

EVALUACIÓN DE IMPACTO DEL PROGRAMA DE SOPORTE PEDAGÓGICO DEL MINISTERIO DE EDUCACIÓN DEL PERÚ

Marjorie Chinen y Juan Bonilla
Mayo, 2017



FORGE

FORTALECIMIENTO
DE LA GESTIÓN DE LA
EDUCACIÓN EN EL PERÚ

Canada

 **GRADE**

EVALUACIÓN DE IMPACTO DEL PROGRAMA DE SOPORTE PEDAGÓGICO DEL MINISTERIO DE EDUCACIÓN DEL PERÚ¹

Documento Técnico

Proyecto FORGE

Autores: Marjorie Chinen² y Juan Bonilla³

Mayo, 2017

El presente documento se realizó por encargo del Proyecto Fortalecimiento de la Gestión de la Educación en el Perú (FORGE) que es implementado por el Grupo de Análisis para el Desarrollo – GRADE con el apoyo técnico y financiero del Gobierno de Canadá a través de Global Affairs Canada. (Proyecto N° A-034597)

* Los autores reconocen la contribución de muchas personas y organizaciones sin los cuales no hubiera sido posible concluir este informe. Nuestro sincero agradecimiento al Grupo de Análisis para el Desarrollo (GRADE) y al proyecto Fortalecimiento de la Gestión de la Educación en el Perú (FORGE) por el apoyo financiero y técnico. En especial queremos agradecer a Martín Benavides y Patricia Arregui. Asimismo, agradecemos al equipo de la Secretaría de Planificación Estratégica del Ministerio de Educación del Perú por sus comentarios y apoyo durante esta evaluación. Ofrecemos también un especial agradecimiento a Fabiola Cáceres y Andrea Salazar. Asimismo, queremos brindarle nuestro agradecimiento al equipo de Soporte Pedagógico por la información brindada para diseñar y entender la intervención. En especial a Vania Minami y a la ex directora de la Dirección General de Educación Básica Regular, Cecilia Ramírez. En AIR, también reconocemos las contribuciones de Álvaro Ballarín con la preparación de las bases de datos, y de Verónica Ruiz de Castilla por sus comentarios y revisión del informe.

¹ Autora correspondiente: Marjorie Chinen es investigadora Afiliada al Grupo de Análisis para el Desarrollo (GRADE), mchinen@grade.org.pe.

² Juan Bonilla es investigador senior en el American Institutes for Research (AIR).

LISTA DE ACRÓNIMOS

| | |
|---------------|--|
| CITS | Comparative interrupted time series |
| DiD | Difference in difference o modelo en diferencias |
| ECE | Evaluación censal de estudiantes |
| FORGE | Fortalecimiento de la Gestión de la Educación en el Perú |
| GRADE | Grupo de Análisis para el Desarrollo |
| IEP | Programa de Educación Integrada |
| MINEDU | Ministerio de Educación del Perú |
| PSM | Propensity score matching |
| SP | Soporte Pedagógico |
| SPE | Secretaria de Planificación Estratégica |

CONTENIDOS

| | |
|--|-----------|
| Resumen ejecutivo | 5 |
| 1. Introducción | 8 |
| Marco conceptual | 9 |
| Organización del documento | 10 |
| 2. Descripción de Soporte Pedagógico | 11 |
| Teoría de cambio | 14 |
| Criterios de focalización de Soporte Pedagógico | 16 |
| 3. Revisión de la literatura | 18 |
| 4. Objetivos y preguntas de investigación | 23 |
| Objetivos del estudio | 23 |
| Preguntas de investigación | 23 |
| 5. Metodología de la evaluación de impacto | 24 |
| Indicadores de impacto y fuentes de datos | 24 |
| Estrategia empírica | 25 |
| Modelo en diferencias y modelo Comparative Interrupted Time Series | 25 |
| <i>Propensity Score Matching</i> de escuelas | 30 |
| 6. Resultados | 42 |
| Impacto en comprensión lectora y matemática | 42 |
| Impacto en los niveles de logro de la ECE | 45 |
| Análisis de robustez | 48 |
| Análisis de heterogeneidad por nivel de implementación | 50 |
| 7. Discusión y recomendaciones | 54 |
| 8. Consideraciones finales | 57 |
| Bibliografía | 59 |
| Apéndice A: Metodología | 62 |
| Apéndice B: Resultados del Método de Emparejamiento | 64 |

RESUMEN EJECUTIVO

La evaluación de la intervención de Soporte Pedagógico se llevó a cabo con financiamiento del Proyecto Fortalecimiento de la Gestión de la Educación en el Perú (FORGE). Soporte Pedagógico (SP), implementado por la Dirección de Educación Primaria del Ministerio de Educación del Perú, es una intervención que busca mejorar las condiciones en el ámbito educativo que permitan disminuir el atraso escolar y mejorar los niveles de los logros de aprendizajes esperados en las áreas de comunicación, matemática, ciencia y ambiente, y personal social. Para lograr sus objetivos, Soporte Pedagógico interviene en cuatro áreas: a) Fortalecimiento de desempeños en docentes y directivos de las instituciones educativas, b) Refuerzo escolar, c) Materiales y recursos educativos y d) Articulación con la familia y la comunidad. En ese sentido, Soporte Pedagógico es una intervención única en América Latina dado que no hay precedentes de un programa que se enfoque simultáneamente en las cuatro líneas de intervención. Hasta el año 2015, SP se implementó solo en instituciones educativas públicas polidocentes completas de educación primaria ubicadas en zonas urbanas.

El presente estudio evaluó el impacto de SP después de poco menos de un año de implementación utilizando exclusivamente datos de fuentes secundarias recogidas por el Ministerio de Educación. Específicamente, se usaron datos históricos sobre el logro de los aprendizajes de los alumnos de segundo grado de primaria en las áreas de comprensión lectora y matemática recogidos a través de la Evaluación Censal de Estudiantes (ECE). Asimismo, se usaron datos históricos del Censo Escolar. Para estimar los efectos de la intervención de Soporte Pedagógico empleamos los métodos cuasi-experimentales de *Comparative Interrupted Time Series* (CITS) y de Diferencia en Diferencias (DiD), combinado con una estrategia de emparejamiento o *propensity score matching* con estratificación geográfica. La combinación de la técnica de emparejamiento con los modelos de CITS y DiD permitió incrementar la comparabilidad de los dos grupos en torno de variables observables y no observables que no cambian en el tiempo dado que las mismas escuelas fueron analizadas a través del tiempo. El modelo CITS utilizó varios años de información histórica de rendimiento académico, lo cual permitió no solo controlar por las diferencias en medias a través del tiempo, sino también por las diferencias en trayectorias (o pendientes) entre los grupos de intervención y comparación. Como resultado, las estimaciones de impacto presentadas en este estudio se presumen no sesgadas de los efectos de la intervención de SP. No obstante, dado el carácter cuasi-experimental del estudio,

se hicieron varias pruebas de robustez para determinar la estabilidad y consistencia de los efectos encontrados.

Los resultados de este estudio sugieren que el programa de SP, después de poco menos de un año de imperfecta implementación, logró mejorar los resultados de rendimiento académico de los estudiantes intervenidos. Es decir, los estudiantes de segundo grado de primaria que fueron intervenidos obtuvieron puntajes más altos en matemática y comprensión lectora que sus contrapartes del grupo de comparación en la Evaluación Censal Escolar. El tamaño del efecto es de aproximadamente 0.1 desviaciones estándar para comprensión lectora, y de 0.09 desviaciones estándar para matemática.

Los resultados también indican que Soporte Pedagógico tuvo un mayor impacto en aquellos estudiantes que se encontraban en el nivel de logro más bajo o nivel “en inicio” de las pruebas de la ECE; es decir, el programa parece estar ayudando más a aquellos estudiantes que tienen los puntajes más bajos de desempeño académico.

El tamaño de estos efectos, si bien son modestos en términos absolutos, son estadísticamente significativos. Estos resultados son encontrados en un contexto en el que SP no logró implementarse con fidelidad y de acuerdo a lo planeado. Asimismo, cabe resaltar que el programa se implementó en un entorno muy heterogéneo: en 16 regiones y en escuelas polidocentes completas y urbanas, es decir, en las escuelas más grandes y por tanto complejas del país.

Cuando comparamos los resultados de este estudio con otros estudios internacionales realizados en países de mediano y bajo ingreso, observamos que están en el rango encontrado en otros programas con componentes similares. Por ejemplo, en una revisión sistemática de la literatura realizada por el International Initiative for Impact Evaluation (Snilstveit et al., 2015) encontraron que el efecto promedio para intervenciones que desarrollan nuevos enfoques de enseñanza (mediante acompañamiento al docente, materiales instructivos y programación escolar) era de 0.23 desviaciones estándar en lenguaje y 0.14 desviaciones estándar en matemática. Además, este estudio reporta un efecto promedio similar a 0.04 en lenguaje, y 0.16 en matemática para intervenciones que intervienen en varios componentes educativos o “programas multinivel” (Snilstveit et al., 2015).

Este estudio también investigó la heterogeneidad de los efectos encontrados al nivel de implementación de dos componentes del programa, tales como el porcentaje de docentes fortalecidos y de acompañantes pedagógicos contratados antes de la primera mitad del año escolar. Específicamente, investigamos si el efecto de SP era mayor

en las unidades ejecutoras en donde más del 90 por ciento de dichas contrataciones se habían realizado antes de mediados del año escolar (o mes de julio de 2015). No obstante, no encontramos diferencias notables en comparación con los efectos obtenidos con toda la muestra. De igual modo, tampoco encontramos diferencias significativas al investigar los efectos del programa entre aquellas escuelas de unidades ejecutoras que a julio de 2015 habían implementado el programa al 100%. Es decir, aun cuando se restringió la muestra a las escuelas intervenidas con “alto” nivel de implementación (y sus escuelas emparejadas), el tamaño del efecto encontrado fue de aproximadamente 0.10 desviaciones estándar para comprensión y matemática, respectivamente. Estos efectos son muy similares a los obtenidos con toda la muestra analítica.

Estos resultados se podrían explicar, en parte, por la poca variabilidad que existía en las escuelas intervenidas entorno a los dos indicadores “próximos” de implementación del programa. Pero también estos resultados podrían estar sugiriendo que la contratación del personal clave del programa es tan solo uno de los muchos componentes o pasos necesarios para implementar el programa con fidelidad. Por ejemplo, si bien el personal clave pudo haberse contratado a tiempo, tal vez no recibió suficiente capacitación para ejercer las labores del programa; o tal vez sí se dieron suficientes capacitaciones, pero los docentes necesitaban más práctica y retroalimentación para realizar su labor de acuerdo a lo planeado por la intervención.

INTRODUCCIÓN

Este informe describe los resultados de la evaluación de impacto de la intervención de Soporte Pedagógico del Ministerio de Educación del Perú. Esta intervención tiene como objetivo mejorar la calidad del servicio educativo brindado en las escuelas polidocentes públicas de primaria para mejorar los niveles de aprendizaje de los estudiantes. Para lograr sus objetivos, Soporte Pedagógico interviene en cuatro áreas: a) Fortalecimiento de desempeños en docentes y directivos de las instituciones educativas, b) Refuerzo escolar, c) Materiales y recursos educativos y d) Articulación con la familia y la comunidad.

Esta evaluación es un estudio de carácter cuasi-experimental que hace uso exclusivo de datos secundarios recogidos por el Ministerio de Educación. El estudio emplea los modelos de *Comparative Interrupted Time Series* y de Diferencia en Diferencias, en combinación con una estrategia de emparejamiento de escuelas para estimar el impacto de Soporte Pedagógico. La combinación de la técnica de emparejamiento con los modelos de CITS y DiD, y el uso de datos históricos y longitudinales, permitió incrementar la comparabilidad de los dos grupos antes de la introducción del programa. La comparabilidad entre los grupos de intervención y comparación fue lograda en términos de características observables y no observables que no cambian en el tiempo, dado que las mismas escuelas fueron analizadas y comparadas a través del tiempo. Asimismo, se logró la comparabilidad de los grupos no solo en los niveles medios de diferentes variables, sino también en las trayectorias (o pendientes) de rendimiento de los dos grupos. Como resultado, las estimaciones de impacto presentadas en este estudio se presumen no sesgadas de los efectos del programa de SP. No obstante, dado el carácter cuasi-experimental del estudio, se hicieron pruebas de robustez para determinar la estabilidad y consistencia de los efectos encontrados.

Los resultados presentados en este informe se basan en datos anuales e históricos recogidos de las escuelas desde el año 2009 usando fuentes secundarias recogidas por el MINEDU como la Evaluación Censal Escolar y el Censo Escolar. Estas bases incluyen una serie de variables tales como el rendimiento académico de los estudiantes en matemática y comprensión lectora, datos de la escuela, de su infraestructura, de los materiales recibidos, entre otros. El Programa de SP comenzó las actividades de implementación en julio de 2014. Los datos de rendimiento que se levantaron después de iniciado el programa se recogieron en el mes de noviembre de 2014 y 2015.

Tal como se detalla a continuación, esta evaluación encontró que el programa SP mejoró

los resultados de comprensión lectora y matemática de los estudiantes de segundo grado de primaria. Estos impactos son de un tamaño modesto pero estadísticamente significativo y en el rango de los efectos encontrados en otros programas educativos con componentes similares.

Dado que la intervención de Soporte Pedagógico se evaluó tan solo después de poco menos de un año de su introducción en diferentes regiones y escuelas del Perú, concluimos que esta intervención tiene el potencial de mejorar los rendimientos de los estudiantes en el mediano plazo. No obstante, dada la complejidad de la intervención será importante que el MINEDU desarrolle mecanismos eficientes de monitoreo de las actividades del programa, y se establezcan procesos claros para poder actuar sobre las dificultades y desafíos que se encuentren en el camino, tanto en las escuelas como en las unidades ejecutoras.

Marco conceptual

En el Perú, los niños y niñas que asisten a la escuela primaria se encuentran generalmente entre los 6 y 11 años. En 2014, el 93.2 por ciento de niños y niñas en esta franja de edades asiste a una escuela de educación primaria (UNICEF, 2014). A inicios del año 2015, este grupo era equivalente a 3 millones y medio de estudiantes distribuidos en diferentes ámbitos geográficos (urbano y rurales), instituciones educativas con diferentes número de docentes (polidocente completa, polidocente incompleta, o multigrado, y unidocente), modalidad (educación básica regular y básica especial) y tipo de gestión (pública de gestión directa, pública de gestión privada, y de gestión privada). Los estudiantes de primaria están distribuidos en seis grados, y la mayor proporción de estudiantes de primaria están matriculados en instituciones educativas polidocentes completas (77 por ciento).

Según el Ministerio de Educación (2015a), en el Perú un 6.5 por ciento de los niños y niñas que están matriculados en la educación primaria asiste a un grado inferior al que le corresponde para su edad. Sin embargo, mientras que en las zonas urbanas el porcentaje de atraso escolar es 4.4 por ciento, los niños y niñas en zonas rurales tienen un atraso del 14.8 por ciento. Esta brecha también se manifiesta en la edad de culminación de primaria: un 88.3 por ciento de las niñas y niños termina la primaria entre los 12 y 13 años, comparado con un 68.6 por ciento en las zonas rurales. El atraso también es mucho mayor para estudiantes de lengua materna indígena: un 16.3 por ciento, frente a un 5.7 por ciento entre los estudiantes que tienen el español como lengua materna. Entre los niños que

tienen una lengua amazónica como primer idioma, el atraso está por encima del 20 por ciento.

Asimismo, los resultados de la ECE en matemática y comprensión lectora muestran una mejora a través del tiempo, especialmente en el año 2014: si bien de 2012 a 2013 los resultados en comprensión lectora y matemática aumentaron 2.1 y 4.1 puntos, de 2013 a 2014 estas mismas pruebas aumentaron 10.5 y 9.1 puntos porcentuales, respectivamente. Sin embargo, esto equivale a que en 2014 solo el 48.4 por ciento de los estudiantes de instituciones educativas polidocentes alcanzó el nivel más alto o satisfactorio¹ de aprendizaje en comprensión lectora; y solo el 28,7 por ciento lo hizo en matemática (Resultados Evaluación Censal, 2014).

En este contexto, el Ministerio de Educación del Perú, mediante la Dirección de Educación Primaria, implementó la intervención de Soporte Pedagógico con el objetivo de brindar apoyo técnico pedagógico a las instituciones educativas primarias polidocentes completas de zonas urbanas. La intervención espera movilizar estrategias metodológicas que garanticen la atención a la diversidad y el uso de las herramientas curriculares y pedagógicas de manera efectiva, y así mejorar la calidad de los aprendizajes y lograr que un mayor porcentaje de estudiantes alcance mejores resultados.

El presente estudio tiene el objetivo principal de evaluar el impacto de Soporte Pedagógico sobre los niveles de logro de aprendizaje en las áreas de comprensión lectora y matemática de segundo grado de primaria.

Organización del documento

El documento presenta a continuación la descripción del programa de Soporte Pedagógico (Capítulo 2), seguido por la revisión de la literatura (Capítulo 3), los objetivos y preguntas de investigación (Capítulo 4) y la metodología del estudio (Capítulo 5). El capítulo 6 presenta los resultados de este estudio y los análisis de robustez. Finalmente, este reporte concluye discutiendo los resultados de la evaluación de impacto en el Capítulo 7 y con algunas consideraciones finales en el Capítulo 8.

1 La prueba diseñada por la Oficina de la Medición de la Calidad de los Aprendizajes del MINEDU cuenta con 3 posibles niveles de logros de aprendizaje: en inicio (no lograron los aprendizajes esperados), en proceso (no lograron los aprendizajes esperados pero están en proceso de lograrlo) y satisfactorio (lograron los aprendizajes esperados).

2. DESCRIPCIÓN DE SOPORTE PEDAGÓGICO

Soporte Pedagógico es una estrategia de apoyo integral a las instituciones educativas públicas polidocentes completas de educación primaria y de zonas urbanas. La intervención tiene como finalidad generar condiciones en el ámbito educativo que permitan disminuir el atraso escolar, y mejorar los niveles de aprendizaje en las áreas de comunicación, matemática, ciencia y ambiente, personal social. SP se define como una intervención que involucra un conjunto de actividades y procesos que fomentan la interacción entre los docentes, directivos, padres y madres de familia, y demás actores de la comunidad educativa para resolver problemas que incidan en la mejora de los aprendizajes de los estudiantes. Estas acciones han sido diseñadas en concordancia con los lineamientos de política curricular vigente y el marco del buen desempeño docente y del directivo.

Específicamente, SP busca el logro de cuatro objetivos² relacionados con los docentes, el equipo directivo, los padres de familia y la comunidad:

- Promover el adecuado desempeño pedagógico del docente que incida favorablemente en el desarrollo de las competencias de los estudiantes.
- Fortalecer prácticas de liderazgo pedagógico para la gestión escolar por parte de los directivos; articulada a la mejora de los desempeños docentes y el desarrollo de competencias en los estudiantes.
- Involucrar a los padres de familia en el acompañamiento y apoyo al proceso de aprendizaje de sus hijos e hijas.
- Brindar a los estudiantes atención a través de las jornadas y encuentros con padres en relación con los aprendizajes logrados durante el grado.

Para lograr dichos objetivos, Soporte Pedagógico cuenta con cuatro distintas líneas de intervención que involucran a los docentes y directivos, estudiantes, los padres de familia y los miembros de la comunidad. Las cuatro líneas de intervención son las siguientes:

2 De acuerdo a la RSG N° 007-2015-MINEDU.

1. Fortalecimiento de desempeños en docentes y directivos de la institución educativa

Esta línea busca generar espacios formativos y de motivación, mediante los cuales se pueda fortalecer el desempeño profesional del docente y de los directivos, teniendo en cuenta los enfoques pedagógicos, las capacidades socio-afectivas, las estrategias de aprendizaje, los recursos educativos, evaluación, y la gestión de aprendizaje. Adicionalmente, esta línea promueve la coordinación y supervisión de los equipos directivos en las instituciones educativas, para que verifiquen y orienten el desarrollo adecuado de los procesos pedagógicos, y el uso oportuno de recursos empleados en el aula con el objetivo de mejorar los aprendizajes de los estudiantes. Estas actividades se desarrollan a través de: a) capacitaciones y talleres para docentes, los cuales buscan fortalecer el desempeño profesional de los docentes sobre la base de la identificación de necesidades pedagógicas, replicando las innovaciones y buenas prácticas pedagógicas; b) acompañamiento pedagógico, el cual consiste en el desarrollo de una secuencia de acciones que busca fortalecer las competencias pedagógicas de los docentes, y promover el logro de los aprendizajes; c) grupos de interaprendizaje, que son espacios de aprendizaje que buscan generar reflexión colectiva en la institución educativa. En estos grupos se busca plantear propuestas y alternativas para mejorar el desempeño docente y fortalecer las capacidades pedagógicas de los directivos en comunicación, matemática, ciencia y ambiente, y personal social; d) programas de actualización en didáctica para docentes, los cuales constan de al menos 400 horas lectivas que permiten una certificación progresiva y la obtención de un diploma de actualización; y e) asesoramiento virtual, el cual consiste en un servicio de consulta permanente en línea con el objetivo de asesorar sobre aspectos pedagógicos y de gestión.

2. Refuerzo escolar

El refuerzo está dirigido a estudiantes de primero, segundo y tercer grado de primaria que muestran retraso para alcanzar el nivel de logro óptimo en matemática y/o comunicación de acuerdo al grado que cursan. Esta línea de intervención considera un conjunto de actividades y estrategias pedagógicas para desarrollar las competencias específicas que requieren mayor acompañamiento para ser alcanzadas. Las sesiones de aprendizaje tienen una duración de 60 minutos, y se ofrecen a grupos con un máximo de seis

estudiantes después del horario oficial de clase. La implementación de refuerzo escolar está a cargo del docente fortaleza, el cual debe realizar las siguientes actividades: a) coordinación para la implementación de la línea de acción con el director y docente de aula; b) la identificación de los estudiantes que necesitan mayor acompañamiento para el logro de las competencias en comunicación y matemática; c) evaluación a los estudiantes sobre las competencias reforzadas; y d) coordinación e información del nivel de avance de logros de los estudiantes en refuerzo escolar a directivos y docentes de la institución educativa.

3. Materiales y recursos educativos

Los materiales y recursos educativos son paquetes didácticos y de ayuda a la labor pedagógica destinados a los docentes, directivos, estudiantes y padres de familia. Para los docentes, estos paquetes didácticos incluyen unidades y sesiones de aprendizaje, rutas de aprendizaje, y el kit de evaluación de segundo y cuarto grado. Para los estudiantes, estos materiales incluyen cuadernos de trabajo, textos escolares y kit de útiles escolares. Para el aula, estos materiales son entregados en la forma de bibliotecas de aula, material concreto de las áreas curriculares de comunicación, matemática, personal social y ciencia y ambiente. Finalmente, para los padres de familia, estos paquetes incluyen guía de padres de familia, fascículos y material concreto para orientar aprendizajes de los hijos e hijas.

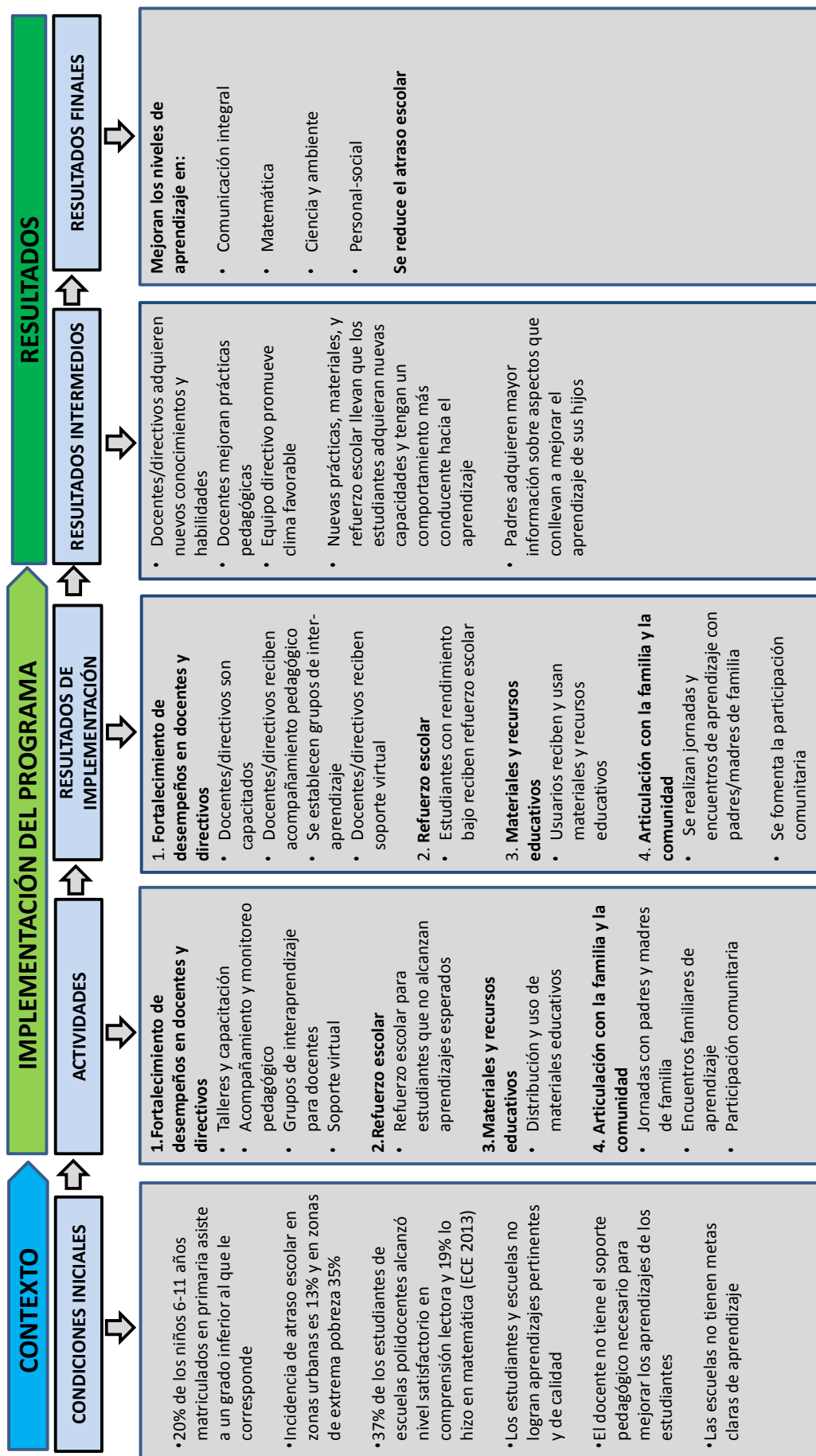
4. Articulación con la familia y la comunidad

Esta línea de acción tiene como objetivo promover la interacción de la comunidad y la familia en torno a actividades que generen condiciones para los aprendizajes deseados en los estudiantes. Para lograr estos objetivos se establecen las siguientes actividades: a) jornadas con padres y madres de familia. Son talleres lúdicos, reflexivos y de intercambio de experiencias donde se brindan estrategias y orientaciones para que los padres puedan acompañar los aprendizajes de sus hijos en situaciones cotidianas; b) encuentros familiares de aprendizaje. Son talleres dirigidos a todas las familias de primero, segundo y tercer grado de primaria en los cuales los padres de familia e hijos pueden compartir, divertirse y aprender; c) finalmente, esta línea de intervención involucra actividades que fomentan la participación comunitaria con el objetivo de establecer vínculos entre la escuela y la comunidad con miras a mejorar el logro de los aprendizajes en los estudiantes.

Teoría de cambio

La evaluación de impacto está basada en la teoría de cambio de la intervención que presentamos en la Figura 1. La teoría de cambio resume de manera simplificada la secuencia lógica de eventos que en principio deberían ocurrir para que se logre el objetivo final de esta intervención, que es la mejora del rendimiento académico de los estudiantes. En primer lugar, la teoría de cambio presenta las condiciones iniciales, es decir, la situación educativa de los estudiantes de primaria de escuelas polidocentes completas de zonas urbanas que motivaron el surgimiento de la intervención. En segundo lugar se presentan las cuatro líneas que Soporte Pedagógico interviene directamente y describe brevemente cada una de esas líneas. Tercero, la teoría de cambio describe lo que comúnmente se denominan “outputs” o, en este caso, resultados de implementación, que representan la realización de las actividades de la intervención, es decir, las actividades que deberían ocurrir para que el programa se implemente con fidelidad. En cuarto lugar se presentan los resultados intermedios, o resultados que deberían ocurrir de acuerdo a esta intervención, para que mejoren el rendimiento académico y se reduzca el atraso escolar. En este caso, los resultados intermedios están relacionados a la mejora de las condiciones en el ámbito educativo. Finalmente, y en quinto lugar, la teoría de cambio presenta los resultados finales que esta intervención busca alcanzar.

Figura 1. Teoría de cambio de la intervención de Soporte Pedagógico



Criterios de focalización de Soporte Pedagógico

Soporte Pedagógico inició actividades aproximadamente en julio de 2014. En este primer año, el programa se focalizó en escuelas primarias, públicas, polidocentes completas, y urbanas de 10 regiones del país. Estas escuelas fueron escogidas de manera que cumplan con la condición de haber participado en la Evaluación Censal de Estudiantes durante los años 2011, 2012 y 2013. Asimismo, se seleccionaron las regiones que mostraban una mejora continua en el resultado promedio de esos tres años (variación positiva en puntajes promedio), pero que todavía tenían una proporción moderada de alumnos en el nivel “no satisfactorio” que se encuentre en los siguientes rangos: que tengan aproximadamente entre 15% y 30% de los alumnos en el nivel “no satisfactorio” en comunicación, y entre 40% y 60% en matemática. Sin embargo, estos criterios no fueron exclusivos, y también se usaron otros criterios adicionales, de carácter político y administrativo, para asignar escuelas al programa. Por ejemplo, se agregó Lima Metropolitana por ser una región atípica y con alta concentración de estudiantes; se incorporó San Martín, dado que SP coincidía con el plan estratégico de la región; Huancavelica y Apurímac, por ser zonas con altos índices de pobreza; Loreto y Ucayali, por factores institucionales; y Piura, por factores políticos.

En el segundo periodo o año 2015 se agregaron nuevas escuelas al programa en siete de las regiones intervenidas en 2014 (Apurímac, Ayacucho, Cusco, Cajamarca, Huancavelica, Lima Metropolitana, Loreto, Piura). Asimismo, se agregaron cuatro regiones: Ica, La Libertad, Lambayeque y Lima Provincias por ser regiones con alta población de estudiantes y docentes en zonas urbanas, y porque presentaban condiciones políticas favorables para la coordinación e implementación del programa.³ Finalmente, aproximadamente en agosto de 2015, se incorporó la UGEL de Islay del departamento de Arequipa a solicitud de esta región.

Si bien el programa priorizó escuelas polidocentes completas, urbanas, públicas, también incorporó algunas escuelas rurales y polidocentes multigrado. En su gran mayoría estas escuelas fueron agregadas a solicitud de las mismas regiones y en reemplazo a otras escuelas propuestas inicialmente por el programa.⁴

En total, el número de escuelas intervenidas en 2015 asciende a 3,180. La Tabla 1 presenta la distribución del número de escuelas intervenidas por región y por año. De las 3,180 escuelas intervenidas, 3,006 (94.5%) correspondían a escuelas urbanas y

3 La región de Madre de Dios no estaba prevista inicialmente para formar parte del programa, pero fue incorporada a partir de una solicitud directa de la región.

4 De acuerdo a la base de focalización de Soporte Pedagógico, en 2014 se intervinieron 22 escuelas rurales, mientras que en 2015 se intervinieron 82 escuelas rurales.

polidocentes completas.

En el 2015, las escuelas con SP representaban aproximadamente el 23 por ciento de las escuelas polidocentes completas, el 20 por ciento de las escuelas urbanas y aproximadamente el 12 por ciento de las escuelas públicas del Perú.

Tabla 1.
Número de escuelas beneficiarias de Soporte Pedagógico por año de inserción al programa⁵

| Departamento | Número de escuelas beneficiarias de Soporte Pedagógico según año de inserción | |
|---------------|---|-------|
| | 2014 | 2015 |
| Apurímac | 43 | 138 |
| Arequipa | 0 | 17 |
| Ayacucho | 57 | 207 |
| Cajamarca | 89 | 221 |
| Callao | 2 | 2 |
| Cusco | 103 | 256 |
| Huancavelica | 26 | 91 |
| Ica | 0 | 115 |
| La Libertad | 0 | 222 |
| Lambayeque | 0 | 114 |
| Lima | 702 | 1,073 |
| Loreto | 119 | 205 |
| Madre de Dios | 0 | 31 |
| Piura | 50 | 327 |
| San Martín | 104 | 116 |
| Ucayali | 33 | 45 |
| Total | 1328 | 3180 |

Fuente. Base de focalización de Soporte Pedagógico.

⁵ Los datos presentados en esta tabla se obtuvieron directamente de la base de focalización de Soporte Pedagógico entregada a los investigadores.

3. REVISIÓN DE LA LITERATURA

La intervención de Soporte Pedagógico es única en América Latina, ya que no hay precedentes de una intervención que se enfoque simultáneamente en las cuatro líneas de intervención. A diferencia de SP, la mayor parte de los programas de mejora educativa en la región se han enfocado en una o dos de estas líneas de intervención. Por lo tanto, esta evaluación contribuye a la literatura al generar evidencias sobre el impacto de corto plazo de intervenir varias áreas de la educación primaria sobre el rendimiento académico de los estudiantes.

No obstante, sí existen evidencias sobre la efectividad de las líneas de intervención de SP por separado. En general, la evidencia sobre la efectividad de intervenir algunas de las líneas de intervención es mixta y procede principalmente de países fuera de América Latina. A pesar de que existen muchos programas en la región que tratan áreas similares a los componentes de SP, la mayor parte de ellos no ha sido sometida a una evaluación rigurosa (Villegas-Reimers, 1998; Navarro & Verdisco, 2000).

La investigación cuantitativa generalmente no ha detectado impactos positivos de programas de capacitación de docentes en el aprendizaje de los estudiantes (Chingos & Peterson, 2011), aunque evidencia más reciente ha cuestionado esta conclusión (Boyd et al., 2009). Según el Banco Mundial, muchos programas de desarrollo de los docentes no son efectivos porque no están bien diseñados ni integrados con el trabajo del profesor (Devarajan & Reinikka, 2003). Por ejemplo, Jacob y Lefgren (2002) evaluaron un programa de desarrollo del profesorado en escuelas con bajo rendimiento en Estados Unidos, y concluyeron que el programa no tuvo ningún efecto positivo significativo en los resultados de los alumnos. Estas conclusiones fueron robustas usando diferentes modelos de estimación y no se encontraron diferencias entre capacidad, género, raza o situación económica. Los autores atribuyeron la falta de resultados en parte a la falta de estructura en la capacitación de los docentes. Chingos y Peterson (2011) sugieren que una dificultad de evaluar la efectividad de estos programas en conjunto es que son muy diferentes entre sí. Según explican estos investigadores, es posible que la efectividad de estos programas dependa de su estructura y contenidos. Por tanto, podría haber programas que sí sean muy efectivos, aunque no se detecte un impacto sistemático de los programas de capacitación.

Boyd et al. (2009) evaluaron 31 programas de desarrollo de profesores de primaria en Nueva York, y sus resultados dan apoyo a esta hipótesis: los autores concluyeron que hay variación significativa en la efectividad de estos programas. Según los autores, los

resultados dependen de si la preparación está directamente vinculada a la práctica de la enseñanza. Por ejemplo, aquellos programas que enfatizan tener más supervisión de las prácticas de enseñanza proporcionan docentes que son mucho más productivos durante su primer año de trabajo. Los efectos son significativos, con una magnitud de efecto similar a la del primer año de experiencia como profesor. Los autores también encontraron evidencia de que los programas que enfatizan contenidos en matemática tienen resultados, aunque estos no son visibles hasta el segundo año de instrucción. Parece ser que aquellos maestros con buena preparación en cuestiones del día a día son relativamente más eficaces en su primer año, mientras que aquellos con un conocimiento fuerte del contenido son capaces de hacer uso de ese conocimiento durante su segundo año.

En cuanto a la línea de intervención de Refuerzo Escolar de Soporte Pedagógico, la investigación cuantitativa indica que estos programas pueden ser efectivos, aunque los efectos generalmente son moderados. Por ejemplo, Lavy y Schlosser (2004) evaluaron un programa en Israel que proporciona instrucción adicional para los estudiantes de bajo rendimiento. El programa tenía como objetivo preparar a los estudiantes para exámenes de matrícula. Para estimar el efecto del programa, se utilizaron escuelas de comparación con características similares que habían sido seleccionadas para llevar a cabo el programa más adelante. Los resultados fueron que el programa aumentó en algo más de tres puntos porcentuales la tasa de matriculación. A pesar de estos resultados, los autores concluyeron que este tipo de intervención es menos rentable que programas basados en incentivos para estudiantes y docentes. Cantrell et al. (2013) también encontraron resultados moderados en un programa de refuerzo en lectura para adolescentes con poca motivación y bajo rendimiento en Estados Unidos. La intervención se ofreció a estudiantes de sexto y noveno grado, y se hizo un seguimiento de sus resultados durante cuatro años. Para estimar la efectividad del programa también se hizo un seguimiento de dos grupos de control (uno para sexto grado y otro para noveno) durante ese periodo de tiempo. Los resultados mostraron efectos positivos de la intervención de lectura en los alumnos de noveno grado, pero no en los de sexto grado.

Banerjee et al. (2007) encontraron resultados más fuertes en dos experimentos realizados en las escuelas en zonas urbanas de la India. El primer programa de educación contrató a mujeres jóvenes para enseñar a los estudiantes con retraso en habilidades básicas de lectura, escritura y aritmética. Como resultado del programa, el puntaje promedio de los niños en las escuelas intervenidas aumentó 0.28 desviaciones estándar. Además, el efecto fue especialmente pronunciado en los niños con resultados de aprendizaje más bajos. La segunda intervención fue un programa de aprendizaje de

matemática por computadora. El programa aumentó las calificaciones de matemática en 0.47 desviaciones estándar. Un año después de que los programas finalizaran, las ganancias iniciales se mantuvieron significativas para los niños intervenidos, aunque el efecto se redujo hasta 0.10 desviaciones estándar.

Otra línea de investigación que sugiere que los programas de refuerzo pueden ser efectivos son los estudios que muestran la efectividad de programas con instrucción adaptada al ritmo de aprendizaje de los alumnos. Un ejemplo de este tipo de programa es el Programa de Educación Integrada (IEP), un programa de lectura para niños de primaria en Sudáfrica (Piper, 2009). El IEP es un programa de formación de docentes pero, a diferencia de otras intervenciones, es un programa de instrucción adaptada a la velocidad de aprendizaje de los estudiantes. Para llevar esto a cabo, los estudiantes tienen que tomar un examen al principio del programa, y se enseña a los profesores a cómo evaluar el progreso de los alumnos y a cómo enseñarles según sus capacidades. Los resultados de la intervención mostraron que los niños en el grupo de intervención aprendieron a leer dos a tres veces más rápido que los niños de las escuelas de control. La magnitud de los efectos del programa es coherente con otros estudios que indican que la adaptabilidad en la instrucción es clave para hacer que los programas de formación del profesorado sean eficaces. Conn (2014) revisó un conjunto de intervenciones pedagógicas en África y encontró que, entre las intervenciones enfocadas en la formación del profesorado, el efecto medio de los programas con instrucción adaptada al ritmo de aprendizaje de los estudiantes fue de 0.42 desviaciones estándar, mientras que aquellos sin este tipo de enseñanza tuvieron un efecto medio de 0.12 desviaciones estándar.

Los resultados de intervenciones basadas en la tercera línea de acción de Soporte Pedagógico, la entrega de materiales y recursos educativos, no han demostrado hasta ahora ser eficaces. Por ejemplo, una evaluación aleatoria llevada a cabo por Glewwe et al. (2002) en zonas rurales de Kenia encontró que proporcionar libros de texto no mejora el promedio de las pruebas. Aunque la intervención sí mejoró los resultados de los que ya eran buenos estudiantes, tuvo muy poco efecto en el resto del alumnado. Sin embargo, este estudio contaba con algunas limitaciones importantes, concretamente a) los libros de texto estaban escritos en inglés, que era la tercera lengua para la mayoría de los estudiantes, y b) el plan de estudios en Kenia está orientado principalmente hacia los estudiantes académicamente fuertes, lo que dificulta que programas diseñados para el resto de los estudiantes sean efectivos.

En cuanto al aumento de la autonomía en las decisiones de las escuelas, parece también tener una efectividad que varía caso por caso. Hanushek et al. (2012), en un estudio con datos de las pruebas PISA que incluía 42 países y datos del año 2000 al

año 2009, encontraron que los efectos de la autonomía escolar son positivos en países desarrollados y negativos en países en vías de desarrollo. Los autores encontraron que estos efectos son robustos con diferentes estimaciones empíricas. Además, el hallazgo de la interacción entre nivel de desarrollo e impacto de la autonomía se manifiesta en los resultados de los estudiantes en matemática, lectura y ciencias. A diferencia de King y Özler (1998), los autores concluyeron que el aspecto de la autonomía escolar que más importa es la decisión de los contenidos académicos. Aunque las decisiones del personal y la autonomía presupuestaria también son relevantes, según Hanushek et al. (2012) estos aspectos guardan menos relación con el aprendizaje de los estudiantes.

Sin embargo, King y Özler (1998) sí encontraron efectos positivos en su estudio sobre la reforma educativa de Nicaragua a principios de los años 90, la cual tenía como objetivo trasladar más poder de gestión a las escuelas. Los resultados mostraron que el rendimiento de los estudiantes tiene una relación positiva con el grado de autonomía en la toma de decisiones de la escuela. Según los autores, este efecto es especialmente fuerte en aquellas escuelas con poder de decisión sobre la elección de sus docentes y el monitoreo de los mismos.

Un mecanismo que puede explicar por qué una mayor autonomía de las escuelas puede ser efectiva es el hecho de que involucrar a padres y miembros de la comunidad puede aumentar la rendición de cuentas de los profesores. Un ejemplo de una intervención de este tipo es el programa Early Grade Reading Assessment Plus, un programa de lectura para niños de primaria (Piper & Korda, 2010). La intervención creó dos tipos diferentes de grupos de tratamiento para diferenciar los efectos de la capacitación y los efectos de aumentar la responsabilidad de los profesores. En el grupo de tratamiento “completo”, los maestros obtuvieron capacitación y se les hizo responsables de los resultados de sus estudiantes. Para llevar esta última parte a cabo, se facilitó a los padres datos del rendimiento de los estudiantes. En el grupo de tratamiento “parcial”, también se hizo responsables a los maestros de los resultados de sus alumnos, pero no hubo componente de capacitación. Los resultados del programa mostraron que los estudiantes en el grupo de tratamiento completo tenían puntuaciones más altas en todas las habilidades de lectura que sus homólogos en los otros grupos. Sin embargo, los estudiantes en el grupo “parcial” de tratamiento, donde el aumento de responsabilidad de los maestros se llevó a cabo sin el componente de capacitación del profesorado, también superaron significativamente los niños en el grupo de control. Ambos resultados combinados sugieren que tanto la formación del profesorado y la rendición de cuentas contribuyen a la mejora de la lectura en edad temprana. Al igual que con la formación de profesores, otras intervenciones basadas en aumentar los incentivos y la responsabilidad de los profesores también han encontrado resultados

similares (Kremer et al 2013; McEwan, 2014; Muralidharan y Sundararaman 2011; Duflo, Hanna, y Ryan 2012).

Birte y otros autores (2015) recientemente publicaron una revisión sistemática de la literatura en el área de educación, que tenía como objetivo sintetizar la evidencia de los efectos de programas de educación en el acceso y el aprendizaje en países de bajo y mediano ingreso. En esta revisión examinaron distintos tipos de programas, y encontraron que los programas denotados como pedagogía estructurada presentaban los efectos más grandes y consistentes sobre el aprendizaje. Las intervenciones de pedagogía estructurada típicamente involucran el desarrollo de nuevos contenidos enfocados en un área en particular, materiales para estudiantes y docentes, y capacitaciones de corto plazo para docentes en la entrega de estos nuevos contenidos. Estas intervenciones tienen como objetivo principal cambiar las prácticas pedagógicas existentes. Los autores encontraron que para que las intervenciones de pedagogía estructurada funcionen se necesita que ocurran varias condiciones. Por ejemplo, que los materiales sean de suficiente calidad, que sean provistos en las cantidades adecuadas y a tiempo en el año escolar. Las capacitaciones tienen que ser de suficiente duración y calidad, y los docentes tienen que poder entender y apoderarse (*buy-in*) de la capacitación. Asimismo, uno de los desafíos más grandes de este tipo de intervenciones es que las metodologías y contenidos introducidos sean aplicados en las aulas con un nivel razonable de fidelidad y por un periodo adecuado.

4. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN

Objetivos del estudio

El principal objetivo de este estudio es evaluar el impacto de corto plazo de la intervención de Soporte Pedagógico usando exclusivamente información existente y disponible del Ministerio de Educación. Asimismo, la evaluación se llevó a cabo sin interrumpir el proceso de implementación del programa. Bajo estas restricciones, fue posible investigar el impacto de la intervención de SP sobre dos áreas curriculares, comunicación integral y matemática, y en un solo grado académico, segundo grado de primaria.

Preguntas de investigación

Este estudio pretende contestar las siguientes preguntas de investigación:

1. ¿Cuál es el impacto de Soporte Pedagógico sobre los resultados de aprendizaje y los niveles de logro de los estudiantes de segundo grado de primaria en el área de comprensión lectora?
2. ¿Cuál es el impacto de Soporte Pedagógico sobre los resultados de aprendizaje y los niveles de logro de los estudiantes de segundo grado de primaria en el área de matemática?
3. ¿Cómo varía el impacto de Soporte Pedagógico cuando solo se examinan las escuelas con “altos niveles de implementación”?

5. METODOLOGÍA DE LA EVALUACIÓN DE IMPACTO

Indicadores de impacto y fuentes de datos

Para contestar las preguntas de investigación usamos los siguientes indicadores de impacto resumidos en la Tabla 2.

Tabla 2. Indicadores de impacto

| Variable | Indicador(es) | Fuente |
|---|---|--|
| Para Soporte Pedagógico | | |
| 1. Resultados de aprendizaje de los estudiantes de segundo grado de primaria en el área de comunicación | - Puntaje promedio a nivel de estudiante en comprensión lectora; - Porcentaje de alumnos por niveles de logro (satisfactorio, en proceso, en inicio) en comprensión lectora. | Evaluación Censal de Estudiantes, 2015 |
| 2. Resultados de aprendizaje de los estudiantes de segundo grado de primaria en el área de matemática | - Puntaje promedio a nivel de estudiante en matemática; - Porcentaje de alumnos por niveles de logro (satisfactorio, en proceso, en inicio) en matemática. | Evaluación Censal de Estudiantes, 2015 |

La Evaluación Censal de Estudiantes (ECE). La ECE es una prueba estandarizada que recoge el MINEDU todos los años en las áreas de comprensión lectora y matemática a los alumnos de segundo grado de primaria. Esta prueba se levanta desde el año 2007 y en principio los resultados son comparables en el tiempo.⁶ El enfoque de la ECE en matemática y comprensión lectora se basa en el supuesto de que estas competencias son habilidades esenciales para el aprendizaje de otras áreas curriculares. La prueba de matemática incluye preguntas sobre números, relaciones y funciones.⁷ Por ejemplo, la prueba podría incluir una hoja con datos sobre la velocidad a la que pueden correr varios animales, en metros por segundo. Luego se hacen preguntas a los estudiantes sobre la distancia que esos animales podrían

6 Las pruebas de los diferentes años se realizan con la misma tabla de especificaciones y utilizan una técnica psicométrica llamada "equiparación" para asegurar que las pruebas de un año a otro utilicen la misma métrica y sean comparables.

7 Si bien el plan de estudios de matemática de segundo de primaria también incluye las áreas de estadística y geometría, se decidió excluirlas para reducir la extensión de la prueba, que es aplicada de manera censal.

recorrer en un tiempo determinado, qué gráfica podría ser coherente con esos datos, entre otras preguntas. Por otra parte, la prueba de comprensión lectora consiste en preguntas de comprensión y reflexión sobre narraciones y textos de diversos temas. Por ejemplo, la prueba podría incluir una narración sobre un empleado de una escuela que recibe un homenaje de sus compañeros de trabajo por haber trabajado diez años allí (Ministerio de Educación, 2015). Luego se hacen preguntas a los estudiantes sobre las emociones del protagonista, los motivos de sus compañeros de trabajo para hacerle el homenaje, el orden de los eventos, etcétera. Los puntajes de las pruebas son computados usando un modelo Rasch, y tienen un índice de confiabilidad superior al 0.80.

Los resultados de aprendizaje de los años 2011-2014 fueron usados en este estudio para estimar las trayectorias previas a la introducción de SP y seleccionar la muestra de escuelas para la evaluación de impacto. Mientras que los resultados de aprendizaje de los años 2009-2015 fueron usados para la especificación de los modelos de impacto.

El Censo Escolar. El Censo Escolar recoge información detallada sobre las características de las escuelas, tales como su ubicación geográfica, instalaciones, acceso a materiales escolares, matrícula, ratio alumnos-docentes, horario escolar, etcétera. El censo se publica entre los meses de abril-mayo de cada año, y se divide en dos partes principales: la información del local escolar y la información sobre la matrícula, docentes y recursos de la escuela.⁸

Estrategia empírica

Modelo en diferencias y modelo Comparative Interrupted Time Series

Para evaluar los efectos de la intervención de Soporte Pedagógico empleamos el método cuasi-experimental denominado *Comparative Interrupted Time Series* (St. Clair, Cook & Hallberg, 2014; Somers, Zhu, Jacob, Bloom, 2013; Bloom, 2001, Bloom 1999). Con este diseño, el efecto del programa se determina al examinar si el grupo de intervención se desvía de su trayectoria base en mayor (igual o menor) medida que el grupo de control. El modelo en diferencias, o *difference in difference* (DiD), es un

⁸ La base de datos original está construida de tal forma que cinco variables especifican de forma única las observaciones. El identificador único de las escuelas son las variables *código modular* y *anexo* y la variable llamada *"nroced"* (número de cédula) indica el nivel educativo de la escuela (primaria, secundaria, etc.). La variable *"cuadro"* indica el tipo de clasificación que se está utilizando para desglosar los datos. Por ejemplo, una pregunta sobre el número de profesores en las escuelas puede estar desglosada por género, por tipo de jornada, etc. Por último, la variable *"tipdato"* indica a cuál de las categorías marcadas por el cuadro se refiere la pregunta, por ejemplo en un cuadro que desglosa por género el tipdato indicaría si la pregunta hace referencia a hombres o a mujeres.

diseño más simplificado (o un caso especial) del CITS dado que el efecto del programa se determina al examinar si el grupo de intervención se desvía de su media base en mayor (igual o menor) medida que el grupo de control. Sin embargo, el modelo CITS controla no solo por las diferencias en medias en varios años, sino también por las diferencias en trayectorias (o pendientes) entre los grupos de intervención y control. Ambos sirven como el contrafactual para evaluar las diferencias en el 2015 en vez de solamente las diferencias de medias (Somers, Zhu, Jacob, & Bloom, 2013). Asimismo, la rigurosidad del modelo en diferencias subyace en el supuesto que el grupo de control provee una estimación válida de los resultados promedio contrafactual para el grupo de tratamiento. Sin embargo, es posible que los grupos de intervención y control tengan “tasas de maduración” (*maturity rates*) distintas y por tanto ese supuesto no se cumpla.⁹ Este supuesto es evaluable con al menos cuatro años de rendimiento previo a la introducción de SP.

El modelo de CITS resuelve las limitaciones del DiD al hacer uso de varios años de rendimiento previos a la introducción del programa (años base). El CITS requiere tener la misma prueba de rendimiento académico para dos grupos y para varios años, antes y después de que se introduzca SP. El efecto de la intervención es estimado a través de la comparación de las diferencias de medias y pendientes de los dos grupos antes de que se inicie la intervención, con las diferencias de medias y pendientes después de la introducción de la misma. Este método se ha encontrado que puede producir estimaciones de impacto muy parecidas a métodos experimentales (St. Clair, Cook & Hallberg, 2014). Sin embargo, no es común tener acceso a varios años de datos históricos (antes de que se introduzca la intervención) y, por ello, el CITS no es un modelo tan popular como el modelo en diferencias. No obstante, esto no representa una limitación para la evaluación de SP dado que estos datos históricos existen en el Ministerio de Educación y estuvieron disponibles para esta evaluación. La Figura 1 del Apéndice A ilustra el modelo CITS.

El modelo CITS permite estimar las trayectorias “base” para cada escuela (o trayectorias para el periodo previo a la introducción de la intervención o 2011-2014); la magnitud por la cual los rendimientos académicos de las escuelas en el año 2015 se desvían de sus trayectorias base; la desviación promedio con respecto a la trayectoria base para los grupos de intervención y de comparación; y, finalmente, el impacto de SP se estima como la diferencia entre la desviación de la trayectoria base en el grupo de intervención, y la desviación de la trayectoria base del grupo de comparación. Si la intervención de SP es efectiva, la desviación

⁹ Maduración o “Maturity” hace referencia a cambios temporales en el tiempo que ocurren naturalmente aun en ausencia de la intervención.

con respecto a la trayectoria base en las escuelas del grupo de intervención será mayor que en el grupo de comparación.¹⁰

En particular, la ecuación estimada para el CITS para la escuela j en el año t es la siguiente:

$$Y_{jt} = \beta_0 + \beta_1 \text{Soporte}_j + \beta_2 \text{Tendencia}_j + \beta_3 \text{Soporte}_j \text{X} \text{Tendencia}_j + \beta_4 I(2015)_j + \beta_5 \text{Soporte}_j \text{X} I(2015)_j + \varepsilon_{jt} \quad (1)$$

Donde:

- Soporte es una variable dicotómica que toma el valor de 1 si la escuela recibe el programa de Soporte Pedagógico, y 0 en caso contrario.
- Tendencia: tendencia lineal antes de la implementación del programa centrada en el año 2014 (Tendencia = Año - 2014). La tendencia lineal fue estimada con datos de 2009 a 2014.
- Soporte X Tendencia: es la interacción entre la variable de Soporte y la tendencia lineal.
- I(2015): es una variable dicotómica que toma el valor de 1 para el año 2015, y el valor de 0 para los años previos.
- Soporte X I(2015): es la interacción entre la variable de Soporte y la variable dicotómica de 2015.

Por tanto:

- β_0 = media de Y para el grupo de comparación en el año 2014.
- β_1 = diferencia de medias entre los grupos de Soporte y de comparación para el año 2014.
- $\beta_0 + \beta_1$ = media de Y para el grupo de Soporte en el año 2014.
- β_2 = pendiente (Tendencia) de Y para el grupo de comparación en el periodo previo a la introducción del programa.
- β_3 = diferencia en las pendientes (Tendencia) entre el grupo de Soporte y el grupo de comparación en el periodo previo a la introducción del programa. Este

¹⁰ Cabe resaltar que este modelo asume que los datos de rendimiento previos a la introducción de SP representan el mejor predictor del rendimiento futuro en ausencia de cambios sistemáticos en el tiempo. Asimismo, el modelo asume que varios años de rendimiento académico predicen mejor el promedio de la IE futuro que una sola observación. Sin embargo, diferentes tipos de trayectorias antes de la introducción de SP pueden generar diferentes proyecciones, por lo que es muy importante determinar de manera cuidadosa la trayectoria base antes de que se inicie SP en 2015.

coeficiente captura las diferencias entre las tendencias del grupo de intervención y de comparación. Lo ideal es que este coeficiente sea igual a cero porque indicaría que las tendencias son paralelas antes de 2015.

- $\beta_2 + \beta_3$ = pendiente (Tendencia) de Y para el grupo de Soporte en el periodo previo a la introducción del programa.
- β_4 = desviación en el año 2015 con respecto a la tendencia del grupo de comparación.
- β_5 = diferencia entre el grupo de Soporte y el grupo de comparación en las desviaciones con respecto a la tendencia en 2015. Este es el coeficiente que captura el impacto del programa en el año 2015.
- $\beta_4 + \beta_5$ = desviación en el año 2015 con respecto a la tendencia del grupo de Soporte.

Por su parte, la ecuación estimada para el DiD para la escuela j en el año t es la siguiente:

$$Y_{jt} = \delta_0 + \delta_1 \text{Soporte}_j + \delta_4 I(2015)_j + \delta_5 \text{Soporte}_j I(2015)_j + \zeta_{jt} \quad (2)$$

Es decir, el modelo de DiD es estimado a partir de tres de las cinco variables del CITS.¹¹ La variable “Soporte” toma el valor de 1 si la escuela recibe el programa de Soporte Pedagógico, y 0 si la escuela es del grupo de comparación. Esta variable mide la diferencia en rendimiento académico entre el grupo de intervención y de comparación en los años anteriores (2013-2014) a la intervención. Es decir, esta variable permite establecer qué tan diferente era el rendimiento académico de los estudiantes de ambos grupos. Segundo, empleamos una variable dicotómica de tiempo “I(2015)” que es igual a 1 si las observaciones son posteriores a la intervención (i.e., las observaciones son del año 2015) y 0 en caso contrario. Esta variable mide el cambio en el tiempo en el rendimiento académico de los estudiantes del grupo de comparación. Si esta variable es positiva es porque la cohorte de estudiantes de segundo de primaria en 2015 del grupo de comparación tuvo un rendimiento mayor en la prueba de la ECE con relación a los estudiantes de segundo de primaria de las cohortes anteriores a 2015. Por último, el coeficiente de la interacción de las dos variables anteriores “Soporte X I(2015)” corresponde al estimador de DiD. Esta última variable captura el efecto causal de la intervención de SP en la medida en que estima el cambio en el tiempo en el rendimiento académico de los estudiantes intervenidos, tras ajustar dicho cambio por lo que hubiera sido en caso de que el programa no se hubiera implementado, es decir, por el cambio en el tiempo en el rendimiento académico de los estudiantes del grupo de comparación.

¹¹ Esto permite ver por qué el modelo de DiD es considerado un caso especial de CITS.

Sin embargo, los impactos de la intervención de SP estimados a partir de la estrategia de DiD solo corresponden al efecto causal del programa siempre y cuando las trayectorias de rendimiento académico de los grupos de intervención y de comparación en el periodo previo a la introducción del programa hayan tenido un comportamiento similar a lo largo del tiempo. Este supuesto se conoce en la literatura de evaluación de impacto como el supuesto de tendencias paralelas. En general, este supuesto no puede ser testado empíricamente y representa una de las mayores limitaciones de los modelos de DiD. No obstante, en la medida en que para esta evaluación de impacto contamos con información histórica de rendimiento académico a partir de 2009, podemos construir un grupo de comparación que presente una trayectoria de rendimiento académico para el periodo anterior a la intervención que resulte muy similar a la trayectoria de rendimiento del grupo de escuelas que recibió el programa de SP. De esta forma, se puede garantizar que el supuesto de tendencias paralelas se cumple por construcción y que, por tanto, los efectos estimados pueden ser interpretados como el efecto causal del programa de SP en el rendimiento académico de los estudiantes.

La ventaja del modelo de CITS sobre el modelo de DiD estándar es que el primero permite validar y controlar empíricamente el supuesto de tendencias paralelas. Específicamente, el modelo CITS permite estimar dos variables adicionales de interés, las cuales no son consideradas en el modelo de DiD. Primero, el coeficiente de la variable "Tendencia" mide la pendiente en el periodo comprendido entre 2009 y 2014 de la variable de resultado de rendimiento (e.g., comprensión lectora) para el grupo de comparación. Por su parte, el coeficiente de la interacción entre la variable dicotómica de Soporte y la tendencia lineal "Soporte X Tendencia" mide si existe alguna diferencia entre las pendientes de las trayectorias de rendimiento de los grupos de intervención y de comparación. Esta última variable nos permite testear empíricamente el supuesto de tendencias paralelas. Es decir, si el coeficiente estimado para esta variable no es estadísticamente diferente de cero significa que las pendientes del grupo de comparación y de intervención son iguales.

Para reducir la posibilidad de que existan diferencias entre los grupos de intervención y de comparación que puedan sesgar las estimaciones de impacto de SP, combinamos la estrategia de CITS y de DiD con la estrategia de emparejamiento denominada *Propensity Score Matching*. Es decir la variable dicotómica "Soporte" en las ecuaciones 1 y 2 solo incluye escuelas emparejadas. La combinación de la técnica de emparejamiento con los modelos descritos permite incrementar la comparabilidad de los dos grupos en términos del rendimiento académico histórico y de otras variables observables relevantes antes de la introducción de SP. Pero, asimismo, permite controlar por características no observables que no cambian en el tiempo dado que las

mismas escuelas son analizadas a través del tiempo, y permite reducir la dependencia de controlar estadísticamente por diferencias en el modelo. En este diseño cuasi-experimental, la disponibilidad de varios años de resultados de aprendizaje antes de la introducción de SP tiene el potencial de reducir sesgo a través de la mejora de la calidad de los pares emparejados o de modelar las diferencias en las trayectorias previas. La estrategia del método de emparejamiento se detalla más adelante en este capítulo.

Es importante anotar que el modelo de CITS no requiere que las tendencias de rendimiento de los grupos de intervención y de comparación tengan que ser paralelas en el tiempo a fin de estimar el efecto de un programa. Sin embargo, en la medida en que en este estudio combinamos los modelos de CITS y de DiD con la estrategia de emparejamiento, cabe esperar las tendencias del grupo de intervención y comparación resulten paralelas por construcción y que, por tanto, los efectos estimados del programa tanto por el modelo de CITS como por el de DiD sean muy parecidos. En todo caso, como se mencionó anteriormente, la gran ventaja de la especificación del modelo de CITS sobre el de DiD es que el primero permite validar empíricamente si el supuesto de tendencias paralelas se cumple.

Todas las especificaciones de los modelos CITS y DiD que se presentan en el Capítulo 6 se estimaron al nivel de la escuela en donde cada observación para una escuela en un año dado está ponderada por el número de estudiantes de la escuela en ese año. Esta estrategia es equivalente a estimar regresiones a nivel del estudiante cuando la variable de tratamiento está definida a nivel de la escuela, como ocurre con el programa de SP. Asimismo, los errores estándar de todas la regresiones están agrupados (*clustered*) a nivel de la escuela con el fin de tener en cuenta la correlación existente en el tiempo de la variable de resultado de una misma escuela.

Propensity Score Matching de escuelas

Con el fin de mejorar las estimaciones de los modelos propuestos, es decir, reducir sesgo en las estimaciones, usamos una combinación de estrategias no paramétricas para seleccionar las escuelas más similares del grupo de intervención y de comparación y por tanto incrementar la comparabilidad de los dos grupos. Es importante recordar que dado que SP se ofrece a toda la escuela, la unidad de análisis es la escuela.

Primero procedimos a restringir el universo de potenciales controles imitando los criterios de focalización de la intervención de SP, es decir, seleccionando solo las escuelas que cumplieran con la característica de ser públicas, urbanas y polidocentes completas dentro de este universo. Esta estrategia permitió hacer una especie de

emparejamiento “exacto” en variables claves de focalización del programa que están altamente correlacionadas con el rendimiento académico de los estudiantes. Pero, a su vez, ayudó a incrementar la comparabilidad de los dos grupos en términos de otras variables relacionadas con las variables de focalización y rendimiento académico que no eran directamente observables en este estudio. Por ejemplo, al excluir del universo de potenciales controles a las escuelas rurales y polidocentes multigrado, se eliminaron de la muestra escuelas que albergan a los estudiantes más pobres y con experiencias escolares y características familiares muy distintas a los estudiantes que atienden las escuelas urbanas y polidocentes completas.

El proceso de emparejamiento exacto en función de las variables de focalización redujo el universo de potenciales controles a 2,609 escuelas que cumplían con la característica de ser públicas, urbanas, polidocentes completas y que contaban a su vez con datos de rendimiento en 2015.¹² Mientras que el grupo de escuelas de intervención que cumplían con esas condiciones se redujo a 2,923. Dado que el ratio de escuelas control a escuelas de intervención es menor a 1, es claro que no existe un conjunto muy grande de potenciales controles y, por tanto, anticipa que no será posible emparejar a todas las escuelas intervenidas.

Segundo, con la muestra de potenciales controles que quedaron después de implementar el emparejamiento exacto, procedimos a seleccionar las escuelas más parecidas entre sí usando un escalar o *propensity score* que resume una serie de características relevantes de las escuelas y de los estudiantes que atienden dichas escuelas (Rosenbaum y Rubin, 1983; Stuart, 2007). Dicho escalar fue usado para emparejar las escuelas de ambos grupos dentro de una distancia máxima determinado por un *caliper*.

Finalmente, para incrementar la comparabilidad de las escuelas emparejadas no solo en términos de las variables de focalización y del *propensity score*, sino también en términos de variables correlacionadas con la ubicación geográfica de las escuelas, emparejamos las escuelas dentro de cada estrato geográfico. De esta manera todos los pares de escuelas emparejadas obtuvieron el mismo valor en la variable “dominio geográfico”.¹³ Los pasos dos y tres de esta estrategia de emparejamiento son explicados a continuación.

12 Se usaron las variables de las bases de la Evaluación Censal Escolar para definir este subgrupo de escuelas. Por lo tanto, este subgrupo corresponde al número de escuelas que cumplían con los tres criterios de focalización y que además se encontraban en la base panel conformada con las bases históricas de la ECE.

13 La variable Dominio proviene de la Encuesta Nacional de Hogares (ENAH) de 2014.

Especificación del modelo de selección

Para estimar el escalar o *propensity score* especificamos un modelo de selección o modelo de *propensity score* (Rosenbaum y Rubin, 1983; Stuart, 2007). Este modelo estima la probabilidad de recibir el programa de Soporte Pedagógico en función de un conjunto de variables observables. La probabilidad estimada tiene un valor en el rango de 0 a 1, donde las escuelas con puntaje cercano a 1 son más propensas a pertenecer al grupo SP que las escuelas con puntaje cercano a 0. En particular, para la especificación del modelo de selección es muy importante incluir variables asociadas con la probabilidad de pertenecer al grupo de intervención y con la variable de resultado de interés, o rendimiento académico de los estudiantes (Stuart, 2009). Dado que el programa tomó en cuenta variables de rendimiento académico para la focalización del programa (como se explicó en el Capítulo 2), resulta relevante incluir variables de rendimiento académico en el modelo de selección. Asimismo, dado que el puntaje en las pruebas de rendimiento es la variable de interés en las regresiones de impacto, la inclusión de datos de rendimiento académico previos a la implementación del programa es clave para la especificación del modelo de selección y para juzgar la comparabilidad de los dos grupos. Adicionalmente, el rendimiento académico de los estudiantes suele estar correlacionado con otras características relevantes del estudiante como las características de los padres, el nivel socioeconómico de la familia, y el entorno en el que viven, y con factores de la escuela y las características de los docentes (Rivkin, Hanushek, & Kain, 2005). Dado que la mayoría de estos factores asociados al rendimiento no son observables en este estudio, el uso de varios años de rendimiento académico histórico ayuda a incrementar la comparabilidad de los dos grupos. Finalmente, el modelo de selección incorpora una serie de variables relevantes que caracterizan a la escuela, los estudiantes que atienden la escuela, y el área en la que se encuentra la institución educativa. En resumen, se especificaron dos modelos de selección con las siguientes características:

Muestra analítica o Muestra A: esta muestra asume que el programa inició las tareas de implementación en el año 2015 y por tanto usa variables previas a 2015, como datos de rendimiento de 2014, para el modelo selección. Es decir, en términos de las variables de rendimiento el modelo incluye los resultados de la ECE de los años 2011, 2012, 2013 y 2014 para matemática y comprensión, así como la proporción de estudiantes en el nivel de logro “satisfactorio” en matemática y comprensión del año 2014, y medidas de dispersión de los resultados en las pruebas de matemática y comprensión lectora en 2014 (o nivel de heterogeneidad de los resultados). Esta especificación resulta importante dado que 2014 es un año en que los rendimientos académicos de los estudiantes de segundo grado de primaria mejoraron notablemente

en el país. Por tanto, incluir los datos de 2014 en la especificación del modelo de selección permite incrementar comparabilidad de los grupos en términos de estos cambios en el tiempo. Asimismo, esta muestra se construye a partir de la inclusión de variables tales como el tamaño de matrícula (de 2013 y 2014), la tasa de aprobación de estudiantes (de 2013 y 2014), el número de docentes en la escuela, el porcentaje de estudiantes indígenas, descripción si la escuela es EIB o no, indicadores sobre la jornada pedagógica, el ratio de alumnos-docentes, la disponibilidad de textos y cuadernos escolares, computadoras e internet. Además, se incluyeron variables sobre la infraestructura del centro tales como características de las paredes (ladrillo), pisos (no tierra) e inodoros (en buenas condiciones), disponibilidad de agua potable, desagüe y electricidad.¹⁴ Finalmente, se incluyeron indicadores distritales que capturan el nivel de pobreza y desnutrición del distrito.¹⁵ Los resultados de este modelo se presentan en el resto del documento.

Muestra para análisis de robustez o Muestra B: incluye variables similares a las de la muestra A, pero asume que el programa inició las tareas de implementación en el año 2014 y por tanto usa variables previas a 2014 para el modelo de selección. Específicamente, esta muestra incluye el valor promedio al nivel de la escuela de los resultados académicos en las áreas de matemática y comprensión lectora para los años 2010, 2011, 2012 y 2013; indicadores que capturan la proporción de estudiantes en los niveles de logro “satisfactorio” e “inicio” en los años 2012 y 2013 para matemática y comprensión, e indicadores que capturan el nivel de dispersión de los resultados académicos de matemática y comprensión lectora en 2013. Asimismo, esta muestra se construye a partir de la inclusión de variables tales como el tamaño de matrícula, la tasa de aprobación de estudiantes, el número de docentes en la escuela, el porcentaje de estudiantes indígenas, descripción si la escuela es EIB o no, indicadores sobre la jornada pedagógica, el ratio de alumnos-docentes, la disponibilidad de textos y cuadernos escolares, computadoras e internet. Además, se incluyeron variables sobre la infraestructura del centro tales como características de las paredes, pisos e inodoros, disponibilidad de agua potable, desagüe y electricidad.¹⁶ Finalmente, se incluyeron indicadores distritales que capturan el nivel de pobreza y desnutrición del distrito.¹⁷ Los resultados del ejercicio de emparejamiento usando esta muestra se presentan en el Apéndice B, mientras que los resultados de impacto usando esta muestra se discuten como análisis de robustez en el Capítulo 6.

14 Estas últimas variables corresponden al año 2014.

15 Estas variables corresponden al año 2009.

16 Estas últimas variables corresponden al año 2013.

17 Estas variables corresponden al año 2009.

El escalar se calcula de forma conjunta para todas las regiones mediante un modelo *Logit*.¹⁸ En particular, la ecuación estimada para el *propensity score* está dada por:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = X_i\beta + Z_d\gamma \quad (3)$$

donde:

- π_{id} es igual a 1 si la escuela i pertenece al grupo de SP y 0 si pertenece al grupo de comparación.
- X_i es un vector de variables de la escuela.
- Z_d es un vector de variables de rendimiento académico histórico para las áreas de matemática y comprensión lectora y para el periodo de 2011-2014.

Emparejamiento de las escuelas

El escalar o *propensity score* estimado a través del modelo de selección se usó para emparejar y seleccionar al subgrupo de escuelas más parecidas entre sí de los grupos de intervención y de comparación. Para seleccionar y emparejar a las escuelas se usó el algoritmo denominado *Nearest Neighbor Matching*. Asimismo, para minimizar la distancia de las escuelas emparejadas y por tanto incrementar la comparabilidad de los pares, se determinó un *caliper* o distancia máxima entre dos escuelas para que se emparejen. Esta estrategia es también conocida como “caliper matching” y es considerada una variante de *Nearest Neighbor Matching* (Smith y Todd, 2003). El *caliper* es definido como el producto de un multiplicador igual a 0.01 y la desviación estándar del *propensity score* estimado.¹⁹ Rosenbaum and Rubin (1985) sugieren un *caliper* de 0.25 desviaciones estándar del *propensity score*, mientras que Cochran y Rubin (1973) recomiendan un *caliper* de 0.2 para reducir 98 por ciento del sesgo.²⁰ Si bien la introducción del caliper restringe el número de escuelas intervenidas que se pueden emparejar, reduce la distancia entre los pares emparejados y por tanto incrementa la comparabilidad de los dos grupos.²¹ Asimismo, “caliper matching” automáticamente usa escuelas dentro y cerca del área de *common support* o área de intersección de las distribuciones de los *propensity scores* en los grupos de intervención y de comparación.

18 Los resultados de la estimación de la probabilidad esperada usando un modelo *logit* son muy similares a los obtenidos al emplear un modelo *probit*. No se emplea el uso de un modelo de probabilidad lineal ya que algunas observaciones pueden obtener probabilidades por fuera del intervalo [0,1].

19 Sin embargo, mientras más pequeño es el *caliper* menor es la probabilidad de emparejar a una escuela.

20 Este *caliper* es recomendado en casos donde la varianza del *propensity score* en el grupo de tratamiento es cerca del doble de la varianza del grupo de control.

21 *Caliper matching* automáticamente usa escuelas dentro y cerca del área de *common support* o área de intersección de las distribuciones de los *propensity scores* en los grupos de intervención y de comparación.

Esto es recomendado en la literatura de *matching* dado que ayuda a reducir sesgo de selección (Heckman, et al, 1997).

En cuanto a la estructura del emparejamiento, esta se determinó como 1-1 y sin reemplazo, es decir, para cada escuela intervenida por SP se seleccionó una escuela de comparación. Otro tipo de estructura en la cual se seleccionan más de un control por escuela intervenida (1-M) comúnmente incrementa sesgo de selección dado que el segundo, tercer o cuarto control tienen por definición una distancia más grande que el primer control emparejado. De esta manera las escuelas comparación son seleccionadas una vez y no se permite que sirvan de control para otras escuelas de SP.²²

Estratificación

Asimismo, para incrementar la comparabilidad de los pares emparejados en términos de características geográficas se emparejaron los casos dentro de estratos geográficos determinados por la variable de dominio geográfico. Esta estrategia permite asegurar la comparabilidad de los pares en torno a variables que comúnmente están asociadas con la ubicación geográfica de la escuela y que no son fácilmente observables con datos secundarios. La variable dominio geográfico permite agrupar y clasificar a las escuelas en ocho (08) estratos geográficos tales como: costa norte, costa centro, costa sur, sierra norte, sierra centro, sierra sur, selva y Lima Metropolitana. De esta manera, todas las escuelas emparejadas tienen un valor exacto en dominio geográfico. Si bien es posible que dentro de estos estratos geográficos todavía existan diferencias importantes entre las escuelas, la combinación de la estratificación con las otras estrategias de emparejamiento ayudó a incrementar la comparabilidad de las escuelas. Asimismo, no fue posible usar estratos más pequeños tales como región, dado que en muchos casos la cantidad de escuelas no intervenidas por el programa era limitada. Pero también porque dado el ratio tan bajo de controles a tratamiento, un mayor nivel de estratificación hubiese reducido altamente el número de escuelas emparejadas.

22 Un beneficio de usar la estructura de emparejamiento 1-1 es que los estimados son más precisos. Sin embargo, un costo es que la muestra de escuelas comparación disponible para emparejar se va reduciendo conforme se emparejan las escuelas. Para reducir la posibilidad de que las escuelas emparejadas al final tengan pares con mayor distancia (y por tanto sean muy diferentes de las escuelas del grupo de intervención), usamos un *caliper* o distancia máxima de emparejamiento.

Resultados del proceso de emparejamiento

Para la muestra analítica o Muestra A se lograron emparejar 1,098 escuelas con Soporte Pedagógico. Esta muestra equivale aproximadamente al 36.5 por ciento de las escuelas de SP, estatales, urbanas y polidocentes completas (N=3,006), y aproximadamente al 38 por ciento de las escuelas de SP que, además de cumplir con las tres primeras condiciones, cumplían con la condición de tener resultados de la ECE para los años 2013, 2014 y 2015 (N=2,923). La Tabla 1 del Apéndice B presenta el porcentaje de escuelas emparejadas por región para diferentes muestras: la muestra original del programa, la muestra de escuelas con SP que incluye solo polidocentes completas y urbanas, y la muestra que cumple con las condiciones anteriores además de las condiciones de ser estatales y con información de rendimiento académico. En promedio se emparejaron más del 45 por ciento de los centros asignados a SP en 10 de las 15 regiones. En Ucayali se emparejaron alrededor de 36 por ciento, mientras que en las regiones de La Libertad, Lambayeque y Piura, ubicadas en la costa norte del Perú, se emparejaron menos del 30 por ciento de las escuelas con SP. Lima Metropolitana y provincia fue la región con menos emparejados en donde aproximadamente un 17 por ciento de las escuelas con SP fueron emparejadas. Asimismo, la Tabla 2 del Apéndice B presenta el porcentaje de escuelas emparejadas para cada dominio geográfico. Para los dominios geográficos de la costa centro, costa sur, sierra centro, sierra sur y la selva, se emparejaron 50 por ciento o más de las escuelas de SP. No obstante, para la región costa norte y Lima metropolitana²³, estos porcentajes son similares a 25 y 13 por ciento, respectivamente.

Las siguientes Figuras 2 y 3 presentan las trayectorias de rendimiento “base” (o previas a la introducción del programa de SP) para las muestras emparejadas del grupo de intervención y de comparación. En estas figuras se puede observar que el conjunto de estrategias descritas logró crear dos grupos casi idénticos en términos de sus trayectorias y los promedios de rendimiento previos a la introducción del programa de SP. En este caso, las mejoras observadas en el grupo de intervención en el año 2015 son más creíbles que se expliquen por la introducción de SP dado que durante los años previos a la implementación del programa los dos grupos presentan promedios y tasas de crecimiento muy similares.

²³ En el indicador de dominio geográfico, Lima metropolitana incluye la provincia del Callao pero no Lima Provincia.

Figura 2.
Trayectoria de la prueba de comprensión lectora antes de introducir la intervención de Soporte Pedagógico para las muestras emparejadas del grupo de intervención y de comparación

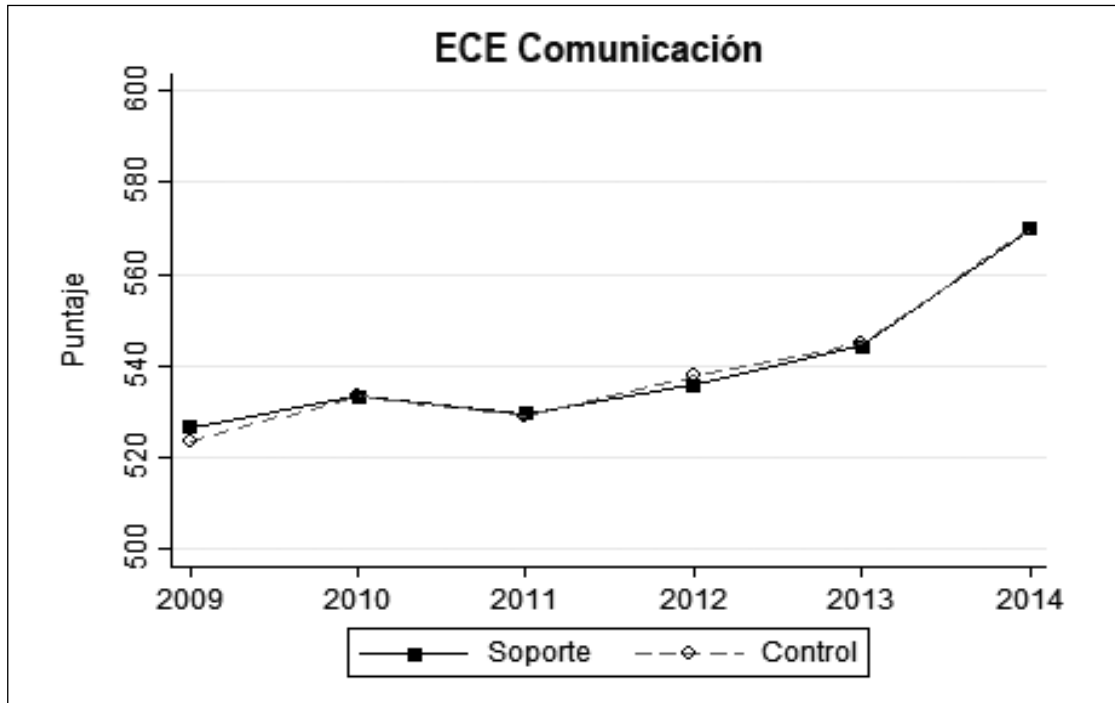
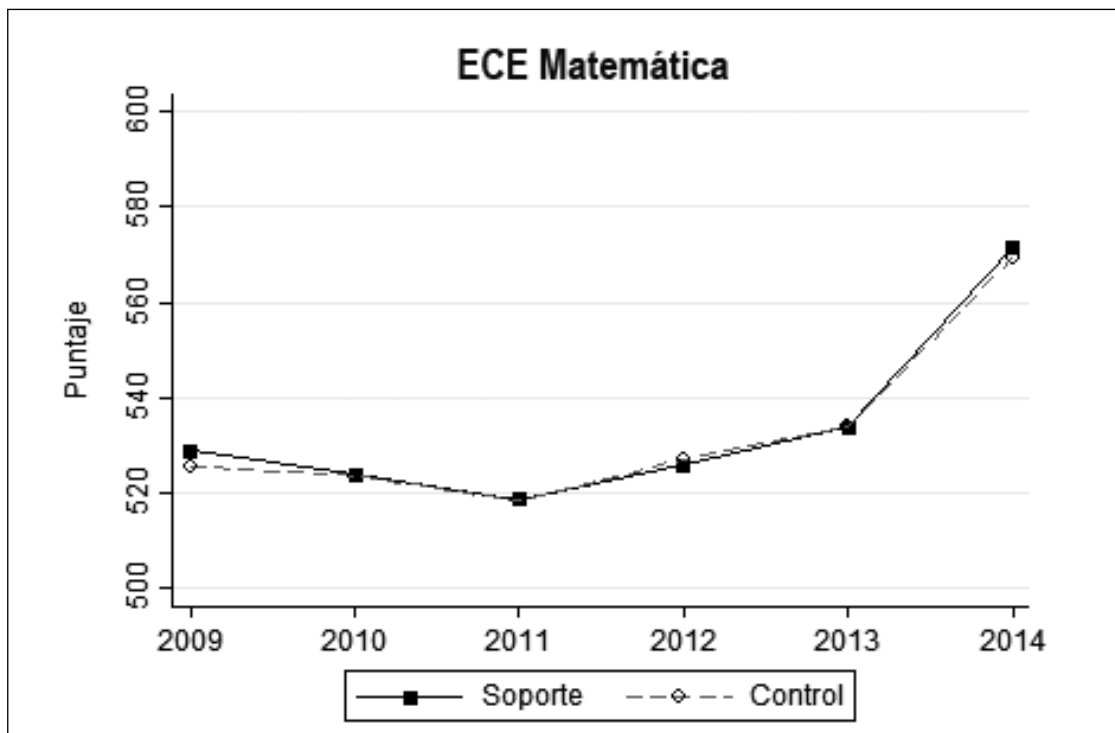


Figura 3.
Trayectoria de la prueba de matemática antes de introducir la intervención de Soporte Pedagógico para las muestras emparejadas del grupo de intervención y de comparación



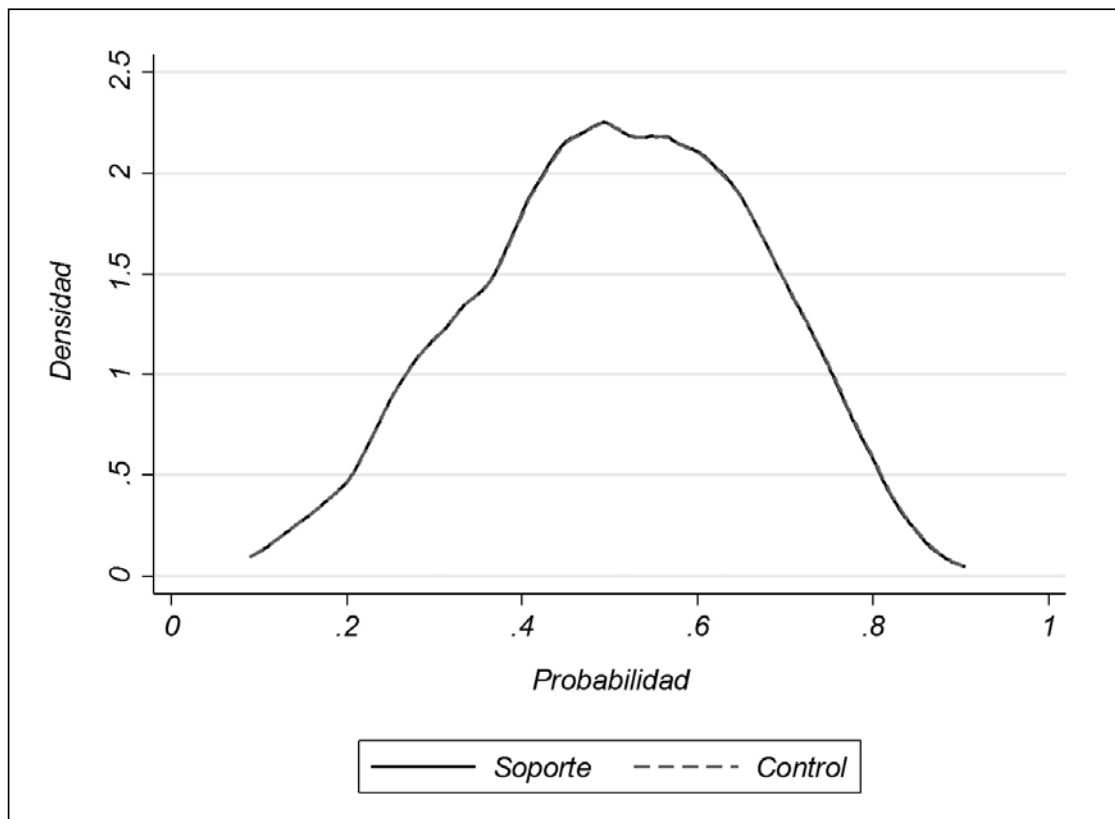
Asimismo, la Tabla 3 presenta la media y la desviación estándar del escalar o valor estimado del *propensity scores* antes y después de emparejar las escuelas. Como es de esperarse, antes de emparejar las escuelas el valor medio del *propensity score* es más alto para las escuelas intervenidas que para las escuelas no intervenidas (0.59 versus 0.46). Sin embargo, después de aplicar el método emparejamiento el valor medio del *propensity score* es idéntico en los dos grupos. Asimismo, el ratio de las varianzas del *propensity score* en los dos grupos es casi igual a 1, como lo recomienda la literatura de *matching* para generar estimaciones de impacto confiables (Rubin, 2001). Estos resultados también se pueden apreciar en la Figura 4, la cual presenta distribuciones idénticas del *propensity score* para las muestras emparejadas del grupo de intervención y comparación.

Tabla 3.
Descriptivos Estadísticos del *Propensity Score* antes y después de aplicar el método de emparejamiento

| | Media | Desviación estandar | N |
|--|-------|------------------------|-------|
| Antes de aplicar el método de emparejamiento: | | | |
| Escuelas sin Soporte Pedagógico | 0.460 | 0.174 | 2,423 |
| Escuelas con Soporte Pedagógico | 0.592 | 0.165 | 2,732 |
| Total muestra con propensity scores | 5,155 | | |
| Después de aplicar el método de emparejamiento: | | | |
| Escuelas sin Soporte Pedagógico | 0.509 | 0.160 | 1,098 |
| Escuelas con Soporte Pedagógico | 0.509 | 0.159 | 1,098 |
| Total muestra emparejada | 2,196 | | |

Fuente. Elaboración propia.

Figure 4.
Distribución del *Propensity Score* para las muestras emparejadas del grupo de intervención y comparación



La Tabla 4 presenta descriptivos estadísticos para las variables de logro de aprendizaje de los alumnos de segundo grado de primaria en las áreas de comprensión lectora y matemática para el periodo 2011-2014. Específicamente, la Tabla 4 presenta el valor medio, el número de escuelas, la diferencia de medias entre las escuelas emparejadas del grupo de intervención y de comparación, el error estándar de la diferencia, el valor p , y el tamaño del efecto (o diferencia de medias estandarizadas). En general los grupos de intervención y de comparación son muy parecidos después del emparejamiento y que en casi todos los casos la diferencia de medias estandarizada es menor a 0.25, como lo recomienda la literatura de *matching*. La Tabla 3 del Apéndice B presenta estadísticos similares para un conjunto extenso de variables de la escuela, y nuevamente confirma la comparabilidad de los dos grupos.

Tabla 4.
Descriptivos estadísticos de las variables de rendimiento para las muestras emparejadas del grupo de intervención y comparación

| Variables | Escuelas | | | | Diferencia | Error estándar | valor-p | Tamaño del Efecto |
|--|------------------------|-------|------------------------|-------|------------|----------------|---------|-------------------|
| | Sin Soporte Pedagógico | | Con Soporte Pedagógico | | | | | |
| | Media | N1 | Media | N2 | | | | |
| Comprensión 2014 | 561.39 | 1,098 | 559.87 | 1,098 | -1.52 | 2.26 | 0.50 | -0.03 |
| Matemática 2014 | 558.87 | 1,098 | 558.57 | 1,098 | -0.29 | 3.46 | 0.93 | -0.00 |
| Comprensión 2013 | 537.13 | 1,098 | 535.92 | 1,098 | -1.21 | 2.16 | 0.57 | -0.02 |
| Matemática 2013 | 526.78 | 1,098 | 524.98 | 1,098 | -1.81 | 2.93 | 0.54 | -0.03 |
| Comprensión 2012 | 528.79 | 1,098 | 525.82 | 1,098 | -2.97 | 2.15 | 0.17 | -0.06 |
| Matemática 2012 | 519.46 | 1,098 | 516.11 | 1,098 | -3.34 | 2.80 | 0.23 | -0.05 |
| Comprensión 2011 | 520.94 | 1,062 | 520.49 | 1,018 | -0.46 | 2.20 | 0.84 | -0.01 |
| Matemática 2011 | 511.65 | 1,062 | 510.28 | 1,018 | -1.38 | 2.90 | 0.64 | -0.02 |
| % logro Inicio Comprensión 2014 | 0.11 | 1,098 | 0.10 | 1,098 | -0.00 | 0.01 | 0.67 | -0.02 |
| % logro Inicio Mate 2014 | 0.39 | 1,098 | 0.38 | 1,098 | -0.00 | 0.01 | 0.96 | -0.00 |
| % logro Satisfactorio Comprensión 2014 | 0.38 | 1,098 | 0.37 | 1,098 | -0.01 | 0.01 | 0.55 | -0.03 |
| % logro Satisfactorio Mate 2014 | 0.26 | 1,098 | 0.26 | 1,098 | -0.00 | 0.01 | 0.90 | -0.01 |
| % logro Inicio Comprensión 2013 | 0.16 | 1,098 | 0.17 | 1,098 | 0.01 | 0.01 | 0.46 | 0.03 |
| % logro Inicio Mate 2013 | 0.49 | 1,098 | 0.50 | 1,098 | 0.01 | 0.01 | 0.61 | 0.02 |
| % logro Satisfactorio Comprensión 2013 | 0.27 | 1,098 | 0.27 | 1,098 | -0.00 | 0.01 | 0.95 | -0.00 |
| % logro Satisfactorio Mate 2013 | 0.16 | 1,098 | 0.16 | 1,098 | 0.00 | 0.01 | 0.93 | 0.00 |
| % logro Inicio Comprensión 2012 | 0.20 | 1,098 | 0.22 | 1,098 | 0.01 | 0.01 | 0.13 | 0.07 |
| % logro Inicio Mate 2012 | 0.51 | 1,098 | 0.52 | 1,098 | 0.02 | 0.01 | 0.13 | 0.06 |
| % logro Satisfactorio Comprensión 2012 | 0.24 | 1,098 | 0.23 | 1,098 | -0.01 | 0.01 | 0.30 | -0.04 |
| % logro Satisfactorio Mate 2012 | 0.12 | 1,098 | 0.12 | 1,098 | -0.00 | 0.01 | 0.94 | -0.00 |
| Desviación Estándar Comprensión 2014 | 64.88 | 1,098 | 65.74 | 1,098 | 0.86 | 0.71 | 0.23 | 0.05 |
| Desviación Estándar Matemática 2014 | 93.85 | 1,098 | 94.78 | 1,098 | 0.93 | 1.05 | 0.38 | 0.04 |
| Desviación Estándar Comprensión 2013 | 65.37 | 1,098 | 65.48 | 1,098 | 0.11 | 0.66 | 0.86 | 0.01 |
| Desviación Estándar Matemática 2013 | 82.24 | 1,098 | 81.13 | 1,098 | -1.12 | 0.92 | 0.23 | -0.05 |

Fuente: Elaboración propia.

Asimismo, el Apéndice B presenta otros resultados de diagnóstico que validan la comparabilidad de los grupos de intervención y comparación formados a partir del método de emparejamiento. Estos resultados incluyen la Tabla 3 que presenta los resultados de balance (el valor promedio de diversas variables, el número de escuelas, la diferencia de medias entre las escuelas emparejadas del grupo de intervención y de comparación, el error estándar de la diferencia, el valor p , y el tamaño del efecto). Adicionalmente, la Figura 2 de dicho apéndice presenta los gráficos de *boxplots* del *propensity score* antes y después del emparejamiento para los grupos de intervención y de comparación. Los resultados de diagnóstico para la muestra creada para los análisis de robustez o Muestra B se presentan en el Apéndice B, en las Figuras 3, 4, 5 y la Tabla 5.

6. RESULTADOS

En este capítulo presentamos los resultados del impacto del programa de SP en el rendimiento académico de los estudiantes de segundo grado de primaria de las escuelas públicas, polidocentes urbanas. Para medir rendimiento académico usamos dos tipos de variables de resultado. En primer lugar, estimamos los efectos de SP en los resultados promedio de las escuelas, tanto en comprensión lectora como en matemática usando la información de la ECE de 2009 a 2015. En segundo lugar, utilizamos el porcentaje de estudiantes que conforman cada uno de las tres niveles de logro (inicial, en proceso y satisfactorio) en las pruebas de la ECE. Para la definición de estas variables también se empleó información del periodo de 2009 a 2015.

Los resultados del impacto muestran que SP generó una mejora modesta pero promisoría en el rendimiento académico de los estudiantes de las escuelas intervenidas. En promedio, los resultados muestran que los estudiantes que recibieron la intervención de SP presentan resultados superiores en las pruebas de la ECE de aproximadamente 0.1 desviaciones estándar en comprensión lectora con relación a los estudiantes del grupo de comparación y de 0.07 desviaciones estándar en Matemática. Por su parte, los resultados en las variables de logro muestran que los estudiantes de SP resultan 3 puntos porcentuales menos probables de pertenecer al grupo de nivel inicial de logro tanto en comunicación como en matemática. Los resultados de impacto son robustos a diferentes tipos de especificaciones, variables de control y tipo de muestras empleadas. Asimismo, los resultados no son heterogéneos al grado de implementación de la intervención.

Impacto en comprensión lectora y matemática

En las Tablas 5 y 6 se presentan los resultados de impacto de SP en comprensión lectora y matemática, respectivamente, usando diferentes metodologías y variables de control para la Muestra A (descrita en el Capítulo 5). La columna (1) muestra los resultados para el modelo DiD que emplea información de las ECE de 2013 a 2015. Para este modelo se asume que los efectos del programa de SP solo pueden ser observados a partir de 2015.²⁴

Los resultados del modelo de DiD en la Tabla 5 para la variable Soporte indican que la diferencia entre el grupo de intervención y de comparación antes de la intervención

²⁴ Este supuesto se investiga con mayor detalle más adelante en este capítulo.

es solo de 0.28 puntos de la prueba de comprensión lectora en favor del grupo de comparación y no es estadísticamente significativa. Este resultado no es sorprendente en la medida en que ello es precisamente lo que se quería lograr a partir del proceso de emparejamiento estadístico: que los estudiantes del grupo de intervención y de comparación fueran lo más parecido posibles en términos de su rendimiento académico antes de la introducción del programa. Por su parte, los resultados de la variable “I(2015)” indican que los estudiantes de segundo de primaria del grupo de comparación que tomaron la ECE en 2015 obtuvieron en promedio 36 puntos adicionales en la prueba de comprensión lectora más que los estudiantes del grupo de comparación de los años anteriores a la intervención. Por último, la interacción de estas dos variables, “Soporte X I(2015)”, indica que el efecto causal del programa de SP en comprensión lectora es de 5.6 puntos en la prueba, resultado que es estadísticamente significativo y que equivale a una ganancia de aproximadamente 0.08 desviaciones estándar.

Tabla 5.
Resultados de impacto de la intervención de Soporte Pedagógico en comprensión lectora

| | DiD | | CITS | |
|-----------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| | (1) | (2) | (3) | (4) |
| Soporte = 1 | -0.28 (2.36) | -1.17 (2.48) | -0.59 (2.30) | 3.03 (2.09) |
| Tendencia Lineal | | 7.85** (0.30) | 8.10** (0.30) | 8.31** (0.29) |
| Soporte X Tendencia | | -0.52 (0.41) | -0.39 (0.41) | -0.49 (0.40) |
| I(2015) = 1 | 36.04** (1.20) | 26.31** (1.38) | 26.07** (1.36) | 24.96** (1.36) |
| Soporte X I(2015) | 5.66** (1.67) | 7.07** (1.91) | 6.65** (1.89) | 6.67** (3.03) |
| Tamaño del Efecto (DE) | 0.08** | 0.10** | 0.10** | 0.10** |
| Características del Modelo | | | | |
| Inicio Programa | 2015 | 2015 | 2015 | 2015 |
| Efectos Fijos | No | No | Dominio | Dominio |
| Otros Controles | No | No | No | Sí |
| Años Usados | 2013-15 | 2009-15 | 2009-15 | 2009-15 |
| N Estudiantes | 296,804 | 694,114 | 694,114 | 691,523 |
| N Escuelas | 2,195 | 2,195 | 2,195 | 2,194 |

Notas: errores estándar agrupados (*clustered*) al nivel de la escuela en paréntesis. Controles usados a nivel de la escuela: variable indicadora si recibió “Acompañamiento Pedagógico Multigrado”, variable indicadora si recibió “Acompañamiento Pedagógico Intercultural Bilingüe”, tasa de aprobación estudiantil, total de docentes, total de alumnos, y porcentaje de alumnos de lengua indígena. Controles usados a nivel distrital: tasa de pobreza, tasa de pobreza extrema, y tasa de desnutrición del 2009. La muestra incluye escuelas emparejadas a partir del método

de *caliper matching* usando información histórica de rendimiento para el periodo 2011-2014 y otras variables que caracterizan a la escuela y a los estudiantes que atienden la escuela. El proceso de emparejamiento fue estratificado usando Dominio Geográfico de la escuela. DE=Desviaciones Estándar

Nivel de significancia: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Los resultados del modelo de DiD de la Tabla 6, que evalúa los efectos del programa en la prueba de matemática de la ECE, son bastante similares a los de la prueba de comprensión lectora. Por una parte, el rendimiento promedio en matemática de los estudiantes del grupo de intervención y de comparación son muy similares en el periodo anterior a la implementación del programa, con solo 1.29 puntos de diferencia en la prueba. Segundo, el grupo de comparación también exhibe un incremento en el tiempo de casi 40 puntos en el rendimiento en la prueba de matemática. Por último, el efecto causal del programa es de 6.5 puntos adicionales, que corresponde a una ganancia de aproximadamente 0.08 desviaciones estándar.

Tabla 6.
Resultados de impacto de la intervención de Soporte Pedagógico en matemática

| | DiD | CITS | | |
|-----------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| | (1) | (2) | (3) | (4) |
| Soporte = 1 | 1.29 (3.48) | 0.59 (3.57) | 0.96 (3.31) | 4.69 (3.10) |
| Tendencia Lineal | | 7.39** (0.48) | 7.74** (0.48) | 7.90** (0.47) |
| Soporte X Tendencia | | -0.08 (0.66) | 0.02 (0.66) | -0.04 (0.65) |
| I(2015) = 1 | 39.53** (1.83) | 32.50** (2.13) | 32.11** (2.11) | 31.24** (2.10) |
| Soporte X I(2015) | 6.54** (2.59) | 7.32** (3.01) | 6.91** (2.99) | 6.85** (2.99) |
| Tamaño del Efecto (DE) | 0.08** | 0.09** | 0.09** | 0.09** |
| Características del Modelo | | | | |
| Inicio Programa | 2015 | 2015 | 2015 | 2015 |
| Efectos Fijos | No | No | Dominio | Dominio |
| Otros Controles | No | No | No | Sí |
| Años Usados | 2013-15 | 2009-15 | 2009-15 | 2009-15 |
| N Estudiantes | 296,804 | 694,114 | 694,114 | 691,523 |
| N Escuelas | 2,195 | 2,195 | 2,195 | 2,194 |

Nota: ver notas de la Tabla 5.

Los resultados del modelo CITS, que estima los efectos de SP en comprensión lectora y matemática, se presentan en las columnas 2 a 4 de las Tablas 5 y 6, respectivamente. En la columna 2 de la Tabla 5, se observa que las escuelas que recibieron el programa de SP obtuvieron 7.04 puntos adicionales (equivalente a 0.10 desviaciones estándar) en la prueba de comprensión lectora con relación a las escuelas del grupo de comparación. Por su parte, el efecto estimado en la prueba de matemática, que se presenta en la columna 2 de la Tabla 6, indica que el grupo de intervención obtuvo 7.32 puntos adicionales (equivalente a 0.09 desviaciones estándar) con respecto al grupo de comparación. Estos resultados son un poco mayores a los estimados a partir del modelo de DiD, aunque la diferencia del efecto estimado entre ambos modelos no es lo suficientemente grande como para resultar estadísticamente significativa. Asimismo, se observa que los efectos para ambas pruebas son robustos a la inclusión de otras variables de control tales como efectos fijos de dominio geográfico (columna 3), o a la inclusión de un gran conjunto de variables exógenas que pueden variar en el tiempo (columna 4), entre las que se incluyen variables indicadoras de otros programas de acompañamiento pedagógico (multigrado e intercultural bilingüe), tasa de aprobación estudiantil, total de docentes, total de alumnos, porcentaje de alumnos de lengua indígena, tasa de pobreza distrital, tasa de pobreza extrema distrital y tasa de desnutrición distrital. La estabilidad de los impactos estimados del programa provee evidencia adicional en favor de la estrategia de identificación empleada.

Por otra parte, resulta interesante discutir los coeficientes estimados tanto para la tendencia lineal, como para la interacción de esta última con la variable dicotómica del año 2015. Con respecto a la primera variable, se puede observar que la pendiente, tanto de las pruebas de comprensión lectora como de matemática, es de aproximadamente 7 puntos adicionales cada año, lo cual indica que cada nueva cohorte de segundo de primaria es sistemáticamente mejor que la anterior cohorte en términos de su desempeño en las pruebas de la ECE. Por su parte, los resultados de la variable “Soporte X Tendencia” no son estadísticamente diferentes de cero, lo cual indica que las pendientes de las trayectorias de las variables de desempeño entre 2009 y 2014 para los grupos de intervención y de comparación son iguales. Este último resultado es la prueba empírica de que el supuesto de las tendencias paralelas se cumple.

Impacto en los niveles de logro de la ECE

Los modelos de DiD y de CITS también pueden ser empleados con el fin de investigar a qué tipo de estudiantes el programa de SP ha beneficiado más. Para ello usamos las variables que describen el nivel de logro en la prueba de la ECE al que pertenecen

los estudiantes de una escuela dada a partir de tres categorías: inicio, proceso y satisfactorio. Por ejemplo, en promedio, de las escuelas que recibieron el programa de SP en 2015, el 27 por ciento estaba en nivel inicio, el 42 por ciento en proceso, y el 31 por ciento en el nivel satisfactorio.

En las Tablas 7 y 8 presentamos los resultados de impacto de SP sobre la proporción de estudiantes que se encuentra en el nivel de logro “inicial” en comprensión lectora y matemática, respectivamente.²⁵ La interpretación de los diferentes modelos y especificaciones son iguales a las discutidas en la sección anterior y por tanto son omitidas aquí. Los resultados para la variable de interés, “Soporte X I(2015)”, indican que las escuelas que recibieron el programa de SP en 2015 son tres puntos porcentuales menos propensas a tener estudiantes en el nivel de logro “inicial” con relación a las escuelas del grupo de comparación. Estos resultados son iguales tanto para la prueba de comprensión lectora como de matemática y, al igual que en las Tablas 5 y 6, los resultados son robustos a los modelos empleados o al tipo de variables de control usadas. Por otra parte, no se encontraron resultados significativos para los niveles de logro “de proceso” y “satisfactorio” entre el grupo de intervención y de comparación para ninguna de las dos pruebas. Estos resultados sugieren que los efectos del programa de SP estimados en la sección anterior se concentran más en aquellos estudiantes que presenten un desempeño bajo en las pruebas de la ECE; es decir, el programa parece estar ayudando más a aquellos estudiantes que tienen los puntajes más bajos de desempeño académico.

25 De igual modo se estimaron modelos para la probabilidad de estar en el nivel satisfactorio. Sin embargo, para dicha categoría no se encontraron diferencias importantes entre los grupos de tratamiento y comparación. Es decir, los efectos estimados ocurren en su totalidad por las diferencias en la composición de los grupos de inicio y proceso.

Tabla 7.
Resultados de impacto de la intervención de Soporte Pedagógico en la probabilidad de estar en nivel de logro “Inicial” en comprensión lectora

| | DiD | CITS | | |
|-----------------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| | (1) | (2) | (3) | (4) |
| Soporte = 1 | -0.00 (0.01) | -0.00 (0.01) | -0.00 (0.01) | -0.02 (0.01) |
| Tendencia Lineal | | -0.02** (0.00) | -0.02** (0.00) | -0.02** (0.00) |
| Soporte X Tendencia | | 0.00 (0.00) | -0.00 (0.00) | -0.00 (0.00) |
| I(2015) = 1 | -0.15** (0.01) | -0.13** (0.01) | -0.13** (0.01) | -0.13** (0.01) |
| Soporte X I(2015) | -0.02** (0.01) | -0.03** (0.01) | -0.03** (0.01) | -0.03** (0.01) |
| Características del Modelo | | | | |
| Inicio Programa | 2015 | 2015 | 2015 | 2015 |
| Efectos Fijos | No | No | Dominio | Dominio |
| Otros Controles | No | No | No | Sí |
| Años Usados | 2013-15 | 2009-15 | 2009-15 | 2009-15 |
| N Estudiantes | 296,804 | 694,114 | 694,114 | 691,523 |
| N Escuelas | 2,195 | 2,195 | 2,195 | 2,194 |

Nota: ver notas de la Tabla 5.

Tabla 8.
Resultados de impacto de la intervención de Soporte Pedagógico en la probabilidad de estar en nivel de logro “Inicial” en matemática

| | DiD | CITS | | |
|-----------------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| | (1) | (2) | (3) | (4) |
| Soporte = 1 | -0.00 (0.01) | -0.00 (0.01) | -0.00 (0.01) | -0.02 (0.01) |
| Tendencia Lineal | | -0.02** (0.00) | -0.02** (0.00) | -0.02** (0.00) |
| Soporte X Tendencia | | 0.00 (0.00) | -0.00 (0.00) | -0.00 (0.00) |
| I(2015) = 1 | -0.15** (0.01) | -0.13** (0.01) | -0.13** (0.01) | -0.13** (0.01) |
| Soporte X I(2015) | -0.02** (0.01) | -0.03** (0.01) | -0.03** (0.01) | -0.03** (0.01) |
| Características del Modelo | | | | |
| Inicio Programa | 2015 | 2015 | 2015 | 2015 |
| Efectos Fijos | No | No | Dominio | Dominio |
| Otros Controles | No | No | No | Sí |
| Años Usados | 2013-15 | 2009-15 | 2009-15 | 2009-15 |
| N Estudiantes | 296,804 | 694,114 | 694,114 | 691,523 |
| N Escuelas | 2,195 | 2,195 | 2,195 | 2,194 |

Nota: ver notas de la Tabla 5.

Análisis de robustez

En esta sección investigamos empíricamente si existe evidencia de que el programa de SP tuvo algún efecto en el año 2014. Este escenario es improbable dado que SP inició actividades en julio de 2014, tan solo 3 meses antes de que se recoja la Evaluación Censal Escolar. Para investigar esto, empleamos la Muestra B, la cual efectúa el emparejamiento de escuelas sin incluir la información de desempeño escolar para el año 2014. De haberse implementado SP en el año 2014, los efectos estimados del programa presentados en las Tablas 5 y 6 podrían estar sesgados hacia abajo.

Los resultados de los análisis de robustez se presentan en las Tablas 9 y 10 para comprensión lectora y matemática, respectivamente. La diferencia más importante entre las especificaciones presentadas en las Tablas 5 y 6, y las presentadas en las Tablas 9 y 10, es que en estas últimas los efectos de SP se estiman tanto para el año 2014 como para el 2015. En particular, la variable de interés para el año 2014 está dada por la interacción “Soporte X I(2014)”. Los resultados de la Tabla 9 indican que el efecto de SP en comprensión lectora en el 2014 fue de aproximadamente 1 punto en la prueba a favor del grupo de comparación, no siendo este resultado estadísticamente significativo.

Tabla 9.
Análisis de robustez con muestra B: resultados de impacto de la intervención Soporte Pedagógico en comprensión lectora

| | DiD | CITS | | |
|---------------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| | (1) | (2) | (3) | (4) |
| Soporte = 1 | -4.50* (2.54) | -4.49* (2.52) | -4.43** (2.26) | 0.36 (2.11) |
| Tendencia Lineal | | 4.51** (0.33) | 4.80** (0.33) | 4.95** (0.32) |
| Soporte X Tendencia | | -0.15 (0.47) | -0.02 (0.46) | -0.11 (0.46) |
| I(2014) = 1 | 24.78** (1.32) | 22.21** (1.35) | 22.17** (1.34) | 21.72** (1.35) |
| I(2015) = 1 | 47.64** (1.45) | 40.57** (1.68) | 40.32** (1.65) | 39.59** (1.68) |
| Soporte X I(2014) | -1.12 (1.86) | -0.99 (1.94) | -1.24 (1.92) | -1.36 (1.92) |
| Soporte X I(2015) | 5.64** (1.95) | 5.92** (2.33) | 5.31** (2.30) | 5.13** (2.30) |
| Tamaño del Efecto (DE) en 2015 | 0.08** | 0.09** | 0.08** | 0.07** |
| Características del Modelo | | | | |
| Inicio Programa | 2014 | 2014 | 2014 | 2014 |
| Efectos Fijos | No | No | Dominio | Dominio |
| Otros Controles | No | No | No | Sí |
| Años Usados | 2013-15 | 2009-15 | 2009-15 | 2009-15 |
| N Estudiantes | 267,266 | 636,985 | 636,985 | 634,882 |
| N Escuelas | 1,865 | 1,865 | 1,865 | 1,864 |

Notas: ver notas de la Tabla 5. La muestra incluye escuelas emparejadas a partir del método de *nearest neighbor* usando información histórica de rendimiento para el periodo 2010-2013 y otras variables de la escuela.

Por su parte, los resultados en la prueba de matemática (Tabla 10) indican también una diferencia negativa, pero no significativa, entre el grupo de intervención y de comparación en 2014. Note que los impactos estimados para comprensión y matemática en el año 2015, si bien son positivos y en algunos casos significativos, resultan inferiores a los hallados al estimar los modelos con la Muestra A. Dado que la muestra B no incluyó información de rendimiento de 2014 en el emparejamiento, los resultados del impacto del programa para 2014 se pueden interpretar como evidencia de que el programa de soporte no tuvo efectos en 2014. Otra posible interpretación es que no hay evidencias de que el programa haya en realidad empezado en 2014, dado que en los primeros meses solo se implementó de manera parcial y dado que el programa se inició tan solo cuatro meses antes de que se recogiera la ECE. Por último, cabe resaltar que los impactos estimados para el año 2015 resultan muy similares a los obtenidos con la Muestra A, con la diferencia de que la estimación de la tendencia que se puede hacer

a partir de la Muestra A es mucho más precisa que con la Muestra B, dado que en la Muestra A se puede emplear un año más en la estimación de la tendencia.

Tabla 10.
Análisis de robustez con muestra B: resultados de impacto de la intervención de Soporte Pedagógico en matemática

| | DiD | CITS | | |
|---------------------------------------|-------------------------------|------------------------------|------------------------------|------------------------------|
| | (1) | (2) | (3) | (4) |
| Soporte = 1 | -3.12 (3.71) | -3.41 (3.42) | -3.64 (3.08) | 1.17 (2.99) |
| Tendencia Lineal | | 1.63** (0.52) | 2.04** (0.52) | 1.99** (0.50) |
| Soporte X Tendencia | | 0.49 (0.75) | 0.65 (0.74) | 0.62 (0.73) |
| I(2014) = 1 | 36.54** (2.12) | 40.21** (2.17) | 40.12** (2.14) | 40.52** (2.15) |
| I(2015) = 1 | 56.03** (2.21) | 58.07** (2.62) | 57.60** (2.59) | 58.00** (2.61) |
| Soporte X I(2014) | -2.41 (2.99) | -2.61 (3.10) | -2.98 (3.06) | -2.99 (3.06) |
| Soporte X I(2015) | 5.05* (3.00) | 4.35 (3.65) | 3.62 (3.61) | 3.41 (3.61) |
| Tamaño del Efecto (DE) en 2015 | 0.06* | 0.05 | 0.05 | 0.04 |
| Características del Modelo | | | | |
| Inicio Programa | 2014 | 2014 | 2014 | 2014 |
| Efectos Fijos | No | No | Dominio | Dominio |
| Otros Controles | No | No | No | Sí |
| Años Usados | 2013-15 | 2009-15 | 2009-15 | 2009-15 |
| N Estudiantes | 267,266 | 636,985 | 636,985 | 634,882 |
| N Escuelas | 1,865 | 1,865 | 1,865 | 1,864 |

Notas: ver notas de la Tabla 5. La muestra incluye escuelas emparejadas a partir del método de *nearest neighbor* usando información histórica de rendimiento para el periodo 2010-2013 y otras variables de la escuela.

Análisis de heterogeneidad por nivel de implementación

Usando dos indicadores “próximos” del nivel de implementación del programa, tales como el porcentaje de docentes fortaleza y el porcentaje de acompañantes pedagógicos contratados, exploramos si el impacto de SP era mayor en las escuelas ubicadas en las unidades ejecutoras en donde el 90 por ciento de dichas contrataciones se realizaron antes del mes de julio de 2015. En particular, se exploró excluir del marco muestral a las escuelas ubicadas en las Unidades Ejecutoras 1, 2, 4 y 5 de Lima Metropolitana por retraso en la contratación de los Docentes Fortaleza para Refuerzo Escolar y

las escuelas de la Unidad Ejecutora 4 de Lima Metropolitana y las de Chiclayo de Lambayeque por demoras en la contratación de los acompañantes pedagógicos.²⁶ Es decir, la muestra de escuelas se redujo en cerca de 220 escuelas con respecto a la muestra completa presentada en las Tablas 5 y 6.

No encontramos diferencias notables con respecto al impacto de SP obtenido con la muestra de escuelas con “alta” implementación y con la muestra de escuelas emparejadas. Es decir, aun cuando se delimitó la muestra a las escuelas intervenidas con “alta” implementación (y sus escuelas emparejadas), el tamaño del efecto encontrado fue de aproximadamente 0.10 desviaciones estándar tanto para comprensión como matemática. Estos resultados se pueden explicar, en parte, debido a la poca variabilidad existente en las escuelas intervenidas en los dos indicadores “próximos” de implementación. Pero también es importante resaltar que la contratación del personal clave del programa es tan solo uno de los muchos pasos necesarios para implementar el programa de acuerdo a lo planeado. El personal pudo haberse contratado a tiempo en su gran mayoría, pero tal vez no se encontraba lo suficientemente preparado o capacitado para ejercer las labores que SP requería; o tal vez las capacitaciones sí ocurrieron pero los docentes necesitaban más práctica y retroalimentación para realizar su labor de acuerdo a lo planeado por la intervención. Las Tablas 11 y 12 presentan los resultados de los análisis de heterogeneidad.

26 Estos datos fueron obtenidos directamente de la Secretaria de Planificación Estratégica.

Tabla 11.
Análisis de heterogeneidad: Resultados de impacto de la intervención de Soporte Pedagógico en comprensión lectora en escuelas con nivel de implementación alta

| | DiD | CITS | | |
|-----------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| | (1) | (2) | (3) | (4) |
| Soporte = 1 | 0.58 (2.61) | 0.20 (2.72) | 0.14 (2.58) | 4.24* (2.33) |
| Tendencia Lineal | | 7.53** (0.33) | 7.75** (0.33) | 7.98** (0.32) |
| Soporte X Tendencia | | 0.06 (0.45) | 0.19 (0.45) | 0.07 (0.44) |
| I(2015) = 1 | 35.81** (1.31) | 27.19** (1.53) | 26.78** (1.51) | 25.45** (1.50) |
| Soporte X I(2015) | 6.71** (1.81) | 7.02** (2.09) | 6.79** (2.07) | 6.88** (2.06) |
| Tamaño del Efecto (DE) | 0.10** | 0.10** | 0.10** | 0.10** |
| Características del Modelo | | | | |
| Inicio Programa | 2015 | 2015 | 2015 | 2015 |
| Efectos Fijos | No | No | Dominio | Dominio |
| Otros Controles | No | No | No | Sí |
| Años Usados | 2013-15 | 2009-15 | 2009-15 | 2009-15 |
| N Estudiantes | 256,565 | 595,577 | 595,577 | 593,308 |
| N Escuelas | 1,973 | 1,973 | 1,973 | 1,972 |

Notas: ver notas de la Tabla 5. La muestra empleada excluye aquellas escuelas pertenecientes a las UGEL con bajo avance en la implementación del programa a julio de 2015 (UGEL 1, 2, 4 y 5 de Lima Metropolitana por retraso en contratación de docentes fortaleza para el componente de Refuerzo Escolar) y las UGEL 4 de Lima Metropolitana y Chiclayo de Lambayeque por demora en la contratación de los acompañantes pedagógicos). Estos resultados fueron obtenidos con la muestra analítica o Muestra A.

Tabla 12.
Análisis de heterogeneidad: Resultados de impacto de la intervención de Soporte Pedagógico en matemática en escuelas con nivel de implementación alta

| | DiD | CITS | | |
|-----------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| | (1) | (2) | (3) | (4) |
| Soporte = 1 | 1.52 (3.87) | 1.66 (3.95) | 1.41 (3.71) | 5.78* (3.45) |
| Tendencia Lineal | | 6.74** (0.52) | 7.08** (0.52) | 7.25** (0.51) |
| Soporte X Tendencia | | 0.55 (0.72) | 0.66 (0.72) | 0.58 (0.71) |
| I(2015) = 1 | 39.13** (1.99) | 33.78** (2.36) | 33.17** (2.33) | 32.07** (2.32) |
| Soporte X I(2015) | 8.30** (2.81) | 7.61** (3.29) | 7.37** (3.27) | 7.44** (3.26) |
| Tamaño del Efecto (DE) | 0.10** | 0.10** | 0.09** | 0.09** |
| Características del Modelo | | | | |
| Inicio Programa | 2015 | 2015 | 2015 | 2015 |
| Efectos Fijos | No | No | Dominio | Dominio |
| Otros Controles | No | No | No | Sí |
| Años Usados | 2013-15 | 2009-15 | 2009-15 | 2009-15 |
| N Estudiantes | 256,565 | 595,577 | 595,577 | 593,308 |
| N Escuelas | 1,973 | 1,973 | 1,973 | 1,972 |

Nota: ver notas de la Tabla 11.

7. DISCUSIÓN Y RECOMENDACIONES

En este informe describimos los resultados de la evaluación de impacto de Soporte Pedagógico, una intervención implementada por el Ministerio de Educación del Perú. La intervención trabaja simultáneamente sobre cuatro líneas: fortalecimiento de desempeños en docentes y directivos de la institución educativa, refuerzo escolar, materiales y recursos educativos, y articulación con la familia y la comunidad. Con ello, SP tiene como objetivo mejorar la calidad del servicio educativo brindado en las escuelas estatales, urbanas y polidocentes completas de primaria.

Para evaluar el impacto de la intervención de SP se utilizaron exclusivamente datos de fuentes secundarias recogidas por el Ministerio de Educación. Específicamente, se usaron los logros de aprendizaje de los alumnos de segundo grado de primaria en las áreas de comprensión lectora y matemática recogidos a través de la Evaluación Censal de Estudiantes entre los años 2009 y 2015. Asimismo, se usaron datos históricos del Censo Escolar.

El mayor reto en este tipo de estudios es que los resultados obtenidos sufran de algún tipo de sesgo de selección. Los estudiantes de las escuelas participantes en el programa de SP pueden ser muy diferentes al estudiante promedio de otras escuelas públicas de educación primaria, por cuanto la estimación de los efectos del programa que resulta de comparar a los estudiantes de SP con otros estudiantes podría estar sesgada. Algunos de estos sesgos pueden ser eliminados al controlar por características observables de los estudiantes, de sus familias, y de las escuelas que atienden, pero la posibilidad de que persistan algunos sesgos debido a características no observables permanece, como la motivación o la habilidad innata de los estudiantes.

Para evaluar el impacto de Soporte Pedagógico usamos una combinación de estrategias empíricas que emplea el modelo de *Comparative Interrupted Time Series* (CITS) y de Diferencia en Diferencias (DiD), con una estrategia de emparejamiento que incluye *propensity score matching* con estratificación geográfica. La estrategia empírica usó varios **años de rendimiento académico** de matemática y comprensión lectora previos a la introducción de SP, para seleccionar al grupo de intervención y comparación de la muestra analítica. La estrategia maximiza el uso de los datos longitudinales al nivel de la escuela, y logra la comparabilidad de los grupos tanto en términos de los valores medios como de las trayectorias de rendimiento académico. De este modo, la estrategia empírica usada permitió controlar simultáneamente tanto por diferencias observables así

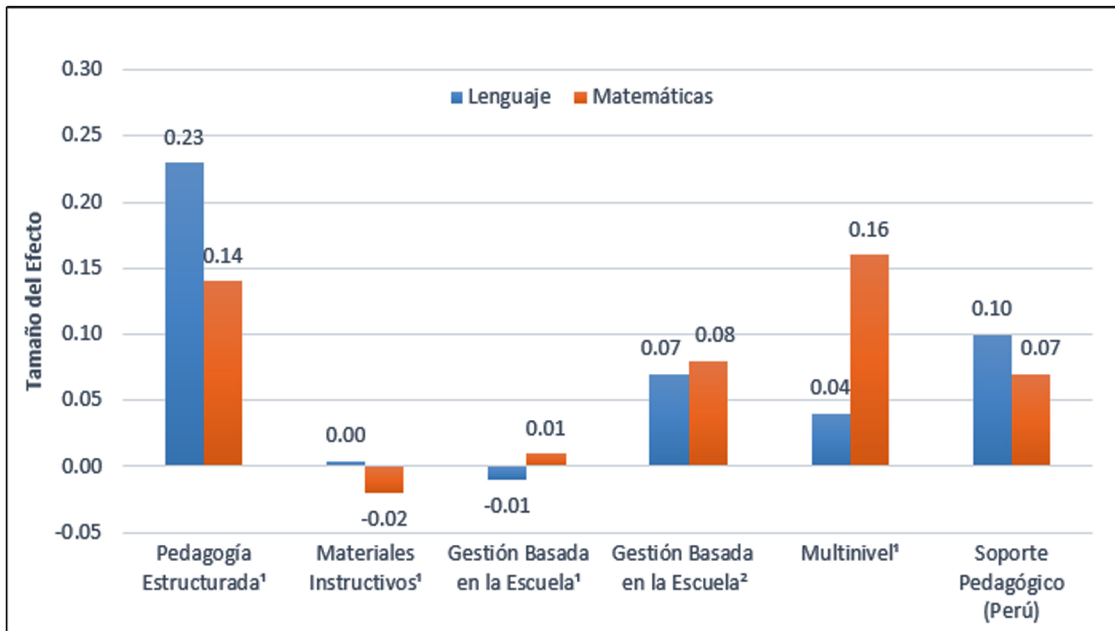
como por diferencias no observables, que no varían en el tiempo, entre los grupos de intervención y de comparación.

Los efectos estimados en este estudio indican que la intervención de SP, después de poco menos de un año de imperfecta implementación, tuvo efectos positivos en el rendimiento académico de los estudiantes, tanto en comprensión lectora como en matemática. Dichos impactos, si bien son modestos, resultan robustos a diferentes tipos de estrategias empíricas, variables de control, y construcción de la muestra final a partir del proceso de emparejamiento. En promedio, los resultados muestran que los estudiantes que recibieron el programa de SP presentan resultados superiores en las pruebas de la ECE de aproximadamente 0.10 desviaciones estándar en comprensión lectora con relación a los estudiantes del grupo de comparación y de 0.09 desviaciones estándar en matemática. Por su parte, los resultados en las variables de “niveles de logro” de la ECE muestran que los estudiantes de SP resultan 3 puntos porcentuales menos probables de pertenecer al grupo más bajo o “en inicio”, tanto en comprensión como en matemática. Es decir, los análisis indican que SP tuvo un mayor impacto en aquellos estudiantes con los puntajes más bajos de desempeño académico.

En comparación con otros estudios, la intervención de SP presenta efectos estimados que son similares a los encontrados en evaluaciones de programas educativos en países de bajos y medianos ingresos. Las revisiones sistemáticas de la literatura realizadas recientemente por Snilstveit, et al. (2015) y Carr-Hill et al. (2015) resumen el tamaño del efecto de diferentes programas educativos que tenían como objetivo mejorar el acceso y el aprendizaje de los estudiantes. La Figura 5 ilustra el tamaño promedio de los efectos encontrados a partir de la metodología de meta-análisis en las áreas curriculares de matemática y lenguaje. Por ejemplo, para las intervenciones clasificadas como “pedagogía estructurada”, que tienen como objetivo cambiar las prácticas en las aulas a través del desarrollo de nuevos contenidos enfocados en un área particular, materiales para estudiantes y docentes, y capacitaciones de corto plazo para docentes en la entrega de estos nuevos contenidos, el tamaño promedio del efecto es de 0.23 desviaciones estándar para lenguaje y 0.14 desviaciones estándar para matemática. Asimismo, para las intervenciones que brindan materiales educativos (ejemplo: libros de texto, cuadernos, lapiceros, computadoras, mobiliario del aula, etcétera), el tamaño promedio es cero. Para los programas clasificados como de “gestión basada en la escuela” (o programas que intentan cambiar la gestión pedagógica e involucran a las comunidades, los padres de familia, y otros actores de interés en las decisiones de la escuela), el tamaño del efecto promedio es 0.07 y 0.08, respectivamente. Finalmente, para programas que combinan diferentes intervenciones denominados “multinivel” el tamaño del efecto es 0.04 y 0.16 para lenguaje y matemática, respectivamente. La

Figura 5 ilustra el tamaño promedio del efecto para diferentes tipos de intervenciones al nivel de la escuela y para matemática y comprensión lectora, incluyendo los efectos encontrados para la evaluación de Soporte Pedagógico.

Figura 5.
Tamaño del efecto en desviaciones estándar de Soporte Pedagógico en comparación con otras intervenciones pedagógicas dirigidas a escuelas primarias de países de bajos y medianos ingresos



Fuente: *Elaboración propia.*

Nota.1.Snilstveit et al. 2015. 2. Carr-Hill et al. 2015.

8. CONSIDERACIONES FINALES

Si bien los resultados reportados en este estudio son robustos a diferentes especificaciones y análisis de sensibilidad, es importante entender los resultados de este estudio sobre la base de sus limitaciones.

- Esta evaluación mide el impacto de SP en 2015 después de tan solo 8 meses de que la intervención iniciara las tareas de implementación ese año²⁷. Asimismo, tomando en cuenta la complejidad y amplia cobertura del programa a nivel nacional, el programa no logró implementarse con fidelidad y como originalmente se planeó. Si la intervención de SP lograra mejorar su implementación con el tiempo, los impactos presentados en este estudio podrían representar un efecto subestimado del programa (o un *lower bound effect*) y en el mediano plazo se esperaría que el impacto del programa se incremente. Sin embargo, dada la complejidad del programa será importante que el MINEDU desarrolle mecanismos eficientes de monitoreo de las actividades del programa, y establezca procesos claros para actuar sobre las dificultades y desafíos encontrados tanto en las escuelas como en las unidades ejecutoras.
- Las escuelas intervenidas fueron seleccionadas por el MINEDU antes de que se realice esta evaluación de impacto usando criterios específicos de focalización que se detallan en el Capítulo 2. Por esta razón, esta evaluación se planteó como un estudio es de carácter cuasi-experimental. No obstante, el estudio explotó el uso de datos históricos y secundarios disponibles para estimar el impacto de SP.
- Con los datos secundarios disponibles, solo fue posible estimar el impacto de la intervención de SP sobre los niveles de aprendizaje de los estudiantes de segundo grado de primaria, y en solo las áreas curriculares de matemática y comprensión lectora. Sin embargo, en 2015 el programa intervino todos los grados de primaria y en diversas áreas curriculares. Por lo tanto, es posible que el programa de SP haya tenido un mayor o menor impacto en otros grados de primaria o en otras áreas curriculares.
- Ninguno de los modelos presentados controla por variables omitidas que cambian en el tiempo, como por ejemplo movimientos de los estudiantes de regreso a escuelas públicas debido a la introducción de SP. La mayor amenaza a la validez interna del modelo CITS son factores que cambian con el tiempo,

²⁷ Los 8 meses se calculan entre el mes en que se iniciaron las clases (marzo) y el mes en el que se recogieron la ECE (noviembre).

que son introducidos al mismo tiempo que SP, y que afectan de manera diferente a los grupos de intervención y de comparación. Sin embargo, no encontramos evidencias concluyentes de la existencia de dichos factores.

- Si bien las estrategias de emparejamiento incrementaron en gran medida la comparabilidad del grupo de intervención y de comparación, solo fue posible emparejar aproximadamente al 38 por ciento de las escuelas intervenidas que contaban con datos de rendimiento académico. No obstante, se lograron emparejar escuelas de todas las regiones intervenidas, y en 10 de las 15 regiones se emparejaron más del 45 por ciento de las escuelas con SP.

BIBLIOGRAFÍA

- Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2005). *Remedying education: Evidence from two randomized experiments in India*. National Bureau of Economic Research.
- Bloom, H. (2001). *Measuring the impacts of whole-school reforms: Methodological lessons from an evaluation of accelerated schools*. Manpower Demonstration Research Corporation, New York.
- Bloom, H. (1999). *Estimating program impacts on student achievement using "short" interrupted time series*. Manpower Demonstration Research Corporation, New York.
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31(4), 416-440.
- Carr-Hill, R., Rolleston, C, Pherali, T., & Schendel, R. (2015). The effects of school-based decision making on educational outcomes in low and middle income contexts: a Systematic Review, 3ie Grantee Final Review. London: International Initiative for Impact Evaluation (3ie).
- Cantrell, S., Almasi, J., Carter, J., & Rintamaa, M. (2013). *Impacts of supplemental reading strategy instruction for adolescents: Results from a Randomized Trial and Follow-up Study*.
- Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, 30(3), 449-465.
- Conn, K. (2014). *Identifying effective education interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations*. Columbia University.
- Cochran, W.G. & Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhya*, 35, 417-446.
- Devarajan, S., & Reinikka, R. (2003). *Making services work for poor people*. World Bank and Oxford University Press.
- Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *The American Economic Review*, 1241-1278.
- Glewwe, P., Kremer, M., & Moulin, S. (2002). *Textbooks and test scores: Evidence from a prospective evaluation in Kenya*. Citeseer.
- Hanushek, E. A., Link, S., & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, 104, 212–232.

- King, E., & Ozler, B. (1998). What's decentralization got to do with learning? The case of Nicaragua's school autonomy reform. *Development Economics Research Group, Working Paper Series on Impact Evaluation of Education Reforms*, (9).
- Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340 (6130), 297–300.
- Lavy, V., & Schlosser, A. (2004). Targeted remedial education for under-performing teenagers.
- McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, 85(3), 353–394.
- Ministerio de Educación, (2015a). Evaluación Censal Escolar: Tendencias. Obtenido de <http://escale.minedu.gob.pe/tendencias>
- Ministerio de Educación, (2015b). Evaluación Censal Escolar: Información para Directores y Docentes. Obtenido de <http://umc.minedu.gob.pe/wp-content/uploads/2015/07/Folleto-ECE-2015-SECUNDARIA-.pdf>
- Muralidharan, K., & Sundararaman, V. (2009). Teacher performance pay: Experimental evidence from India. *National Bureau of Economic Research*.
- Navarro, J. C., & Verdisco, A. E. (2000). *Teacher training in Latin America: Innovation and trends*. Inter-American Development Bank.
- Piper, B. (2009). Impact study of SMRS using early grade reading assessment in three provinces in South Africa. *Integrated Education Programme, RTI International*.
- Piper, B., & Korda, M. (2010). Early grade reading assessment (EGRA) Plus: Liberia. Program evaluation report: Prepared for USAID/Liberia. Research Triangle Park, NC: RTI International.
- Resultados de la Evaluación Censal de Estudiantes 2014. Obtenido de: <http://umc.minedu.gob.pe/wp-content/uploads/2015/02/ECE-2014-Web-270215-27febv2.pdf>
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 417–458.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central of the propensity score in observational studies for causal effects. *Biometrika*, 17, 41-55.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using a multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33-88.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2, 169-188.

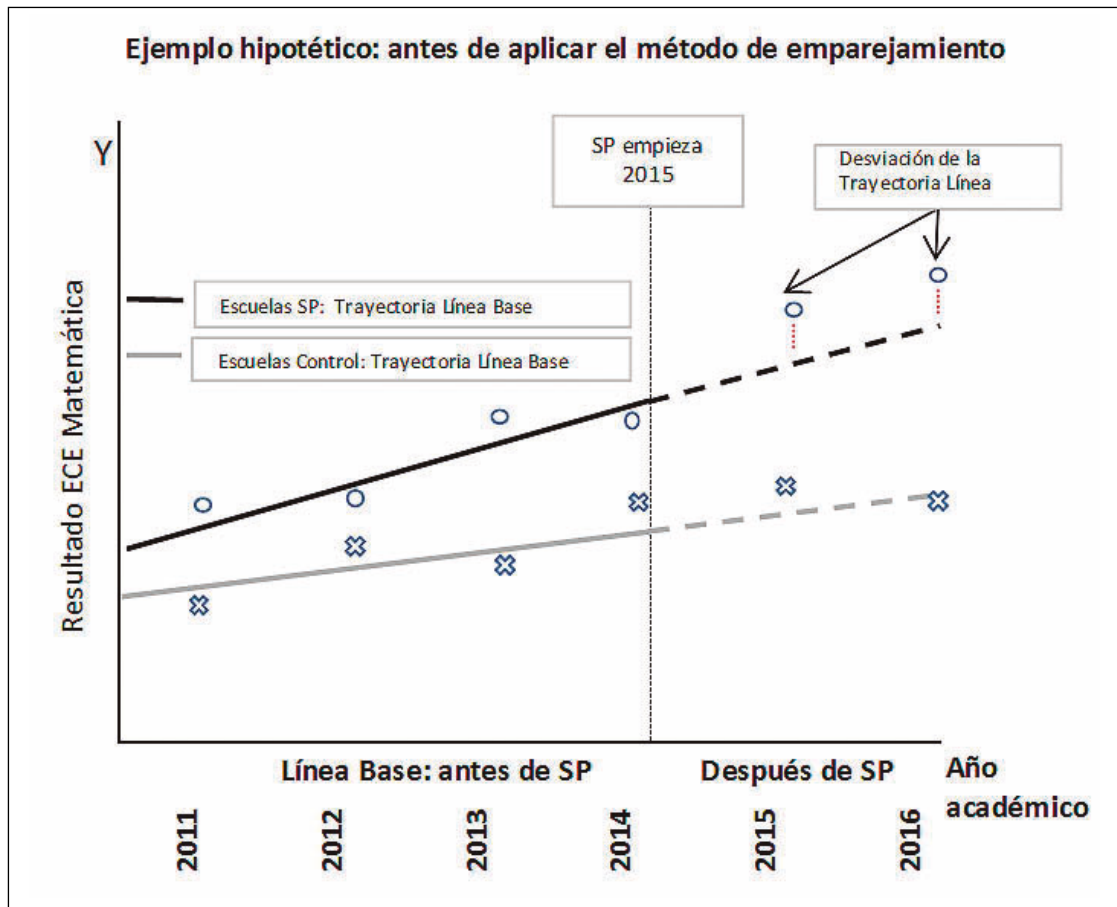
- St. Clair, T., Cook, T., & Hallberg, K. (2014). Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a randomized experiment. *American Journal of Evaluation*, 35(3): 311-327.
- Smith, J., & Todd, P. (2003). Does matching overcome Lalonde's critique of nonexperimental estimators?
- Snilstveit, B, Stevenson, J, Phillips, D, Vojtkova, M, Gallagher, E, Schmidt, T, Jobse, H, Geelen, M, Pastorello, M, and Eyers, J. (2015). Interventions for improving learning outcomes and access to education in low- and middle-income countries: A systematic review, 3ie Final Review. London: International Initiative for Impact Evaluation (3ie).
- Somers, M., Zhu, P., Jacob, R., & Bloom, H. (2013). The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation (MDRC working paper in research methodology). MDRC: New York, NY.
- Stuart, E.A. (2010). Matching Methods for Causal Inference: A review and a look forward. *Statistical Science* 25(1): 1-21.
- Stuart, E. A. (2007). Estimating Causal Effects Using School-Level Data Sets. *Educational Researcher*, 36(5), 187-198.
- UNICEF. (2014). Estado de la niñez en el Perú. Obtenido de http://www.unicef.org/peru/spanish/children_3787.htm.
- Villegas-Reimers, E. (1998). *The Preparation of Teachers in Latin America: Challenges and trends*. Human Development Department, World Bank, Latin America and the Caribbean Regional Office.

APÉNDICE A: METODOLOGÍA

El modelo CITS usa datos de varios años antes de la introducción de SP para determinar la trayectoria de rendimiento académico de “la línea base” para cada grupo. La línea base o periodo previo a la introducción de SP está conformada por los años entre 2011 y 2014. Los años de 2015 y 2016 corresponden al periodo denominado “después” de la introducción de SP. La línea sólida representa la trayectoria estimada antes de la introducción de SP, mientras que la línea con puntos representa el valor proyectado. El impacto de la intervención es evaluado al determinar si el rendimiento académico de 2015 del grupo de intervención se desvía de la trayectoria de la línea base por una mayor magnitud que la del grupo de comparación.

La Figura 1 ilustra una situación hipotética en donde las trayectorias base de los grupos de intervención y de comparación tienen tasas de crecimiento distintas y por tanto la ganancia observada en el grupo de intervención en el año 2015 se explica en parte porque este grupo venía creciendo a una tasa más rápida (que el grupo de comparación) y no necesariamente porque la intervención es más efectiva. El modelo de CITS permite controlar por las diferencias entre las tasas de crecimiento del grupo de intervención y de comparación.

Figura 1.
Ejemplo hipotético: estimación del impacto del programa de Soporte Pedagógico usando el modelo *Comparative Interrupted Time Series*



APÉNDICE B: RESULTADOS DEL MÉTODO DE EMPAREJAMIENTO

Muestra analítica del estudio: Muestra A

Tabla 1.
Proporción de escuelas emparejadas con Soporte Pedagógico por región

| | Grupo 1 | Grupo 2 | Grupo 3 | Grupo 4 | % Emparejados | | |
|---|---------|---------------------------------|---|-------------|---------------|-------------|-------------|
| Número de escuelas con Soporte Pedagógico | | | | | | | |
| Región | Total | Polidocente completas y urbanas | Polidocente completas, urbanas, estatales y con rendimiento académico | Emparejadas | Del grupo 1 | Del grupo 2 | Del grupo 3 |
| 1 Apurímac | 138 | 125 | 114 | 77 | 56% | 62% | 68% |
| 2 Arequipa | 17 | 17 | 17 | 11 | 65% | 65% | 65% |
| 3 Ayacucho | 207 | 183 | 153 | 116 | 56% | 63% | 76% |
| 4 Cajamarca | 221 | 213 | 208 | 102 | 46% | 48% | 49% |
| 5 Cusco | 256 | 234 | 223 | 147 | 57% | 63% | 66% |
| 6 Huancavelica | 91 | 82 | 75 | 60 | 66% | 73% | 80% |
| 7 Ica | 115 | 114 | 114 | 60 | 52% | 53% | 53% |
| 8 La Libertad | 222 | 221 | 220 | 60 | 27% | 27% | 27% |
| 9 Lambayeque | 114 | 114 | 113 | 24 | 21% | 21% | 21% |
| 10 Lima (metropolitana, callao y provincia) | 1,075 | 1,024 | 1014 | 177 | 16% | 17% | 17% |
| 11 Loreto | 205 | 183 | 177 | 86 | 42% | 47% | 49% |
| 12 Madre De Dios | 31 | 30 | 30 | 14 | 45% | 47% | 47% |
| 13 Piura | 327 | 318 | 317 | 80 | 24% | 25% | 25% |
| 14 San Martín | 116 | 105 | 105 | 67 | 58% | 64% | 64% |
| 15 Ucayali | 45 | 43 | 43 | 16 | 36% | 37% | 37% |
| Total | 3,180 | 3,006 | 2,923 | 1,098 | 35% | 37% | 38% |

Tabla 2.
Proporción de escuelas emparejadas por dominio geográfico

| Dominio geográfico | Número de escuelas con Soporte Pedagógico | | % emparejadas |
|------------------------------|---|-------------|---------------|
| | Con rendimiento 2015 | Emparejados | |
| 1 Costa Norte | 634 | 157 | 25% |
| 2 Costa Centro | 234 | 117 | 50% |
| 3 Costa Sur | 17 | 11 | 65% |
| 4 Sierra Norte | 169 | 72 | 43% |
| 5 Sierra Centro | 223 | 171 | 77% |
| 6 Sierra Sur | 312 | 211 | 68% |
| 7 Selva | 449 | 244 | 54% |
| 8 Lima Metropolitana, Callao | 885 | 115 | 13% |
| Total | 2,923 | 1,098 | 38% |

Figura 2.
Boxplots del *Propensity Score* antes y después de aplicar el método de emparejamiento o *Matching*

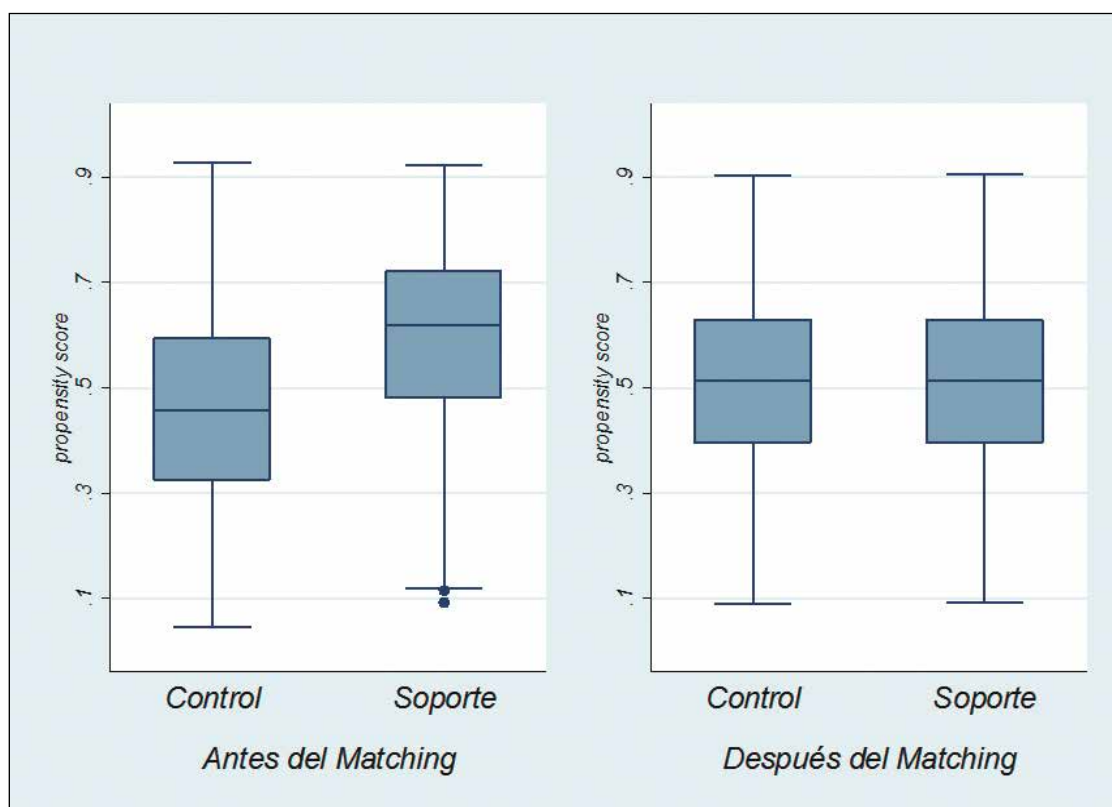


Tabla 3.
Características de las escuelas emparejas

| Variables | Sin Soporte Pedagógico | | Con Soporte Pedagógico | | Difer | ES | valor-p | Tamaño del efecto |
|--|------------------------|-------|------------------------|-------|-------|------|---------|-------------------|
| | Media | N1 | Media | N2 | | | | |
| 2014 total_alumnos | 270.99 | 1,098 | 301.90 | 1,098 | 30.91 | 9.19 | 0.00 | 0.14 |
| 2013 total_alumnos | 272.97 | 1,098 | 305.30 | 1,098 | 32.33 | 9.09 | 0.00 | 0.15 |
| 2014 tasa_aprobación | 72.56 | 1,098 | 72.98 | 1,098 | 0.42 | 0.32 | 0.19 | 0.06 |
| 2013 tasa_aprobación | 74.28 | 1,098 | 74.40 | 1,098 | 0.12 | 0.36 | 0.73 | 0.01 |
| 2014 % lengua indígena | 7.75 | 1,098 | 11.73 | 1,098 | 3.98 | 1.13 | 0.00 | 0.15 |
| 2014 con cuadernos escolares | 0.87 | 1,098 | 0.87 | 1,098 | -0.00 | 0.01 | 0.75 | -0.01 |
| 2014 con textos escolares | 0.89 | 1,098 | 0.89 | 1,098 | 0.00 | 0.01 | 0.89 | 0.01 |
| 2014 N computadoras | 48.26 | 1,098 | 49.70 | 1,098 | 1.44 | 1.72 | 0.40 | 0.04 |
| 2014 N computadoras con internet | 45.24 | 1,098 | 46.54 | 1,098 | 1.29 | 1.64 | 0.43 | 0.03 |
| 2014 total_docentes | 13.23 | 1,098 | 14.83 | 1,098 | 1.60 | 0.35 | 0.00 | 0.19 |
| 2014 ratio_alum_docentes | 19.51 | 1,098 | 19.39 | 1,098 | -0.12 | 0.26 | 0.64 | -0.02 |
| 2014 jornada_pedagógica | 8.08 | 1,098 | 8.09 | 1,098 | 0.00 | 0.05 | 0.92 | 0.00 |
| 2014 pared_ladrillo | 0.85 | 1,098 | 0.84 | 1,098 | -0.01 | 0.02 | 0.60 | -0.02 |
| 2014 piso_no_tierra | 0.99 | 1,098 | 1.00 | 1,098 | 0.00 | 0.00 | 0.53 | 0.03 |
| 2014 Inodoros | 1.48 | 1,098 | 1.37 | 1,098 | -0.11 | 0.18 | 0.54 | -0.03 |
| 2014 Desagüe | 0.81 | 1,098 | 0.86 | 1,098 | 0.05 | 0.02 | 0.00 | 0.15 |
| 2014 Electricidad_CP | 0.99 | 1,098 | 0.99 | 1,098 | -0.00 | 0.00 | 0.70 | -0.02 |
| 2014 Agua_Potable | 0.89 | 1,098 | 0.93 | 1,098 | 0.04 | 0.01 | 0.00 | 0.13 |
| 2014 escuela EIB | 0.10 | 1,098 | 0.13 | 1,098 | 0.03 | 0.01 | 0.01 | 0.10 |
| Tasa Pobreza del Distrito en el 2009 | 0.39 | 1,098 | 0.43 | 1,098 | 0.04 | 0.01 | 0.00 | 0.18 |
| Tasa Pobreza Extrema del Distrito en el 2009 | 0.14 | 1,098 | 0.17 | 1,098 | 0.03 | 0.01 | 0.00 | 0.19 |
| Tasa de Desnutrición del Distrito en 2009 | 24.08 | 1,098 | 28.56 | 1,098 | 4.49 | 0.67 | 0.00 | 0.28 |

Fuentes: ECE, Censo Escolar, Soporte Pedagógico.

Tabla 4.
Clasificación de las regiones según dominio geográfico

| Región | Costa | | | Sierra | | | Selva | Lima | Total |
|-----------------|------------|------------|-----------|------------|------------|------------|------------|------------|--------------|
| | Norte | Centro | Sur | Norte | Centro | Sur | | M/politana | |
| Amazonas | 0 | 0 | 0 | 15 | 0 | 0 | 29 | 0 | 44 |
| Ancash | 0 | 73 | 0 | 0 | 38 | 0 | 0 | 0 | 111 |
| Apurímac | 0 | 0 | 0 | 0 | 0 | 83 | 0 | 0 | 83 |
| Arequipa | 0 | 0 | 14 | 0 | 0 | 100 | 0 | 0 | 114 |
| Ayacucho | 0 | 0 | 0 | 0 | 113 | 0 | 13 | 0 | 126 |
| Cajamarca | 5 | 0 | 0 | 81 | 0 | 0 | 52 | 0 | 138 |
| Callao | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 116 | 116 |
| Cusco | 0 | 0 | 0 | 0 | 0 | 145 | 15 | 0 | 160 |
| Huancavelica | 0 | 0 | 0 | 0 | 67 | 0 | 0 | 0 | 67 |
| Huánuco | 0 | 0 | 0 | 0 | 38 | 0 | 22 | 0 | 60 |
| Ica | 0 | 86 | 0 | 0 | 0 | 0 | 0 | 0 | 86 |
| Junín | 0 | 0 | 0 | 0 | 59 | 0 | 26 | 0 | 85 |
| La Libertad | 69 | 0 | 0 | 38 | 0 | 0 | 0 | 0 | 107 |
| Lambayeque | 85 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 86 |
| Lima M/politana | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 114 | 114 |
| Lima Provincias | 0 | 75 | 0 | 0 | 13 | 0 | 0 | 0 | 88 |
| Loreto | 0 | 0 | 0 | 0 | 0 | 0 | 116 | 0 | 116 |
| Madre De Dios | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 14 |
| Moquegua | 0 | 0 | 2 | 0 | 0 | 5 | 0 | 0 | 7 |
| Pasco | 0 | 0 | 0 | 0 | 14 | 0 | 8 | 0 | 22 |
| Piura | 107 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 116 |
| Puno | 0 | 0 | 0 | 0 | 0 | 83 | 5 | 0 | 88 |
| San Martín | 0 | 0 | 0 | 0 | 0 | 0 | 143 | 0 | 143 |
| Tacna | 0 | 0 | 6 | 0 | 0 | 5 | 0 | 0 | 11 |
| Tumbes | 48 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 48 |
| Ucayali | 0 | 0 | 0 | 0 | 0 | 0 | 45 | 0 | 45 |
| Total | 314 | 234 | 22 | 144 | 342 | 421 | 488 | 230 | 2,195 |

Analisis de robustez: Muestra B

En esta sección se presentan los resultados para la Muestra B que se usan para los estudios de robustez. Nótese que, de manera similar, el ejercicio de emparejamiento generó dos grupos muy similares, como se ilustra en las Figuras 3, 4 y 5. La Tabla 5 presenta los descriptivos estadísticos de diversas variables después del emparejamiento y muestra que las diferencias de medias estandarizadas son en su gran mayoría menor a 0.25.

Figura 3.
Trayectorias de las muestras emparejadas para matemática
(Muestra B)

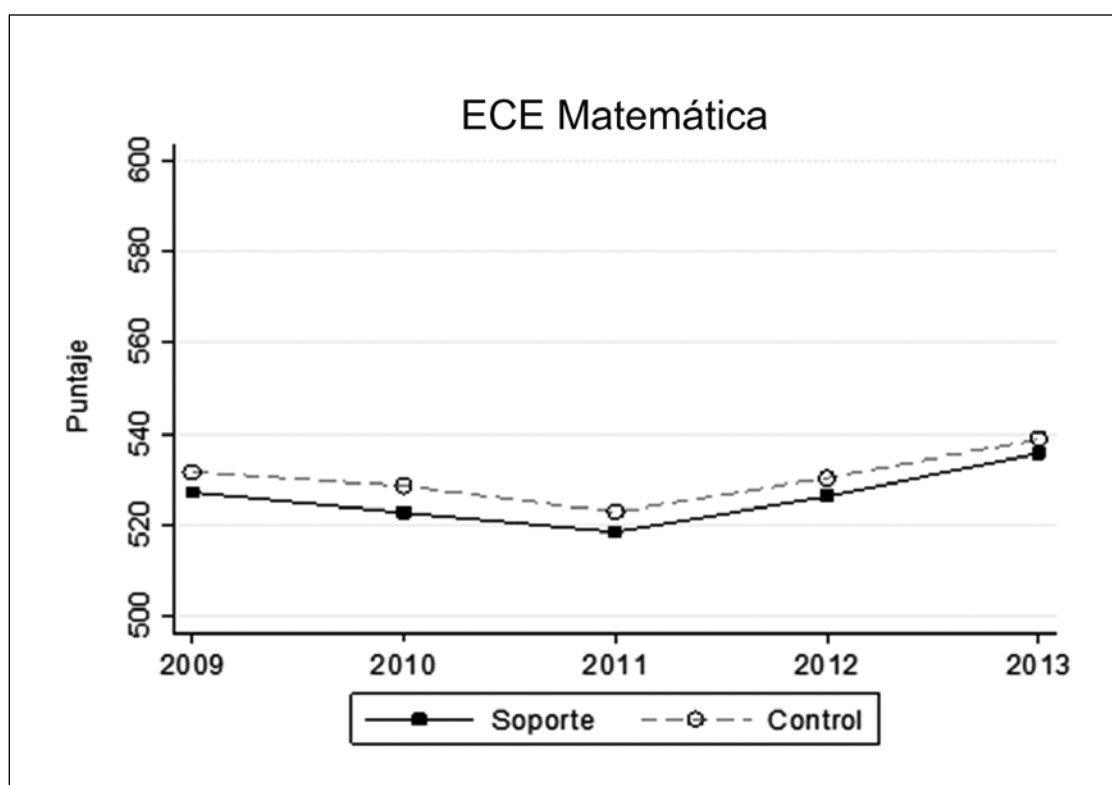


Figura 4.
Trayectorias de las muestras emparejadas para comprensión
(Muestra B)

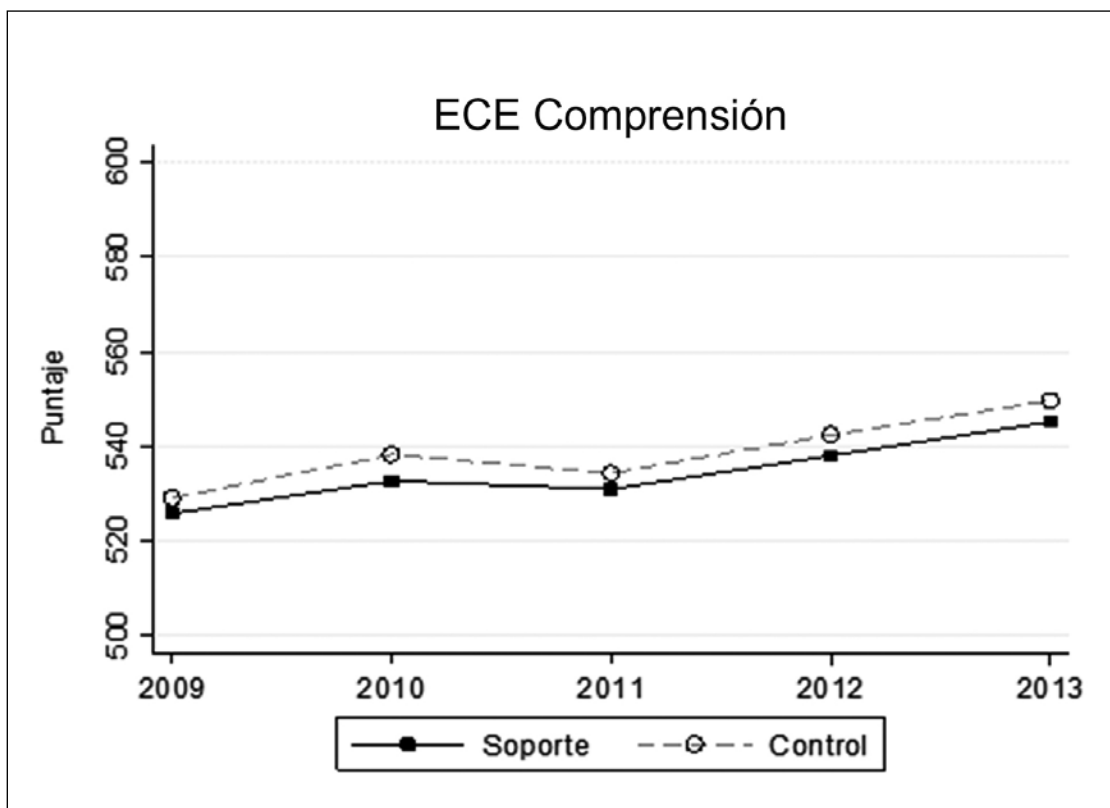


Figura 5.
Distribución del *Propensity Score* de las muestras emparejadas
(Muestra B)

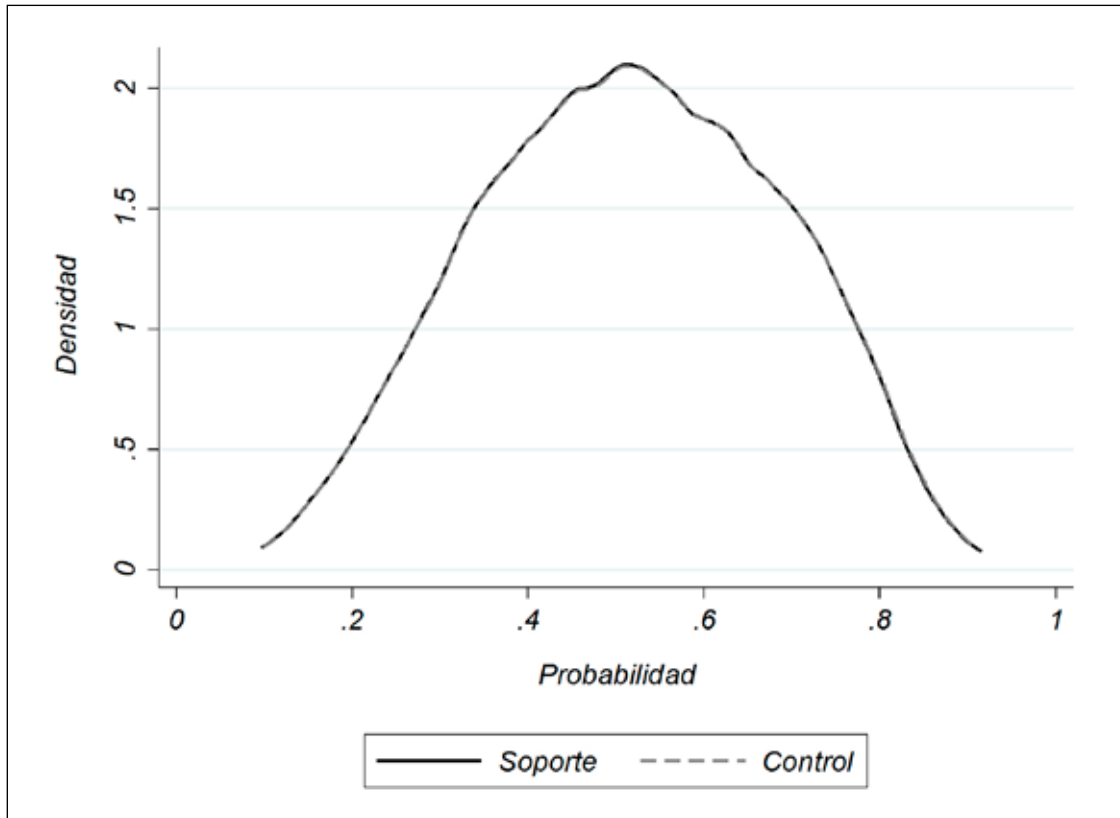


Tabla 5.
Características de las escuelas emparejas (Muestra B)

| Variables | Sin Soporte Pedagógico | | Con Soporte Pedagógico | | Dif | ES | Valor-p | Tamaño Del Efecto |
|--|------------------------|-----|------------------------|-----|-------|------|---------|-------------------|
| | Media | N1 | Media | N2 | | | | |
| Media ECE Comprensión Lectora 2013 | 542.06 | 933 | 537.15 | 933 | -4.92 | 2.26 | 0.03 | -0.10 |
| Media ECE Matemática 2013 | 531.50 | 933 | 525.96 | 933 | -5.54 | 3.09 | 0.07 | -0.08 |
| Media ECE Comprensión Lectora 2012 | 533.77 | 933 | 528.07 | 933 | -5.69 | 2.28 | 0.01 | -0.12 |
| Media ECE Matemática 2012 | 522.14 | 933 | 517.87 | 933 | -4.28 | 2.94 | 0.15 | -0.07 |
| Media ECE Comprensión Lectora 2011 | 525.64 | 933 | 521.01 | 933 | -4.64 | 2.31 | 0.04 | -0.09 |
| Media ECE Matemática 2011 | 514.66 | 933 | 509.21 | 933 | -5.45 | 2.92 | 0.06 | -0.09 |
| Media ECE Comprensión Lectora 2010 | 530.77 | 933 | 522.95 | 933 | -7.82 | 2.38 | 0.00 | -0.15 |
| Media ECE Matemática 2010 | 521.10 | 933 | 514.57 | 933 | -6.53 | 3.21 | 0.04 | -0.09 |
| % de estudiantes con logro en Com = Inicio en 2013 | 0.15 | 933 | 0.16 | 933 | 0.02 | 0.01 | 0.04 | 0.09 |
| % de estudiantes con logro en Mate = Inicio en 2013 | 0.47 | 933 | 0.49 | 933 | 0.03 | 0.01 | 0.03 | 0.10 |
| % de estudiantes con logro en Com = Satisfactorio en 2013 | 0.29 | 933 | 0.27 | 933 | -0.02 | 0.01 | 0.05 | -0.09 |
| % de estudiantes con logro en Mate = Satisfactorio en 2013 | 0.17 | 933 | 0.16 | 933 | -0.01 | 0.01 | 0.40 | -0.04 |
| % de estudiantes con logro en Com = Inicio en 2012 | 0.19 | 933 | 0.20 | 933 | 0.01 | 0.01 | 0.15 | 0.07 |
| % de estudiantes con logro en Mate = Inicio en 2012 | 0.49 | 933 | 0.51 | 933 | 0.02 | 0.01 | 0.09 | 0.08 |
| % de estudiantes con logro en Com = Satisfactorio en 2012 | 0.26 | 933 | 0.23 | 933 | -0.03 | 0.01 | 0.00 | -0.13 |
| % de estudiantes con logro en Mate = Satisfactorio en 2012 | 0.13 | 933 | 0.12 | 933 | -0.01 | 0.01 | 0.12 | -0.07 |
| % de estudiantes con logro en Com = Inicio en 2011 | 0.22 | 933 | 0.23 | 933 | 0.01 | 0.01 | 0.36 | 0.04 |
| % de estudiantes con logro en Mate = Inicio en 2011 | 0.55 | 933 | 0.56 | 933 | 0.02 | 0.01 | 0.16 | 0.07 |
| % de estudiantes con logro en Com = Satisfactorio en 2011 | 0.24 | 933 | 0.21 | 933 | -0.02 | 0.01 | 0.01 | -0.12 |
| % de estudiantes con logro en Mate = Satisfactorio en 2011 | 0.12 | 933 | 0.11 | 933 | -0.01 | 0.01 | 0.09 | -0.08 |
| Desviación Estándar ECE Comprension Lectora 2013 | 66.38 | 933 | 65.68 | 933 | -0.70 | 0.72 | 0.33 | -0.05 |
| Desviación Estándar ECE Matemática 2013 | 83.33 | 933 | 81.84 | 933 | -1.49 | 0.96 | 0.12 | -0.07 |
| Desviación Estándar ECE Comprension Lectora 2012 | 67.20 | 933 | 65.39 | 933 | -1.81 | 0.73 | 0.01 | -0.11 |
| Desviación Estándar ECE Matemática 2012 | 81.64 | 933 | 78.84 | 933 | -2.80 | 1.07 | 0.01 | -0.12 |
| Desviación Estándar ECE Comprension Lectora 2011 | 66.35 | 933 | 64.47 | 932 | -1.88 | 0.73 | 0.01 | -0.12 |

| | | | | | | | | |
|---|--------|-----|--------|-----|-------|-------|------|-------|
| Desviación Estándar ECE Matemática 2011 | 81.60 | 933 | 79.63 | 932 | -1.97 | 1.07 | 0.07 | -0.09 |
| 2013 total_alumnos | 294.77 | 933 | 316.88 | 933 | 22.11 | 10.44 | 0.03 | 0.10 |
| 2012 total_alumnos | 298.55 | 918 | 325.00 | 916 | 26.45 | 10.65 | 0.01 | 0.12 |
| 2011 total_alumnos | 308.96 | 933 | 336.02 | 932 | 27.05 | 10.78 | 0.01 | 0.12 |
| 2013 tasa_aprobación | 74.57 | 933 | 74.62 | 933 | 0.05 | 0.35 | 0.89 | 0.01 |
| 2012 tasa_aprobación | 74.71 | 918 | 74.30 | 916 | -0.41 | 0.38 | 0.28 | -0.05 |
| 2013 % lengua indígena | 4.71 | 933 | 10.15 | 933 | 5.44 | 1.06 | 0.00 | 0.24 |
| 2013 con cuadernos escolares | 0.89 | 933 | 0.87 | 933 | -0.02 | 0.02 | 0.25 | -0.05 |
| 2013 con textos escolares | 0.05 | 933 | 0.05 | 933 | 0.01 | 0.01 | 0.52 | 0.03 |
| 2013 N computadoras | 49.09 | 933 | 47.91 | 933 | -1.17 | 1.71 | 0.49 | -0.03 |
| 2013 N computadoras internet | 19.28 | 933 | 18.36 | 933 | -0.92 | 1.47 | 0.53 | -0.03 |
| 2013 total_docentes | 13.88 | 933 | 14.90 | 933 | 1.02 | 0.38 | 0.01 | 0.12 |
| 2013 ratio_alum_docentes | 20.33 | 933 | 20.31 | 933 | -0.02 | 0.29 | 0.94 | -0.00 |
| 2013 jornada_pedagógica | 8.08 | 933 | 8.09 | 933 | 0.01 | 0.05 | 0.85 | 0.01 |
| 2013 pared_ladrillo | 0.87 | 933 | 0.84 | 933 | -0.03 | 0.02 | 0.11 | -0.07 |
| 2013 piso_no_tierra | 0.99 | 933 | 0.99 | 933 | 0.00 | 0.00 | 0.76 | 0.01 |
| 2013 Inodoros | 1.45 | 933 | 1.36 | 933 | -0.09 | 0.19 | 0.65 | -0.02 |
| 2013 Desagüe | 0.79 | 933 | 0.83 | 933 | 0.04 | 0.02 | 0.04 | 0.09 |
| 2013 Electricidad_CP | 0.99 | 933 | 0.99 | 933 | 0.00 | 0.00 | 0.82 | 0.01 |
| 2013 Agua_Potable | 0.89 | 933 | 0.92 | 933 | 0.02 | 0.01 | 0.07 | 0.08 |
| 2013 Escuela EIB | 0.08 | 933 | 0.12 | 933 | 0.04 | 0.01 | 0.00 | 0.15 |
| Tasa Pobreza del Distrito en 2009 | 0.35 | 933 | 0.41 | 933 | 0.06 | 0.01 | 0.00 | 0.26 |
| Tasa Pobreza Extrema del Distrito en 2009 | 0.11 | 933 | 0.16 | 933 | 0.04 | 0.01 | 0.00 | 0.28 |
| Tasa de Desnutrición del Distrito en 2009 | 22.27 | 933 | 27.52 | 933 | 5.24 | 0.72 | 0.00 | 0.33 |

Fuentes: ECE, Censo Escolar, Soporte Pedagógico.