions”, en *The Review of Economics and Statistics*, Vol. 91, pp. 457-477.
- Leal, Fernando (2003) *Introducción al estudio del español desde una perspectiva multidisciplinaria*. Universidad de Guadalajara.
- Leeuw, Frans y Jos Vaessen (2009) “NONIE Guidance on Impact Evaluation”; Draft Version for Discussion at the Cairo conference March-April, 2009. Network on Networks on Impact Evaluation.
- Luftig, Richard L. (2010) “When a little bit means a lot: The effects of a short-term reading program on economically disadvantaged elementary schoolers”, en *Reading Research and Instruction*, Vol. 42, No. 4, pp. 1-13.
- Maluf, María Regina y Renan Sargiani (2013) “Lo que la neurociencia tiene que decir sobre el aprendizaje de la lectura”, en *Revista de Psicología de Arequipa*, Vol. 3, No. 1, pp. 9-23.
- Mcbride-Chang, Catherine (2015) *Children’s Literacy Development: A Cross-Cultural Perspective on Learning to Read and Write*. Routledge.
- Mceachin, Andrew y Allison Atteberry (2017) “The Impact of Summer Learning Loss on Measures of School Performance”, en *Education Finance and Policy*, Vol. 12, No. 4, pp. 468-491.

- Melosh, Geraldine M. (2003) *Summer Books! Stemming Reading Loss in High-Poverty Second Graders Through Access to 192esultados192 Books During the Summer Vacation Break*. Doctoral dissertation, University of Florida, EUA.
- Meyer, Bruce D. (1995) “Natural and Cuasi-experiments in Economics”, en *Journal of Business and Economic Statistics*, Vol. 13, pp. 151-161.
- Miller, Elizabeth (2018) *The Effects of a Summer Literacy Program on Mitigating Summer Reading Loss in Elementary Students*. UMI Dissertation No. 10810020, Lamar University, EUA.
- Mraz, Maryann y Timothy V. Rasinski (2007) “Summer Reading Loss”, en *The Reading Teacher*, Vol. 60, No. 8, pp. 784-789.
- National Reading Panel (2000) *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. National Institute of Child Health and Human Development, Washington D.C.
- Neuman, Susan B. (2009) *Changing the Odds for Children at Risk*. Teachers College Press, Nueva York.
- Neuman, Susan B. y Donna Celano (2006) “The knowledge gap: Implications of leveling the playing field for low-income and middle-income children”, en *Reading Research Quarterly*, Vol. 41, No. 2, pp. 176-201.
- (2012) *Giving our children a fighting chance: Poverty, literacy, and the development of information capital*. Teachers College Press.
- Nieder, Andreas, Ilka Diester y Oana Tudusciuc (2006) “Temporal and spatial enumeration processes in the primate parietal cortex”, en *Science*, Vol. 313, No. 5792, pp.1431-1435.
- Noguerol, Artur (2003) “Leer para pensar, pensar para leer: la lectura como instrumento para el aprendizaje en el Siglo XXI”, en *Lenguaje*, No. 31, pp. 36-58.

- O'Connor, Rollanda E., Deborah Fulmer, Kristin R. Harty, and Kathryn M. Bell (2005) "Layers of Reading Intervention in Kindergarten Through Third Grade Changes in Teaching and Student Outcomes", en *Journal of Learning Disabilities*, 38(5), pp. 440-455.
- Ortega-Hesles, María Elena (2012) "Lessons from the Pilot Study of a Cluster Randomized Trial: Summer Reading Intervention Targeting Third Grade Students in Mexico", *Qualifying paper*, Harvard Graduate School of Education.
- Pagan, S. L. (2010) "Children reading for pleasure: Investigating predictors of reading achievement and the efficacy of a paired-reading intervention to foster children's literacy Skills". *Doctoral dissertation UMI No. 70556*.
- Pagan, Stephanie y Monique Sénéchal (2014) Involving parents in a summer book reading program to promote reading comprehension, fluency, and vocabulary in grade 3 and grade 5 children. *Canadian Journal of Education*, 37(2), n2.
- Petersen, Steven E.; Peter T. Fox, Abraham Z. Snyder y Marcus E. Raichle (1990) "Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli", en *Science*, Vol. 249, No. 4972, pp. 1041-1044.
- Public Impact (2018) *Closing achievement gaps in diverse and low-poverty schools: An action guide for district leaders*. Oak Foundation, Carolina del Norte, EUA y Ginebra, Suiza.
- Audenbush, Stephen; Jessaca Spybrook, Howard Bloom, Richard Congdon, Carolyn Hill y Andres Martinez (2011) *Optimal Design Software for Multi-level and Longitudinal Research* (Version 3.01) [Software]. Disponible en <http://wtgrantfoundation.org/resource/optimal-design-with-empirical-information-od>. Consultado el 7 de marzo de 2019.
- Ravallion, M. (1999) *The mystery of the vanishing benefits: Ms Speedy analyst's introduction to evaluation* (Vol. 2153). World Bank Publications.

- Reimers, Fernando y Jenny Eva Jacobs (2008) *Leer (comprender y aprender) y escribir para comunicarse. Desafíos y oportunidades para los sistemas educativos* pp. 11-61.
- Rosselli, Mónica; E. Matute y A. Ardila (2006) “Predictores neuropsicológicos de la lectura en español”, en *Revista de Neurología*, Vol. 42, No. 4, pp. 202-210.
- Sakaiya, Taichi (1994) *Historia del futuro. La sociedad del conocimiento*. Editorial Andrés Bello, Santiago.
- Seawright, Marcia (2017) *The Effects of a Summer Reading Program on Reading Achievement and Reading Motivation*. UMI Dissertation No. 10264389. University of South Carolina, EUA.
- Seol, Sunmi; Aaron Sharp y Paul Kim (2011) “Stanford Mobile Inquiry-based Learning Environment (SMILE): using mobile phones to promote student inquires in the elementary classroom”, en *Proceedings of the 2011 International Conference on Frontiers in Education: Computer Science & Computer Engineering*, pp. 270-276.
- Seward, Carrie (2009) “Evaluating the effectiveness of a short-duration reading intervention on grade one phonological awareness and word reading”, en *Theses and Dissertations (Comprehensive)*. 1071.
- Snowling, Margaret J. y Charles Hulme (Eds.) (2008) *The Science of Reading: A Handbook*. John Wiley & Sons.
- Tarkiainen, Antti, Piers L. Cornelissen, y Riitta Salmelin (2002) “Dynamics of visual feature analysis and object-level processing in face versus letter-string perception”, en *Brain*, Vol. 125, No. 5, pp. 1125-1136.
- Van Andel, Ashley Kristina (2011) “Summer library reading programs and literacy: An assessment of children’s reading progress after having participated in a summer library reading program”. *Theses and Dissertations (Comprehensive)*. 1035.

- Vilalta, Carlos (2005) “Cómo enseñar autocorrelación espacial”, en *Economía, Sociedad y Territorio*, Vol. 5, No. 18, pp. 323-333.
- White, Thomas G; James S. Kim; Helen Chen Kingston y Lisa Foster (2014) “Replicating the Effects of a Teacher-Scaffolded Voluntary Summer Reading Program: The Role of Poverty”, en *Reading Research Quarterly*, Vol. 49, No. 1, pp. 5-30.
- Wilkins, Chuck, Russell Gersten, L. E. Decker, Leslie Grunden, Sarah Brasiel, K. Brunnert, y Madhavi Jayanthi (2012) *Does a summer reading program based on Lexiles affect reading comprehension? Final Report*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.
- Wood, David R., Jérôme Seymour Bruner y Gail Ross (1976) “The role of tutoring in problem solving”, en *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, Vol. 17, No. 2, pp. 89-100.

Anexo 3A. Efectos encontrados en diversas intervenciones de lectura durante el verano

Autor y año	Población objetivo	Método	Intervención	Impacto (d) 95% CI
Melosh (2003)	Alumnos de segundo grado en el nivel de primaria.	Diseño experimental.	Entrega de libros con un sistema tutorial de acompañamiento.	0.108 (-0.505 a 0.721)
Kim (2006)	Alumnos de cuarto grado en el nivel de primaria.	Diseño experimental.	Entrega de ocho libros durante el verano.	0.083 (-0.095 a 0.261)
Kim (2007)	Alumnos entre primer y quinto grado en el nivel de primaria.	Diseño experimental.	Entrega de diez libros durante el verano.	0.185 (-0.051 a 0.420)
Kim y White (2008)	Alumnos entre tercero y quinto grado en el nivel de primaria.	Diseño experimental.	Entrega de ocho libros, con acompañamiento tanto de padres como de profesores externos.	0.089 (-0.194 a 0.371)
Seward (2009)	Alumnos de preprimaria.	Diferencias en diferencias evaluadas con RMANCOVA.	Campamento de verano para los niños, acompañado de talleres para padres.	0.553 (0.025 a 1.081)
Allington <i>et. al.</i> (2010)	Alumnos de Primaria.	Diseño experimental.	Entrega de libros durante tres veranos.	0.139 (0.027 a 0.251)
Butler (2010)	Alumnos entre segundo y cuarto grado en el nivel de primaria.	Diferencias en diferencias evaluadas con ANCOVA.	Envío de un libro semanal con seguimiento.	0.708 (0.202 a 1.214)
Kim y Gurian (2010)	Alumnos de cuarto grado en el nivel de primaria.	Diseño experimental.	Envío de 10 libros durante las vacaciones, con acompañamiento y entrenamiento para padres.	-0.021 (-0.323 a 0.281)
Pagan (2010)	Alumnos de tercero y quinto grado en el nivel de primaria.	Diseño experimental.	Envío de un libro semanal con acompañamiento de los padres.	0.11 (-0.410 a 0.630)
Van Andel (2011)	Alumnos entre segundo y quinto grado en el nivel de de primaria.	Diferencias en diferencias evaluadas con MANCOVA.	Programa de lectura en una biblioteca.	0.05 (-0.511 a 0.611)
Wilkins, Chuck, et al. (2012)	Alumnos de tercer grado en el nivel de primaria.	Diseño experimental.	Entrega de libros (elegidos a través de Lexile Framework®) y recordatorios.	0.02 (-0.044 a 0.096)

Fuente: cálculos propios.

Anexo 3A. Efectos encontrados en diversas intervenciones de lectura durante el verano (Cont.)

Autor y año	Población objetivo	Método	Intervención	Impacto (d) 95% CI
Ortega-Hesles (2012)	Alumnos de tercer y cuarto grado en el nivel de primaria.	Diseño experimental.	Entrega de libros y acompañamiento para padres.	0.06 (-0.135 a 0.26)
Guryan, Kim, Capotosto, Quinn, Chen, Kingston, Foster y Cooc (2015) ⁵⁰	Alumnos de tercer grado en el nivel de primaria.	Diseño experimental.	Entrega de libros, acompañamiento para padres y llamadas del profesor.	0.11 (0.09 a 0.126)
White, Kim, Chen Kingston y Foster (2015)	Alumnos de tercer grado en el nivel de primaria.	Diseño experimental.	Entrega de libros, acompañamiento para padres y llamadas del profesor.	0.11 (0.06 a 0.16)
Kim, Guryan, White, Quinn, Capotosto y Kingston (2016)	Alumnos de tercer y cuarto grado en el nivel de primaria.	Diseño experimental.	Entrega de libros, acompañamiento para padres y llamadas del profesor, tras 9 meses.	0.05 (0.01 a 0.094)

Fuente: cálculos propios.

⁵⁰ Los efectos encontrados en Guryan (2015), White *et. al.* (2015) y Kim *et. al.* (2016) corresponden a la interacción del tratamiento con medidas socioeconómicas de nivel de pobreza alto en la población tratada.

Anexo 3B. Carta a los Directores en las escuelas seleccionadas.



24 de octubre de 2011

Estimado Director/a,

La escuela que usted dirige fue seleccionada al azar para participar en un estudio sobre prácticas de lectura. El propósito es aprender sobre los hábitos de lectura entre los niños de primaria del municipio de -----, Aguascalientes. Para poder encontrar maneras para ayudar a sus estudiantes y a otros niños a leer mejor necesitamos de su apoyo.

De esta manera, en el marco del programa “Verano de Lectura”, que dio inicio en el período vacacional pasado, los próximos 24 al 26 de octubre un equipo de trabajo del Centro de Investigación y Docencia Económicas, A.C. (CIDE) visitará su institución. El objetivo será llevar a cabo la aplicación de los reactivos de español de la prueba ENLACE 2011 a los grupos **de quinto grado de primaria** que fueron seleccionados previamente para el proyecto, así como un cuestionario sobre hábitos de lectura.

La aplicación tendrá una duración de dos horas y quince minutos divididos en dos sesiones de cuarenta y cinco minutos, cada una, antes del receso y una después. Los materiales que se aplicarán y entregarán a los alumnos durante la dinámica son: a) una prueba de español, b) una hoja de respuestas, c) un cuestionario sobre lectura para alumnos.

La preparación de la prueba deberá dar inicio a las 8am en el caso del turno matutino con aplicación de 9am a 12:15 con receso a las 10:30am. En el turno vespertino, la preparación será de 1:30pm, con inicio de la prueba a las 2:30pm, con término aproximado a las 5:15pm.

Al término de la aplicación entregaremos a los estudiantes un breve cuestionario que llevarán a sus casas para pedir que sus padres lo contesten y que deberán regresar a sus profesoras al día siguiente.

Por este motivo, mucho agradeceré se les permita la preparación y aplicación de la prueba a los aplicadores que se identifiquen con esta carta.

En caso de tener cualquier duda nos puede contactar directamente al teléfono [REDACTED] en la Ciudad de México, o bien en la siguiente dirección de correo electrónico: [REDACTED].

Sin otro particular, agradezco de antemano su contribución a este proyecto,

Atentamente,

Dr. Sergio Cárdenas Denham

Profesor- Investigador

División de Administración Pública

Centro de Investigación y Docencia Económicas, A. C. (CIDE).

Anexo 3C. Bitácora de las visitas a domicilio por parte de los capacitadores

Información del levantamiento y visitas realizadas		
D1. Nombre del alumno		
D2. Escuela a la que asiste		
D3. Nombre del entrevistador		
	Control	Visita UNO
D4. Fecha		
D5. Hora de visita		
D6. Código de resultado (consulte la lista siguiente)		
D7. Por favor explique el código asignado en caso de ser necesario		

Lista de códigos de resultados para visitas realizadas
1. Entrevista completa, con entrega de libros y capacitación a cuidador
2. Entrevista incompleta, con entrega de libros sin capacitación
3. Entrevista incompleta, no recibe libros pero acepta capacitación
4. Menor presente sin padre, tutor o cuidador
5. Familia ausente temporalmente
6. El padre, tutor o cuidador se negó recibir información y libros
7. Vivienda deshabitada
8. No conocen al alumno
9. El alumno ya no vive con la familia
10. El alumno nunca ha vivido en este domicilio
11. Falleció el alumno
12. Toda la familia migró fuera de la localidad
13. Domicilio incompleto o inexistente
14. Otra (explique...)

Anexo 3D. Construcción del índice de niveles socioeconómicos AMAI.

A fin de construir un indicador del nivel socioeconómico de los hogares, se optó por utilizar la clasificación ofrecida por la Asociación Mexicana de Inteligencia de Mercado y Opinión Pública (AMAI). Dicha clasificación, conocida como Índice de Niveles Socio Económicos (NSE)⁵¹, parte de un algoritmo “que clasifica a los hogares en siete niveles, considerando ocho características o posesiones del hogar y la escolaridad de la persona que más aporta al gasto”.

Las ocho características son:

- Número de habitaciones
- Número de baños completos
- Número de focos
- Número de autos
- Regadera
- Estufa
- Tipo de piso

El índice otorga un puntaje determinado a la presencia o ausencia de las distintas características en el hogar; utilizando para ello rangos, y se construye sumando el puntaje obtenido por cada hogar, de acuerdo con dichas características. La Tabla A5.1 muestra el puntaje asociado a cada característica:

⁵¹ Información disponible en la página <http://nse.amai.org/nseamai2/>

Tabla 3D.1 Puntaje asociado a cada característica, de acuerdo con la regla AMAI

CANTIDAD	PUNTOS				CANTIDAD	PUNTOS																						
	TV A COLOR	AUTOMÓVIL	BAÑOS	COMPUTADORA		PISO DIFERENTE DE TIERRA O CEMENTO	REGADERA	ESTUFA																				
No Tener	0	0	0	0	No Tener	0	0	0																				
1	26	22	13	17	Tener	11	10	20																				
2	44	41	13	29	<table border="1"> <thead> <tr> <th colspan="2">ESCOLARIDAD</th> </tr> <tr> <th>NIVEL</th> <th>PUNTOS</th> </tr> </thead> <tbody> <tr> <td>Sin Instrucción</td> <td>0</td> </tr> <tr> <td>Primaria o secundaria completa o incompleta</td> <td>22</td> </tr> <tr> <td>Carrera técnica preparatoria completa o incompleta</td> <td>38</td> </tr> <tr> <td>Licenciatura completa o incompleta</td> <td>52</td> </tr> <tr> <td>Postgrado</td> <td>72</td> </tr> </tbody> </table>				ESCOLARIDAD		NIVEL	PUNTOS	Sin Instrucción	0	Primaria o secundaria completa o incompleta	22	Carrera técnica preparatoria completa o incompleta	38	Licenciatura completa o incompleta	52	Postgrado	72						
ESCOLARIDAD																												
NIVEL	PUNTOS																											
Sin Instrucción	0																											
Primaria o secundaria completa o incompleta	22																											
Carrera técnica preparatoria completa o incompleta	38																											
Licenciatura completa o incompleta	52																											
Postgrado	72																											
3	58	58	31	29																								
4	58	58	48	29																								
<table border="1"> <thead> <tr> <th colspan="2">FOCOS</th> <th colspan="2">CUARTOS</th> </tr> <tr> <th>CANTIDAD</th> <th>PUNTOS</th> <th>CANTIDAD</th> <th>PUNTOS</th> </tr> </thead> <tbody> <tr> <td>6 - 10</td> <td>15</td> <td>0 A 4</td> <td>0</td> </tr> <tr> <td>11 - 15</td> <td>27</td> <td>5 A 6</td> <td>8</td> </tr> <tr> <td>16 - 20</td> <td>32</td> <td>7 ó MÁS</td> <td>14</td> </tr> <tr> <td>21 +</td> <td>46</td> <td></td> <td></td> </tr> </tbody> </table>		FOCOS		CUARTOS		CANTIDAD	PUNTOS	CANTIDAD	PUNTOS	6 - 10	15	0 A 4	0	11 - 15	27	5 A 6	8	16 - 20	32	7 ó MÁS	14	21 +	46					
FOCOS		CUARTOS																										
CANTIDAD	PUNTOS	CANTIDAD	PUNTOS																									
6 - 10	15	0 A 4	0																									
11 - 15	27	5 A 6	8																									
16 - 20	32	7 ó MÁS	14																									
21 +	46																											

Fuente: Perezbolde, Guillermo (2014). Cómo se calcula el Nivel Socio Económico (NSE) para determinar el target. Página web disponible en <http://www.merca20.com/como-se-calcula-el-nivel-socio-economico-nse-para-determinar-el-target/>

Para determinar en cuál de los siete niveles socioeconómicos cae cada hogar, existen rangos en los puntajes, que se presentan en la Tabla A6.2. Los niveles D y E -en los que se encuentra el menor porcentaje de la población- son quienes carecen de todos, o casi todos, los bienes para satisfacer sus necesidades. En el nivel D+ se encuentran quienes cuentan, al menos, con la mínima infraestructura sanitaria; mientras que el nivel C se caracteriza por contar con infraestructura básica en entretenimiento y tecnología.

Finalmente, los niveles C+, B y A tienen garantizadas todas las necesidades de calidad de vida; distinguiéndose únicamente entre ellos por su capacidad para contar con ahorros y realizar inversiones.

Tabla 3D.2 Nivel socioeconómico obtenido por cada hogar, de acuerdo a su puntaje

NIVEL	PUNTOS
E	Hasta 60
D	Entre 61 y 101
D+	Entre 102 y 156
C	Entre 157 y 191
C+	Entre 192 y 241
A/B	Entre 242 y más

Fuente: Perezbolde, Guillermo (2014, *op. Cit.*).

La batería de preguntas tomadas del *Cuestionario de padres sobre hábitos de lectura de sus hijos*, que se eligieron para dar cuenta de estos indicadores fueron:

- A) ¿Cuál es el máximo grado de estudios del padre?
- B) ¿Cuántos cuartos tiene su vivienda incluyendo la cocina?
- C) ¿Tienen excusado, retrete, sanitario, letrina u hoyo negro?
- D) ¿Hay luz eléctrica de manera regular en su vivienda?
- E) ¿En su vivienda tienen automóvil o camioneta?
- F) ¿En su vivienda tienen regadera?
- G) ¿En su vivienda tienen estufa de gas?
- H) ¿De qué material es la mayor parte del piso de su vivienda?

Con excepción de las preguntas C) y D), el resto cumple exactamente con las especificaciones que dicta la clasificación. En el caso del número de baños, se tomó la decisión de que la respuesta afirmativa a la pregunta C) otorgara el puntaje promedio que se asignaría a los distintos niveles de la clasificación; mientras que una respuesta negativa no otorgara ningún puntaje. La pregunta D), utilizada como aproximación al número de focos en la vivienda, siguió el mismo patrón. Cuando el encuestado respondía afirmativamente se asignó el puntaje promedio de las distintas clasificaciones; mientras que, cuando la respuesta era negativa, se otorgó un puntaje de cero.

Anexo 3E. Ejercicio con diferentes puntos de corte para eliminar valores aberrantes

Tabla 3E.1. Diferencias en puntajes de la prueba ENLACE antes y después del verano (Con 2.5 desviaciones estándar en ambos lados de la distribución)

	Participantes (n =675)			
	Pre ENLACE	Post ENLACE	Diferencia	Prueba T pareada
Grupo de Control (n =411)	489.44 (88.43)	492.52 (91.13)	3.07 (58.30)	0.49 [0.31]
Grupo de Tratamiento (n =264)	494.30 (95.14)	498.69 (95.25)	4.39 (58.77)	0.53 [0.30]
Diferencia (tratamiento – control)	4.86 (91.07)	6.17 (92.74)	1.31 (91.82)	0.18 [0.43]

Nota: entre paréntesis se presentan las desviaciones estándar, y entre corchetes la significancia de los estadísticos t. Fuente: cálculos propios con base en la información recopilada por el CIDE.

Tabla 3E.2. Diferencias en puntajes de la prueba ENLACE antes y después del verano (Con 2 desviaciones estándar en ambos lados de la distribución)

	Participantes (n =660)			
	Pre ENLACE	Post ENLACE	Diferencia	Prueba T pareada
Grupo de Control (n =402)	489.66 (88.78)	493.86 (90.71)	4.20 (55.28)	1.52* [0.06]
Grupo de Tratamiento (n =258)	495.40 (95.02)	500.95 (93.94)	5.55 (55.54)	1.61* [0.05]
Diferencia (tratamiento – control)	5.74 (91.24)	7.09 (91.98)	1.35 (55.34)	0.31 [0.38]

Nota: entre paréntesis se presentan las desviaciones estándar, y entre corchetes la significancia de los estadísticos t. Fuente: cálculos propios con base en la información recopilada por el CIDE.

Tabla 3E.3. Diferencias en puntajes de la prueba ENLACE antes y después del verano (Con 1.5 desviaciones estándar en ambos lados de la distribución)

	Participantes (n =598)			
	Pre ENLACE	Post ENLACE	Diferencia	Prueba T pareada
Grupo de Control (n =364)	493.10 (88.57)	498.13 (90.56)	5.03 (45.70)	2.10** [0.02]
Grupo de Tratamiento (n =234)	500.56 (95.39)	505.63 (93.97)	5.07 (45.57)	1.70** [0.05]
Diferencia (tratamiento – control)	7.46 (91.29)	7.5 (91.90)	0.04 (45.61)	0.01 [0.50]

Nota: entre paréntesis se presentan las desviaciones estándar, y entre corchetes la significancia de los estadísticos t. Fuente: cálculos propios con base en la información recopilada por el CIDE.

CAPÍTULO 4

Conclusiones generales

“Es una coincidencia tan bizarramente improbable que algo tan alucinantemente útil se haya desarrollado sólo por casualidad que algunos pensadores han decidido verlo como la prueba final e irrefutable de la no existencia de Dios.

El razonamiento va más o menos así: ‘Me niego a probar que existo’, dice Dios, ‘porque la prueba niega la fe, y sin la fe no soy nada.’”

Adams Douglas. *Guía del autoestopista galáctico*.

El propósito de esta tesis fue aportar al estudio de implementaciones de la metodología de evaluación de impacto en México. En cada uno de los tres capítulos que la componen se buscó, respectivamente, 1) conocer el grado de implementación de la MEI en el contexto mexicano; 2) presentar las ventajas de un diseño cuasiexperimental frente a un estudio econométrico tradicional en el contexto de evaluación de políticas educativas y 3) mostrar la implementación y resultados obtenidos a través de un estudio que partió de la base de un diseño experimental, aplicado al campo de las políticas educativas.

Mediante la técnica de saturación, en el primer capítulo se recopiló información respecto a las investigaciones que han aplicado la metodología de evaluación de impacto en nuestro país. Los resultados evidencian la preponderancia del programa *PROGRESA/Oportunidades/Prospera* para conducir la agenda de investigación aplicada de esta metodología entre las décadas de los noventa y principios del 2000. Sin embargo, una vez aislado este conjunto de investigaciones, se observa que el desarrollo de la MEI en México ha seguido los patrones tradicionales

encontrados para otros países: un primer momento de desarrollo de investigación que recurre a métodos cuasiexperimentales aplicados a grandes bases de datos como fuente principal de evidencia; un segundo momento con agendas de investigación que parten de diseños experimentales -beneficiadas por estímulos legales gubernamentales aplicados a generar evidencia de efectos de programas públicos- y un tercer momento en el que la academia y otras organizaciones no gubernamentales han ganado preponderancia en el diseño de experimentos que contribuyan al cúmulo de evidencia respecto al efecto de intervenciones en contextos locales y muestras específicas.

Los estudios encontrados dieron cuenta que, hasta el año 2013, las investigaciones más novedosas recaían en este último grupo, quedando espacio para que sean los actores adicionales al sector gubernamental quienes detonen una nueva ola de investigaciones cada vez de mayor alcance y resultados con mayor validez externa. Esa sigue siendo una interrogante que corresponde a una nueva agenda de investigación en las décadas por venir. Lo que no está sujeto a debate, es la relevancia que los comités de ética tendrán sobre las agendas futuras de evaluación e implementación de políticas, aspecto que se recalca como una tarea a 207resultados207 por parte de la creciente comunidad evaluadora.

Para aportar al conjunto de estudios que aplican la MEI en nuestro país, en el segundo capítulo se presentó la evaluación de una política de aleatorización a los turnos matutino y vespertino por parte de las autoridades de uno de los Estados más pequeños de la república mexicana. Tras aproximar los efectos mediante 207resulta cuadrados ordinarios y un modelo panel de efectos fijos, se discute la validez de los resultados encontrados mediante estas técnicas de estimación; llegándose a la conclusión que las escuelas elegidas para contrastar a las que recibieron el tratamiento (la aleatorización) difieren grandemente en sus características respecto de las escuelas tratadas. Por ello, se evalúa la pertinencia de utilizar un modelo cuasiexperimental de

aparejamiento y diferencias en diferencias para obtener un mejor grupo de comparación y con ello resultados con mayor validez.

Los 20 resultados mostraron que los efectos benéficos de la política en términos de reducción en las brechas de desempeño entre los turnos vespertino (el más desfavorecido) y matutino se ven opacadas frente a un incremento en las brechas de las tasas de reprobación, extraedad y repitencia para los alumnos que recibieron la intervención. Estos efectos eran hasta ahora desconocidos, y llaman a realizar un estudio más robusto sobre los efectos de intervenciones destinadas a generar una mayor equidad en las escuelas mexicanas; que partan idealmente de la base de un diseño experimental.

En el tercer capítulo de la tesis se evalúa una intervención, nuevamente de política educativa, que parte de una aleatorización. Comenzando con un marco teórico sobre la importancia del desempeño lector -y la relevancia que cobran los meses de verano para explicar las diferencias en desempeño entre unos alumnos y otros en el nivel de primaria- se presentó la justificación para llevar a cabo la intervención. Consistió en la entrega de tres libros a cada alumno de quinto grado en el nivel de primaria (en el grupo de tratamiento), acompañado de sesiones de capacitación a los padres de familia de los alumnos seleccionados. Elegidos a partir de una lotería al interior de quince planteles de uno de los municipios del Estado de Aguascalientes, la intervención se llevó a cabo acudiendo directamente a los domicilios de los menores, a fin de no contaminar los grupos de tratamiento y control.

Los resultados -medidos a través de los exámenes de la prueba ENLACE de la materia de Español que los menores presentaron al final del año escolar- se contrastaron con los resultados de una prueba ENLACE para el mismo grado escolar que presentó una cohorte de ciclos escolares previos. A pesar del cuidado que se tuvo para garantizar la comparabilidad entre grupos de tratamiento y control, se encontraron diferencias estadísticamente significativas en el

desempeño de las niñas entre grupos previo a la aleatorización. Por lo demás, los grupos resultaron estadísticamente indistinguibles, por lo que se presentaron resultados aglutinados para niños y niñas.

Estos resultados indican que el tratamiento no tuvo un efecto estadísticamente significativo; aunque la dirección y magnitud del impacto fueron acordes con los efectos esperados basados en la literatura nacional e internacional en la materia. En principio, existen dos explicaciones para estos resultados: el tamaño de la muestra elegida para el estudio, y el breve periodo de exposición al tratamiento. Adicionalmente, el nivel socioeconómico de los alumnos mostró ser comparativamente alto frente a los grupos más desfavorecidos que serían, típicamente, quienes tendrían los mayores beneficios de una intervención como ésta.

Las principales implicaciones de las investigaciones contenidas en esta tesis se espera que sean en el campo de la literatura sobre evaluación de impacto en México, así como en el cúmulo de evidencia robusta sobre los efectos de políticas, particularmente educativas, en el contexto mexicano; que sean idealmente extrapolables a otros contextos de América Latina y países en vías de desarrollo, y marquen una dirección sobre el futuro de esta veta de investigación de cara a su relevancia.

GLOSARIO DE SIGLAS

2SLS	Mínimos cuadrados en dos etapas
AGE	Apoyos a la Gestión Escolar.
ANUIES	Asociación Nacional de Universidades e Instituciones de Educación Superior.
ANCOVA	Análisis de Covarianza.
AMAI	Asociación Mexicana de Inteligencia de Mercado y Opinión Pública.
ATE	Efecto promedio del tratamiento.
ATT	Efecto promedio del tratamiento sobre los tratados.
Boltrab	Servicio Bolsa de Trabajo del Servicio Nacional de Empleo.
CIDE	Centro de investigación y Docencia Económicas.
CIMO	Programa Comprehensivo de Modernización y Calidad.
Conafe	Consejo Nacional de Fomento Educativo.
Coneval	Consejo Nacional de Evaluación de la Política de Desarrollo Social.
DIDM	Aparejamiento de diferencias en diferencias.
DSS	Escolaridad de doble turno.

GLOSARIO DE SIGLAS

ENA	Encuesta Nacional de Lectura.
ENLACE	Evaluación Nacional del Logro Académico en Centros Escolares.
Econlit	<i>American Economic Association's Literature Database.</i>
ERIC	<i>Education Resources Information Center.</i>
Estrategia 100 x 100	Estrategia Integral para el Desarrollo Social y Económico de los Municipios con Menor Índice de Desarrollo Humano.
FONAES	Fondo Nacional de Apoyo para las Empresas de Solidaridad.
Fondo PyME	Fondo de Apoyo para la Micro, Pequeña y Mediana Empresa.
JSTOR	Repositorio <i>Journal Storage.</i>
Liconsa	Leche Industrializada Conasupo, S.A. de C.V.
MEI	Metodología de evaluación de impacto.
NBER	<i>National Bureau of Economic Research.</i>
Oportunidades	Programa de Desarrollo Humano <i>Oportunidades.</i>
PARE	Programa para Abatir el Rezago Educativo
PEC	Programa Escuelas de Calidad.
PROBECAT/SICAT	Programa de Becas de Capacitación para Desempleados.

GLOSARIO DE SIGLAS

Procampo	Programa de Apoyos Directos al Campo.
PROGRESA	Programa de Educación, Salud y Alimentación.
Prosoft	Programa para el Desarrollo de la Industria del Software y la Innovación.
Prospera	Programa de Inclusión Social.
RMANCOVA	Análisis de covarianza por mediciones repetidas.
SAEBE	Sistema de Apoyos Económicos a Buscadores de Empleo.
Sedesol	Secretaría de Desarrollo Social
SEP	Secretaría de Educación Pública.
SICAT	Sistema de Capacitación para el Trabajo.
SOFES	Sociedad De Fomento a la Educación Superior.
USAER	Unidad de Servicio de Apoyo a la Educación Regular.