



ove

Oficina
de Evaluación
y Supervisión

Documento
de trabajo

OVE/WP-01/12

La pedagogía en ciencias y Medio ambiente: Evidencias experimentales en Perú

Septiembre 2012



ove

Oficina
de Evaluación
y Supervisión

La pedagogía en ciencias y Medio ambiente: Evidencias experimentales en Perú

Diether Beuermann
Emma Naslund-Hadley
Inder J. Ruprah
Jennelle Thompson

Septiembre 2012

© Banco Interamericano de Desarrollo,
www.iadb.org

La información y las opiniones que se presentan en esta publicación son exclusivamente de los autores y no expresan ni implican el aval del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representan.
Este documento puede reproducirse libremente a condición de que se indique que es una publicación del Banco Interamericano de Desarrollo.

La pedagogía en ciencias y Medio ambiente Evidencias experimentales en Perú¹

Resumen

En las sociedades de hoy basadas en el conocimiento, nunca antes fueron tan esenciales la comprensión de conceptos científicos básicos y la capacidad de estructurar y resolver problemas de ciencia. En consecuencia, en este documento evaluamos una metodología innovadora para la enseñanza de ciencias y medio ambiente en escuelas públicas primarias, en las que la enseñanza tradicional (centrada en el docente) se reemplazó por actividades centradas en el alumno, que emplean juegos LEGO. Documentamos mejoras positivas y significativas con desviaciones estándar de 0,18 en calificaciones de pruebas estandarizadas. Dichos resultados positivos están concentrados principalmente en los varones, que estuvieron por encima de la mediana del desempeño académico inicial de referencia.

Palabras clave: ciencias, medio ambiente; Perú

JEL: I21; I28; I29; O15; O31

1) Introducción

La matrícula escolar ha aumentado considerablemente en los últimos 10 años en la mayoría de los países en desarrollo, pero el avance en la calidad de la educación no ha sido tan marcado. La evidencia sugiere que el incremento en la participación escolar no se traduce automáticamente en un aumento de la competencia en las habilidades básicas (Glewwe and Kremer, 2006; Glewwe et al, 2011). Y, como es de suponer, Perú no constituye una excepción. Las tasas de cobertura para la educación inicial, primaria y secundaria han aumentado significativamente durante la última década, y en general se consideran satisfactorias, para un país en desarrollo y de ingresos medios.ⁱ

Sin embargo, en términos de calidad, en las pruebas nacionales, regionales e internacionales, los estudiantes peruanos obtuvieron reiteradamente bajas calificaciones. En 2009, una prueba nacional para alumnos de segundo grado reveló que sólo el 13,8 por

¹ Los autores agradecen al Fondo Japonés para la Reducción de Pobreza del Banco Interamericano de Desarrollo (BID) por haber financiado el piloto.

ciento alcanzó los logros de aprendizaje esperados en matemática.ⁱⁱ En el Segundo Estudio Regional Comparativo y Explicativo (SERCE), los estudiantes de Perú se desempeñaron por debajo del promedio regional, tanto en matemática como en ciencias naturales (UNESCO, 2008). En matemática de tercer grado, más de la mitad de la población estudiantil alcanzó sólo un desempeño mínimo. Otro resultado del estudio SERCE fue que el promedio de las calificaciones de los estudiantes del sector urbano era tres veces superior al del alumnado del sector rural. En las pruebas del Programa para la Evaluación Internacional de Alumnos (PISA), entre 65 países participantes, Perú quedó clasificado en el puesto 60 en matemáticas y en el 63 en ciencias naturales (OECD, 2011). Esta evaluación de alumnos que están próximos a culminar el ciclo secundario (alrededor de los 15 años de edad) y cuyo principal objetivo es determinar el grado de adquisición de las habilidades necesarias para participar plenamente en la sociedad del conocimiento sugiere que la mayoría de los estudiantes peruanos no están preparados para ingresar al mercado laboral ni para iniciar estudios terciarios. Por ende, determinar cómo mejorar la calidad educativa es una cuestión de medidas políticas crítica para Perú. Existe un caudal creciente de investigación que respalda el cambio de la educación tradicional liderada por el docente hacia un aprendizaje centrado en el alumno complementado con cierto nivel de indagación, como forma de maximizar la adquisición de conocimientos (Healy, 1990; Lowery, 1998). Recientemente, por medio de un análisis de 37 estudios experimentales y no experimentales de instrucción basada en la indagación, Furtak et al (2012) encuentran que metodologías dirigidas por docentes en base a la indagación fueron relativamente mas efectivas que enfoques centrados totalmente en los alumnos con baja indagación. En ese sentido, se requieren

investigaciones adicionales para definir cuál es el grado de indagación que resulta más efectivo para los distintos cursos y contenidos. Además, el número de evaluaciones rigurosas de los distintos enfoques para la enseñanza de ciencias naturales es particularmente escaso. En vista de estos antecedentes, en 2008 el Ministerio de Educación del Perú solicitó la asistencia del Banco Interamericano de Desarrollo para validar un enfoque pedagógico centrado en el alumno complementado con indagación, para la educación de ciencias y medio ambiente. Esta asociación dio como resultado el desarrollo de un método de indagación guiada (Colburn, 2000), para el aula de ciencias naturales, con énfasis en actividades centradas en el alumno bajo la guía del docente.

En lugar de enseñar simplemente a memorizar la historia de la ciencia y los hechos científicos (como se hace tradicionalmente), este nuevo enfoque se concentra en el desarrollo del pensamiento científico y en la comprensión de lo que el alumno puede lograr con ese conocimiento. La metodología se basa en la curiosidad de los niños y en su propensión natural a explorar el mundo que los rodea. Este nuevo paradigma educativo se aplica a tres módulos: nuestro ambiente, el cuerpo humano y nuestro mundo físico, y se implementó de manera piloto en clases de tercer grado de 53 escuelas (grupo de tratamiento), del departamento de Lima, y se evaluó mediante un diseño experimental (53 escuelas, como grupo de control, que continuaron con el enfoque pedagógico tradicional). El tratamiento consistió en preparar material didáctico y módulos de capacitación docente y proporcionar apoyo en el aula, así como instrumentos de evaluación continua de los estudiantes. En términos de equipamiento, se proporcionó instrumental de laboratorio escolar, además de materiales de aprendizaje LEGO, para complementar los experimentos dentro y fuera del aula. Si bien ya se había evaluado el empleo de LEGO, la

mayoría de los estudios anteriores consistía en muestreos reducidos por periodos de estudio limitados, que se caracterizan principalmente por ser evaluaciones cualitativas no suficientemente rigurosas. No obstante, esos estudios cualitativos determinaron que el empleo de estos materiales aumenta el pensamiento crítico, desarrolla la habilidad de solución de problemas y mejora la colaboración entre estudiantes (Noble, 2001). En Perú, Iturrizaga (2000) observó, a partir de datos de los establecimientos educativos de tratamiento y de control, que los puntajes en los exámenes mejoraban considerablemente en matemática, lectura, tecnología y coordinación viso-motriz, aunque los grupos no fueron seleccionados en forma aleatoria. Además, Hussain et al. (2006), concluyen, a partir de datos de Suecia, que los estudiantes de quinto grado que participaron en este enfoque pedagógico se desempeñaron mejor en matemática. El alcance limitado de estos estudios y la debilidad de su metodología sugieren que se necesitan más investigaciones sobre el impacto causal en el aprendizaje estudiantil.

En este documento, informamos sobre los resultados de una evaluación diseñada en forma experimental para un programa piloto en Perú, que se basan en evaluaciones estandarizadas tomadas a los grupos de tratamiento y de control, así como también en encuestas realizadas a directores de escuela, docentes, estudiantes y padres. El presente informe se estructura de la siguiente manera: La sección 2 describe los antecedentes del programa, las áreas de aprendizaje comprendidas y el contexto en el cual se implementó. La sección 3 presenta la estrategia de investigación y su puesta en práctica, y la calidad de los datos. La sección 4 presenta los resultados y su interpretación. Por último, la sección 5 presenta las conclusiones.

2) El programa

2.1. Antecedentes

Hasta mediados del siglo XX, el diseño curricular para ciencias naturales entendía que al ingresar al sistema educativo el niño era una pizarra en blanco sin nociones o creencias acerca de diferentes hechos y fenómenos. En consecuencia, al docente se le exigía que transfiriera a los alumnos su conocimiento de los conceptos científicos. Este estilo de enseñanza frontal posteriormente fue remplazado en la literatura por una diversidad de enfoques de aprendizaje activo, entre otros, las dramatizaciones, los debates y los grupos de aprendizaje colaborativo.

En América Latina, la discusión entre métodos centrados en el docente o aprendizaje centrado en el alumno aún persiste tanto a nivel ideológico como filosófico. Si bien algunos sistemas educativos han actualizado sus planes de aprendizaje, en la práctica, el estilo de enseñanza con el docente delante de la clase sigue predominando en toda la región. Perú no constituye una excepción, y si bien el programa de estudios de ciencias para los distintos niveles enfatiza el aprendizaje basado en la indagación, los estilos de enseñanza tradicionales de pizarra y tiza y clases expositivas siguen predominando y priorizando la memorización de los conceptos científicos y de la historia de la ciencia, en lugar del desarrollo de habilidades de indagación y pensamiento crítico.

En tercer grado, el programa tradicional de ciencia y medio ambiente abarca tres áreas temáticas. Primero, el área sobre el "cuerpo humano" que presenta las estructuras, funciones e interacciones de los distintos sistemas. En segundo lugar, "nuestro ambiente" que introduce a los estudiantes a las explicaciones científicas acerca de cómo funcionan los ecosistemas. En términos de contenido, comprende tres temas: (i) ecosistemas, que incluye el desarrollo y

crecimiento de las plantas y de los seres vivos (organismos, relación entre animales y plantas, animales vertebrados e invertebrados); (ii) biodiversidad, que abarca los animales y plantas nativos y exóticos; y (iii) la protección de plantas, animales y habitats. En tercer lugar, un módulo denominado "nuestro mundo físico" comprende temas tales como el planeta Tierra y sus características, fuerzas y movimientos, electricidad, luz y color y magnetismo (Ministerio de Educación, 2008).

2.2. La intervención

La intervención desarrolló metodologías centradas en el alumno complementadas con indagación, para enseñar las mismas áreas y los mismos temas comprendidos en el enfoque tradicional liderado por el docente. Bajo el protocolo de indagación desarrollado, el docente reta a los alumnos proveyéndoles con un problema a resolver así como los materiales necesarios. Los estudiantes deben seguidamente elaborar sus propios procedimientos, registrar y reportar los resultados. El maestro facilita el proceso de aprendizaje por medio de la motivación de sus estudiantes y ayudándolos a explorar nuevas ideas así como formular preguntas interesantes. Durante las conversaciones, el maestro presenta los nombres formales de los distintos conceptos. Luego los estudiantes aplican los conceptos a nuevas situaciones.

El diseño aprovechó diversos paradigmas de aprendizaje activo (Bonwell and Eison, 1991) que enfatizan la importancia de comprender las ciencias, en lugar de memorizar conceptos aislados. Entre otros, la construcción de nuevos conocimientos a partir de lo que los estudiantes ya saben; el desarrollo de la capacidad de pensamiento crítico; la enseñanza a través de la indagación y el abordaje de distintos estilos de aprendizaje de los alumnos.

Dentro del área "cuerpo humano", el modelo estaba dirigido a estimular la curiosidad y a la deducción de conclusiones lógicas a partir de los temas tratados. En "nuestro ambiente", se introdujo a los alumnos al trabajo de campo, ayudándolos a generar sus propias explicaciones científicas acerca de cómo funcionan los ecosistemas, sobre la base de la observación empírica, en lugar de su simple presentación en clase. En el tópico "mundo físico", el programa se diseñó para que los estudiantes adquirieran más competencias en la formulación atenta de preguntas de investigación, en la realización de experimentos y en la interpretación de los datos relacionados con los temas tratados.

Para lograrlo, se desarrollaron diversos materiales para comprender las pautas curriculares de los tres módulos. Como por ejemplo, planes de clase, diarios de actividades y equipos aúlicos simples con microscopios, lupas, jarras graduadas, balanzas y consumibles diversos para enseñar los módulos "ambiente" y "cuerpo humano", además de los juegos educativos LEGO para enseñar el módulo "mundo físico".

Los importantes vacíos pedagógicos y de contenido de los docentes peruanos constituyeron un desafío en la tarea de llevar con éxito el nuevo programa de estudios basado en la indagación, a las aulas de ciencias. En consecuencia, una piedra angular de este programa fue ayudar a que los docentes desarrollen los conocimientos adecuados para los contenidos de ciencias de tercer grado y a que aprendan a permitir aprendizaje del estudiante a través de la indagación. La capacitación para los maestros comprendió dos tipos de actividades: (i) talleres interactivos para desarrollar conocimientos en los contenidos curriculares, sugerir una diversidad de métodos que faciliten y promuevan las habilidades de razonamiento de los estudiantes, y ayudar a los maestros a desarrollar actividades aúlicas que permitan la participación de los alumnos en las investigaciones científicas; y (ii) asesoramiento técnico y

tutoría pedagógica dentro y fuera del aula, por ejemplo a través de sesiones demostrativas, seguimiento individual y clases compartidas entre colegas.

Además del desarrollo de conocimientos pedagógicos y sustantivos, los talleres y la asistencia personalizada hicieron hincapié en el desarrollo de evaluaciones formativas, para ayudar a los docentes a crear planes de aprendizaje individuales para los estudiantes. Todas las sesiones de capacitación y tutoría de docentes se diseñaron para que fueran lo más concretas y de aplicación práctica posible, priorizando las actividades paso a paso, por encima de las conversaciones filosóficas acerca de las definiciones del aprendizaje basado en la indagación.

2.3. El contexto

Para comprender el contexto en el cual se desarrollará el programa, nos basamos en los datos recopilados por estudios sociodemográficos propios. Recopilamos esta información en las escuelas participantes del programa piloto y en las familias de los estudiantes, a través de encuestas.

Casi un tercio de los estudiantes relevados informó que trabajan en explotaciones agrícolas o micro-emprendimientos familiares locales, en algún momento durante el año escolar. El porcentaje de alumnos que indicó que vende productos en las calles es del orden del 8 al 10 por ciento, según la época del año de que se trate. Sorprendentemente, de acuerdo con las respuestas, hay más estudiantes procedentes de zonas rurales que urbanas que trabajan en la calle, incluso en el área metropolitana de Lima.

Por ejemplo, el 11 por ciento de los alumnos de zonas rurales informó que vende productos en las calles los fines de semanas durante el período escolar, mientras que de

los estudiantes urbanos sólo lo hace un 6 por ciento. La encuesta a los padres mostró que el 46 por ciento de los jefes de hogar no había terminado la educación secundaria y que sólo un 19 por ciento ha tenido algún tipo de educación terciaria. Los padres relevados informan que sus ingresos promedio son de S/. 500.00 (casi US\$180) por mes; es decir menos que el salario mínimo vigente al momento del estudio.ⁱⁱⁱ Si bien existen diversas razones que hacen dudar de la cifra exacta de ingresos, queda en claro que la mayoría de los estudiantes provienen de familias de bajos ingresos.

3) Estrategia de investigación: diseño, implementación y datos

3.1. Marco conceptual

El marco conceptual típico que se invocó para el análisis de la educación es el de la metodología de función de producción, cuya especificación convencional para una función lineal variable en el tiempo es:^{iv}

$$A_t = \beta_0 + \beta_{1t} \cdot S_t + Q_t' \beta_{qt} + C_t' \beta_{ct} + H_t' \beta_{ht} + I_t' \beta_{It} + \mu_{at} \quad (1)$$

Donde A_t son las destrezas aprendidas a lo largo del tiempo t , S_t es años de escolarización adquirida por tiempo t , Q_t es un vector de características de la escuela y de docentes, C_t es un vector de características del alumno, H_t es un vector de las características del hogar, I_t es un vector de entradas escolares relacionadas con control parental, y μ_{at} es un término de error, y donde μ_{at} responde a variables para las cuales no hay datos y o existen errores de medición en las destrezas adquiridas o en las variables explicativas. El impacto causal de una determinada variable explicativa de las destrezas puede estimarse en forma consistente sólo si μ_{at} no se correlaciona con dicha variable explicativa. Es improbable

que esta situación perturbe los datos observados, entre otros factores, por sesgos de especificación, de selección y desgaste muestral o en la medición.

Para superar estos factores y recuperar el efecto causal del tratamiento, utilizamos datos generados a partir de una intervención randomizada, que consistió en la introducción de un ligero cambio en un conjunto de escuelas seleccionadas al azar, y la ausencia de dicho cambio en otras escuelas también elegidas en forma aleatoria. Las pruebas randomizadas correctamente definidas pueden superar los problemas mencionados precedentemente, y revelar los efectos causales a través de la diferencia en el resultado de interés, que en nuestro caso son las calificaciones entre los grupos de tratamiento y de control. La mejor manera de emplear experimentos para informar acerca de medidas políticas es evaluar más las medidas de política, tal como sostienen Glewwe et al. (2011); esta metodología resulta muy prometedora cuando la teoría brinda escasas pautas, como por ejemplo en cuanto a qué tipos de materiales pedagógicos son los más efectivos.

En consecuencia, nos concentramos en proporcionar pruebas consistentes respecto de la efectividad de una metodología de enseñanza, y no en estimar una forma particular de función de producción educativa, o en determinar la ponderación relativa de distintos componentes de esa misma función. En síntesis, aceptamos en forma endógena un dato de entrada determinado con el fin de evaluar sus efectos en el rendimiento académico, al mismo tiempo que se mantienen constantes todos los otros datos de entrada observables y no observables que podrían estar presentes en la función de producción educativa desconocida.^v A tales fines, una intervención randomizada constituye un enfoque confiable. La literatura sobre evaluación de programas educativos ha utilizado esta metodología en forma extensa con el fin de analizar los efectos en el aprendizaje de

distintos elementos educativos tales como el uso de computadoras (Barrera-Orsorio and Linden, 2009; Malamud and Pop-Eleches, 2011; Cristia et al, 2012), tamaño y seguimiento de la clase (Duflo et al, 2011), y programas informáticos para educación (Banerjee et al, 2007; Carrillo et al, 2010).

3.2. Diseño

La estrategia de la investigación consta de tres componentes: Primero, un diseño experimental de evaluación para evaluar el impacto causal del nuevo enfoque pedagógico en los logros académicos de los estudiantes. Segundo, encuestas a directores, docentes, padres y alumnos para obtener información socio-demográfica de las escuelas y el entorno familiar. Tercero, una evaluación cualitativa con el fin de comprender más detalladamente el contexto del tratamiento y la evaluación.^{vi}

El diseño experimental abarcó 106 establecimientos del Departamento de Lima, con una asignación aleatoria, es decir 53 escuelas de tratamiento y 53 sin tratamiento, con un total de 2771 alumnos de tercer grado en total. La muestra fue estratificada de acuerdo con la ubicación establecimiento escolar (urbana, metropolitana y rural). Por razones presupuestarias, se incluyeron sólo dos aulas por escuela. En aquellos establecimientos donde había más de dos aulas para ese grado, se seleccionaron al azar sólo dos, en presencia del director. Los estudios realizados con anterioridad sugieren que existe una alta correlación entre aulas en una misma escuela. Por ende, se pierde muy poca información adicional al incluir únicamente dos aulas por escuela.^{vii} Se tomaron exámenes de referencia para ciencia y medio ambiente, y también en matemática y

comprensión lectora, al inicio y nuevamente al final del tercer grado de ese año (ciclo lectivo 2010 de abril a diciembre).^{viii}

Se realizaron encuestas a directores, docentes y padres. El cuestionario para los directores de escuela estuvo diseñado para obtener información sobre el número de alumnos y docentes, las instalaciones, equipamiento y materiales didácticos, y el entorno escolar. La consulta entre docentes recopiló datos de referencia sobre distintos aspectos de la enseñanza de las ciencias y del medio ambiente en la escuela, y de sus alumnos. Los alumnos recibieron una encuesta socio-demográfica, que llevaron a sus casas en sobres cerrados para que fueran contestadas por sus padres y devueltas.

Las herramientas de investigación cuantitativa fueron complementadas con una evaluación cualitativa, que consistió en visitas in situ a ocho escuelas, cuatro del grupo de tratamiento y cuatro del grupo de control. Estos establecimientos fueron elegidos de modo tal que abarcasen todos los estratos del estudio: rural, urbano, metropolitano, y escuelas pequeñas y medianas. La visita se realizó entre agosto y septiembre de 2010. Este análisis adicional incluyó entrevistas a docentes, directores y alumnos del tercer grado. El objetivo era transmitir información adicional sobre el contexto del experimento, cómo podría afectarlo, y consecuentemente, los factores que contribuían a sus éxitos y fracasos. Utilizamos las observaciones recopiladas durante esta etapa al momento de analizar los resultados cuantitativos, a fin de interpretarlos.

3.3. Implementación

La implementación del experimento se desvió de su plan inicial de varias maneras. Primero, estaba previsto que se administraría durante todo el año académico 2010 (de

abril a diciembre); No obstante, las demoras en la distribución de los materiales y en las sesiones de capacitación determinaron que el tratamiento se aplicara efectivamente sólo durante cinco meses, antes de evaluar a los alumnos (de julio a noviembre).

En segundo lugar, no fue posible distribuir los materiales educativos para los módulos "cuerpo humano" y "medio ambiente" debido a complicaciones logísticas en los procesos de importación. Los juegos educativos LEGO para el módulo "mundo físico" fueron los únicos que se utilizaron realmente durante los cinco meses del período de implementación. Por último, el tiempo de capacitación de los docentes rurales fue inferior al de sus pares en los establecimientos escolares urbanos (20 horas en comparación con 60 horas). Esto se debe a que el sistema de apoyo docente impartido por la web no estaba disponible en las escuelas rurales, dado que carecen de conexión a Internet. Además, en las zonas rurales la capacitación comenzó después, y por ese motivo esas escuelas contaron sólo con dos meses de implementación tras la capacitación. La suma de todas estas situaciones implicó que el tratamiento durase menos de lo planeado y que su intensidad y duración fueran incluso inferiores en las escuelas rurales que en las urbanas.

3.4. Datos

La asignación aleatoria ejecutada adecuadamente debe eliminar, en promedio, todos los factores de confusión posibles, tanto observables como no observables. La Tabla 1 presenta el resumen estadístico de las características físicas de los establecimientos escolares y los resultados de las evaluaciones al inicio, según el estrato del tratamiento.

Tabla 1: Diferencias iniciales entre grupos de tratamiento y de control

	Tratamiento (1)	Control (2)	Diferencia (3)	Observaciones (4)
Panel A: Características de la escuela				
Cantidad de docentes - Primario	12,31 (1,60)	10,84 (1,29)	1,47 (2,05)	106
Con agua corriente	0,73 (0,06)	0,70 (0,07)	0,03 (0,09)	106
Tiene teléfono	0,48 (0,07)	0,60 (0,07)	-0,12 (0,10)	106
Con conexión a Internet	0,31 (0,07)	0,22 (0,06)	0,09 (0,09)	106
Lugar para sembrar plantas	0,94 (0,03)	0,98 (0,02)	-0,04 (0,04)	106
Sala de informática	0,87 (0,05)	0,86 (0,05)	0,01 (0,07)	106
Laboratorio de ciencias	0,85 (0,05)	0,96 (0,03)	-0,11* (0,06)	106
Sala de actividades artísticas	0,91 (0,04)	0,98 (0,02)	-0,08 (0,05)	106
Sala de música	0,91 (0,04)	0,98 (0,02)	-0,08 (0,05)	106
Panel B: Desempeño (todas las escuelas)				
Ciencias y medio ambiente	0,12 (0,08)	0,00 (0,08)	0,12 (0,11)	2771
Habilidad oral	0,12 (0,06)	0,00 (0,07)	0,12 (0,09)	2771
Habilidad matemática	0,07 (0,06)	0,00 (0,06)	0,07 (0,09)	2771
Panel C: Desempeño (escuelas con 1 o 2 secciones)				
Ciencias y medio ambiente	-0,06 (0,11)	-0,06 (0,08)	0,00 (0,13)	1487
Habilidad oral	-0,04 (0,09)	-0,10 (0,09)	0,06 (0,13)	1487
Habilidad matemática	-0,04 (0,08)	-0,11 (0,09)	0,07 (0,12)	1487
Panel D: Desempeño (escuelas con 3 o más secciones)				
Ciencias y medio ambiente	0,32 (0,09)	0,07 (0,14)	0,25 (0,17)	1284
Habilidad oral	0,30 (0,05)	0,12 (0,08)	0,18* (0,09)	1284
Habilidad matemática	0,19 (0,09)	0,13 (0,08)	0,06 (0,12)	1284
Errores estándar estimados agrupados por escuela entre paréntesis. * Significativo en 10 %; ** significativo en 5 %; *** significativo en 1%.				

No existen diferencias significativas en la infraestructura de los establecimientos escolares entre los grupos de tratamiento y de control (Panel A). Del mismo modo, se observa la ausencia de diferencias estadísticas entre ambos grupos en las calificaciones que obtuvieron los estudiantes en las pruebas (Panel B).^{ix} Dado que en las escuelas que cuentan con tres o más divisiones de tercer grado, la estrategia de nuestra investigación muestreó solamente dos secciones aleatorias, es importante demostrar que tal procedimiento no contaminó ni sesgó potencialmente nuestro diseño. Por consiguiente, comparamos las escuelas de tratamiento y de control con respecto a los grados iniciales de referencia, según la cantidad de secciones de tercer grado. El panel C muestra que no hubo diferencias significativas entre los establecimientos escolares de tratamiento y de control con una o dos secciones. Del mismo modo, el panel D indica que no se observaron diferencias iniciales al considerar las escuelas con tres o más secciones. De esta manera, podemos utilizar con seguridad la muestra sin preocupaciones con respecto a posibles sesgos que pudieran haberse introducido por incluir muestras no aleatorias diferenciales en las escuelas con tres o más secciones.

Asimismo, las calificaciones de las pruebas iniciales proporcionan información importante para evaluar posibles impactos heterogéneos en la distribución de puntajes. De este modo, se plantea una pregunta evaluativa importante acerca de si el nuevo modelo pedagógico puede ayudar a cerrar las brechas de aprendizaje. Con el fin de responder a este interrogante, estimaremos los efectos de la intervención para los subgrupos. La posible heterogeneidad de los impactos, y si acentúan o atenúan las desigualdades originales en los puntajes, resulta crítica en el diseño de medidas políticas con respecto a una potencial uniformidad o diferenciación en la ampliación del programa.

El desgaste relativamente alto es un problema latente para la evaluación. La muestra final en número de alumnos que rindieron las pruebas descendió un 14,4%, o sea 2.373 estudiantes de los 2.771 en la prueba inicial de referencia.^x Este desgaste de la muestra original podría producir un sesgo si los estudiantes que abandonaron son sistemáticamente diferentes de los que permanecieron. El sesgo de desgaste debilitaría la validez interna del estudio, y la reducción de la muestra podría disminuir el poder estadístico de los exámenes. La atribución compromete la validez interna del diseño experimental, porque los miembros de los grupos de tratamiento y de control para los cuales se cuenta con datos de seguimiento pueden ser sub-muestras no aleatorias de los grupos originales. En consecuencia, presentamos a continuación los resultados obtenidos al evaluar si la intervención ocasionó retiros, si el desgaste fue ortogonal a los resultados de interés y si el poder estadístico de la muestra fue suficiente.

Con el propósito de examinar si el desgaste no estuvo relacionado sistemáticamente con el tratamiento, se estimó la siguiente regresión:

(2)

Donde L_{ij} es igual a uno si el estudiante i en la escuela j no fue evaluado al final; si no, tiene valor cero (“Alumno que ha terminado sus estudios”); T_j es igual a uno si la escuela j participó en el tratamiento; si no, tiene valor cero. El parámetro β será estadísticamente indistinguible de cero si las tasas de desgaste no fueron sistemáticamente diferentes entre ambos grupos.

La columna uno de la Tabla 2 indica el estimado de β a partir de la ecuación (2). El coeficiente estimado para β es estadísticamente indistinguible de cero, demostrando que las tasas de desgaste en los dos grupos no fueron sistemáticamente diferentes.

exactamente iguales (desviaciones estándar de 0,19), a pesar de la reducción del tamaño de la muestra de alumnos. Esto se explica porque la asignación del tratamiento se realizó a nivel de los establecimientos escolares y el poder del experimento está, por consiguiente, impulsado principalmente por el número de establecimientos participantes (que fue el mismo al inicio y al final).

4) Resultados e interpretación

4.1. Efectos generales

El impacto directo esperado a partir del programa es una mejora en las calificaciones de los exámenes de ciencia y ambiente. Con el propósito de medir el impacto del programa en los puntajes de las pruebas, se estimó la siguiente regresión:

$$Y_{ij,t} = \delta + \delta_1 \cdot Y_{ij,t-1} + \beta \cdot T_j + \varepsilon_{ij,t} \quad (4)$$

Donde $Y_{ij,t}$ es el resultado estandarizado del examen, al final, para el estudiante i de la escuela j ; e $Y_{ij,t-1}$ es la calificación al inicio.^{xiii} El parámetro β mide el impacto del programa (expresado en términos de desviaciones estándar con respecto al grupo de control) en los resultados de las pruebas. En la Tabla 3 se presentan los resultados de la regresión estimada.

Tabla 3: Resultados generales - Ciencias y medio ambiente

Variables dependientes:	Ciencias y medio ambiente							
	Cuerpo humano		Medio ambiente		Mundo físico		Generales	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Efectos generales								
Efecto del tratamiento	0,02 (0,09)	-0,03 (0,06)	0,08 (0,10)	0,03 (0,07)	0,22** (0,10)	0,18** (0,08)	0,15 (0,11)	0,10 (0,07)
Observaciones	2536	2373	2536	2373	2536	2373	2536	2373
Panel B: Efectos por género								
Niñas	-0,06 (0,10)	-0,06 (0,07)	-0,03 (0,11)	-0,03 (0,07)	0,10 (0,11)	0,10 (0,09)	0,02 (0,12)	0,02 (0,08)
Efecto adicional en varones	0,14 (0,09)	0,06 (0,08)	0,20** (0,08)	0,11 (0,07)	0,24*** (0,08)	0,16** (0,08)	0,25*** (0,09)	0,15** (0,07)
Observaciones	2536	2373	2536	2373	2536	2373	2536	2373
Panel C: Efectos geográficos								
Rural	-0,23 (0,14)	-0,18 (0,12)	-0,20 (0,18)	-0,16 (0,13)	0,02 (0,16)	0,09 (0,13)	-0,14 (0,18)	-0,08 (0,13)
Urbano	0,07 (0,10)	0,01 (0,07)	0,16 (0,12)	0,10 (0,07)	0,26** (0,12)	0,20** (0,09)	0,21 (0,13)	0,14* (0,08)
Observaciones	2536	2373	2536	2373	2536	2373	2536	2373
Panel D: Efectos geográficos por género								
Rural - niñas	-0,29 (0,18)	-0,21 (0,14)	-0,30 (0,19)	0,23* (0,13)	-0,10 (0,18)	-0,03 (0,15)	-0,26 (0,20)	-0,18 (0,14)
Rural - varones	-0,17 (0,16)	-0,15 (0,15)	-0,10 (0,20)	-0,09 (0,17)	0,14 (0,17)	0,21 (0,14)	-0,02 (0,20)	0,02 (0,17)
Urbano - niñas	-0,00 (0,11)	-0,02 (0,07)	0,06 (0,12)	0,06 (0,08)	0,15 (0,14)	0,14 (0,11)	0,10 (0,14)	0,08 (0,09)
Urbano - varones	0,13 (0,11)	0,04 (0,09)	0,25* (0,14)	0,15 (0,09)	0,36*** (0,13)	0,26** (0,10)	0,32** (0,14)	0,20** (0,10)
Observaciones	2536	2373	2536	2373	2536	2373	2536	2373
Controles de referencia iniciales	No	Si	No	Si	No	Si	No	Si

Errores estándar estimados agrupados por escuela entre paréntesis. * Significativo en 10 %; ** significativo en 5 %; *** significativo en 1%. Todas las calificaciones se expresan como desviaciones estándar respecto del grupo de control. Las estimaciones en las columnas uno, tres, cinco y siete se obtienen se establecieron a partir del modelo sin control de las calificaciones iniciales. En consecuencia, utilizamos la totalidad de las observaciones, 2536, en la línea final, incluyendo a los alumnos que no fueron evaluados en la prueba de inicio. Las estimaciones en las columnas dos, cuatro, seis y ocho están controladas por los resultados iniciales de cada alumno, en consecuencia incluyen alumnos evaluados en las pruebas iniciales y en las finales, es decir 2373 estudiantes.

Tal como se demuestra en el Panel A de la Tabla 3, no existen impactos en los módulos “cuerpo humano” y “medio ambiente” (Columnas 1 a 4) y los puntajes generales de las evaluaciones (Columnas 7 y 8). No obstante, se registra un efecto importante en el módulo “mundo físico”, con un impacto estimado de desviaciones estándar de 0,18 (Columna 6). Estos resultados expresan que este nuevo método pedagógico es más eficaz que los empleados tradicionalmente en las escuelas peruanas, a pesar de que la nueva metodología se aplicó menos tiempo que lo planeado. Además, el tamaño del efecto es mayor que el observado en la mayoría de los estudios sobre intervenciones destinadas a mejorar los puntajes de las evaluaciones a nivel escolar (ver Glewwe et al., 2011).

No sorprende encontrar efectos solamente en la sección del “mundo físico”, dado que los materiales didácticos para los demás módulos no se utilizaron de manera eficaz. Además, en esa sección efectivamente se hizo uso de los juegos educativos LEGO. Asimismo, la evidencia cualitativa corrobora que el programa despertó el interés de los alumnos con relación a este módulo. En efecto, los docentes destacaron que la asistencia siempre fue un problema en las escuelas. No obstante, cuando al inicio del programa se anunciaba con anterioridad en qué días se utilizarían los juegos educativos LEGO, los alumnos que registraban ausencias frecuentes comenzaron a asistir con regularidad en esas fechas. Luego, los docentes dejaron de anunciar con anticipación la utilización de los juegos LEGO, produciéndose un aumento de la asistencia de todos los días. Los docentes informaron también que los niños comenzaron a identificar las máquinas que sus padres usan para trabajar y a comprender conceptos científicos tales como la polea.

Es importante determinar si se registraron impactos diferenciales para distintos segmentos de la población de estudio. Dada la ausencia de cualquier heterogeneidad en

los efectos, es posible ampliar el módulo “mundo físico” de manera uniforme, pero si se registran efectos heterogéneos, toda ampliación requerirá modificaciones al enfoque pedagógico empleado para los distintos sub-grupos. Consideramos los impactos diferenciales posibles por género, ubicación geográfica de la escuela y desempeño inicial de referencia. Además, pretendemos determinar si hubo efectos de derrame del tratamiento en “matemáticas” y “comprensión lectora”.

4.2. Impactos por género

Un importante aspecto es determinar si la intervención tuvo impactos diferentes por género. Los resultados pueden verse en la Tabla 3 – Panel B. Para obtener los efectos diferenciales por género se estimó la siguiente regresión:

$$Y_{ij,t} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Male_{ij} + \beta_1 \cdot T_j + \beta_2 \cdot T_j \cdot Male_{ij} + \varepsilon_{ij,t} \quad (6)$$

Donde la variable categórica $Varón_{ij}$ equivale a uno si el alumno i en la escuela j es varón. El parámetro β_1 captura los impactos del programa en las niñas mientras que β_2 capta los impactos adicionales en los varones con respecto a las niñas.

De acuerdo con los resultados anteriores, los efectos son nulos en los sub-temas “cuerpo humano” y “medio ambiente”. Sin embargo, los impactos en el módulo “mundo físico” tienen efecto exclusivamente en los niños (Columnas 5 y 6), es decir, un impacto adicional con desviaciones estándar de 0,24 a 0,16, en tanto para las niñas el impacto es estadísticamente indistinguible de cero. Más aún, el efecto del programa en los puntajes generales en los exámenes de “ciencias” y “medio ambiente” es significativamente mayor que en las niñas con desviaciones estándar de 0,15, en tanto no se presenta impacto en las calificaciones generales de las niñas (Columna 8).

Por ende, el tratamiento acentuó la inequidad de género. Muchos posibles factores puede haber contribuido a producir este efecto diferencial. La variación entre géneros puede ser el resultado de diferencias en la atracción de las áreas de contenido específicas dentro de los tres módulos, incluidos los problemas concretos de ciencias que los alumnos debieron resolver. Dada la limitada cantidad de equipos de ciencias y juegos educativos LEGO por clase (solamente uno para compartir dentro del aula), es posible que los varones los monopolizaran. Si bien no se registró sistemáticamente datos acerca del uso de los materiales didácticos, las visitas al aula durante las entrevistas cualitativas sugieren que, de hecho, hubo monopolización por parte de los varones.

4.3. Impactos por ubicación geográfica

Además, los datos permiten determinar si existen los impactos son diferentes por ubicación geográfica de la escuela. La asignación aleatoria se estratificó geográficamente de modo tal que de los 53 establecimientos de tratamiento (control), 29 (28) se encontraban en áreas urbanas y 24 (25) en zonas rurales. El Panel C de la Tabla 3 muestra los impactos estimados por ubicación geográfica.

Este panel indica por separado las escuelas rurales y las urbanas. La Columna 6 revela un efecto significativo en las áreas urbanas equivalente a desviaciones estándar de 0,2, pero ningún efecto en las zonas rurales para el módulo “mundo físico”. No percibimos efectos significativos en ninguno de los restantes módulos considerados de manera individual, pero sí existe un cierto efecto en el puntaje total con desviaciones estándar de 0,14 (Columna 8). Estos efectos diferenciales reflejan el hecho de que el programa piloto tuvo una aplicación más intensiva en las áreas urbanas. La capacitación de los docentes rurales

fue menor que la de los docentes urbanos (20 y 60 horas, respectivamente). Además, la capacitación de los maestros rurales se implementó después que la de sus colegas urbanos de modo tal que el período efectivo de implementación final del nuevo método de enseñanza en las áreas rurales fue de menos de dos meses antes del examen.

Por lo tanto, dado que el programa se implementó mejor en las áreas urbanas, es interesante evaluar si sus efectos variaron por género dentro de estas áreas geográficas. Por ejemplo, el Panel D presenta los efectos diferentes. En primer lugar, no vemos efectos diferentes en las áreas rurales ya que prácticamente todos los impactos estimados son estadísticamente casi iguales a cero. En las áreas urbanas no vemos ningún efecto en las niñas. Sin embargo, los efectos en los niños parecen ser significativos para “mundo físico” y el puntaje general (Columnas 5 a 8). El efecto general para los exámenes asciende a desviaciones estándar de 0,2 (Columna 8). Este resultado proporciona evidencia de que el programa, implementado con adecuada capacitación para los docentes (60 horas) y al menos 5 meses de aplicación, arroja resultados positivos de aprendizaje en el desempeño general para el segmento que aparentemente monopoliza los juegos educativos LEGO (es decir, los varones).

4.4. Distribución de los impactos

La aplicación puede ampliar o ajustar la distribución de los puntajes en los exámenes. El Panel A de la Tabla 4 muestra impactos diferenciales por cuartiles de resultados en los exámenes iniciales de referencia. Es decir, los alumnos fueron clasificados de acuerdo con sus calificaciones iniciales y la distribución se categorizó en cuartiles. Se estimaron las siguientes regresiones:

$$Y_{ij,t} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Q1_{ij,t-1} + \delta_3 \cdot Q2_{ij,t-1} + \delta_4 \cdot Q3_{ij,t-1} + \beta_1 \cdot T_j \cdot Q1_{ij,t-1} + \beta_2 \cdot T_j \cdot Q2_{ij,t-1} + \beta_3 \cdot T_j \cdot Q3_{ij,t-1} + \beta_4 \cdot T_j \cdot Q4_{ij,t-1} + \varepsilon_{ij,t} \quad (5)$$

Donde $Q1_{ij,t-1}$, equivale a uno si el alumno i en la escuela j se encontraba entre los centiles 1 y 25 de la distribución de resultados de los exámenes de referencia; si no, tiene un valor igual a cero; $Q2_{ij,t-1}$ equivale a uno si el alumno i en la escuela j se ubicaba entre el centil 26 y 50 en los resultados iniciales, o si no, tiene valor igual a cero; $Q3_{ij,t-1}$ equivale a uno si el alumno i en la escuela j se encontraba entre el centil 51 y 75 de la distribución inicial, y si no, tiene valor cero; $Q4_{ij,t-1}$ equivale a uno si el alumno i en la escuela j se hallaba por encima del centil 75, o de otro modo, tiene el valor igual a cero. El resto de las variables se definen del mismo modo. En este entorno, las estimaciones de los parámetros β_1 , β_2 , β_3 y β_4 representan los impactos del programa dentro de cada cuartil de la distribución inicial.

Tabla 4: Resultados según puntaje en pruebas de línea base – Totalidad de la muestra

Variables dependientes:	Ciencias y medio ambiente			
	Cuerpo humano	Medio ambiente	Mundo Físico	Generales
	(1)	(2)	(3)	(4)
Panel A: Efectos por cuartil ó Totalidad de la muestra				
Cuartilo 1	-0,09 (0,09)	0,04 (0,08)	0,09 (0,10)	0,03 (0,09)
Cuartilo 2	-0,06 (0,11)	0,09 (0,08)	0,11 (0,10)	0,07 (0,10)
Cuartilo 3	0,00 (0,08)	-0,09 (0,09)	0,27*** (0,10)	0,11 (0,09)
Cuartilo 4	0,01 (0,06)	0,03 (0,08)	0,25** (0,10)	0,15* (0,08)
Observaciones	2373	2373	2373	2373
Panel B: Efectos por cuartil - Todas las niñas				
Cuartilo 1	-0,14 (0,13)	-0,01 (0,12)	0,08 (0,13)	-0,01 (0,13)
Cuartilo 2	-0,08 (0,13)	-0,02 (0,11)	0,04 (0,14)	-0,02 (0,12)
Cuartilo 3	0,01 (0,10)	-0,14 (0,10)	0,15 (0,11)	0,03 (0,10)
Cuartilo 4	-0,04 (0,10)	0,03 (0,10)	0,10 (0,14)	0,05 (0,11)

Tabla 4: Resultados según puntaje en pruebas de línea base – Totalidad de la muestra

Variables dependientes:	Ciencias y medio ambiente			
	Cuerpo humano	Medio ambiente	Mundo Físico	Generales
	(1)	(2)	(3)	(4)
Panel C: Efectos por cuartil - Todos los varones				
Cuartilo 1	-0,04 (0,12)	0,09 (0,10)	0,10 (0,11)	0,07 (0,11)
Cuartilo 2	-0,03 (0,13)	0,20* (0,10)	0,17 (0,10)	0,14 (0,10)
Cuartilo 3	0,00 (0,11)	-0,05 (0,11)	0,37*** (0,12)	0,18 (0,12)
Cuartilo 4	0,06 (0,09)	0,03 (0,11)	0,40*** (0,11)	0,24** (0,10)
Observaciones	2373	2373	2373	2373

Errores estándar estimados agrupados por escuela entre paréntesis. * Significativo en 10 %; ** significativo en 5 %; *** significativo en 1%.

De manera similar a lo observado en los resultados globales, no hay impacto para los módulos “cuerpo humano” y “medio ambiente” (Columnas 1 y 2). Sin embargo, los impactos del módulo “mundo físico” están concentrados para los alumnos que se encontraban en la mitad superior de la distribución de puntajes iniciales (Columna 3). Por lo tanto, la intervención acentúa la distribución, donde los estudiantes con calificaciones más altas en las mediciones iniciales sacan más provecho que los alumnos con puntajes preliminares relativamente más bajos. Más aún, parece existir algún efecto en las puntuaciones generales, pero sólo para el cuartil superior (Columna 4). Este resultado tiene importantes consecuencias en cuanto al diseño de las políticas. Una opción posible sería evitar acentuar la inequidad de los impactos de la intervención clasificando a los alumnos de acuerdo con su preparación inicial (puntajes en los exámenes de referencia), abordaje que a menudo se denomina de “seguimiento”. Como demuestra el estudio de Duflo et al. (2011), cuando se divide a los alumnos en grupos de capacidades similares,

de acuerdo con las mediciones iniciales, mejora el rendimiento de toda la distribución, y no sólo de los que tienen mejores calificaciones. La clasificación permite ajustar mejor la velocidad de enseñanza a las necesidades de cada alumno.

Sin embargo, también puede suceder que sólo los alumnos con mayores habilidades puedan asimilar las competencias ofrecidas por la intervención. Por lo tanto, incluso realizando el seguimiento, los alumnos con menores capacidades no sacarían provecho del programa. Aun cuando se presente dicha situación, sería apropiado emplear el programa para los alumnos más capacitados con un diseño de seguimiento. Pero sería necesario desarrollar un método alternativo para los alumnos con menor rendimiento. Desafortunadamente, nuestra intervención no incorporó ningún seguimiento y, por lo tanto, no podemos contribuir a este tema, que quedará pendiente para futuras investigaciones.

4.5. Resultados por género, ubicación y distribución de los puntajes iniciales

Anteriormente habíamos notado que los impactos positivos se encontraban en la mitad superior de la distribución inicial de las calificaciones y que no había un impacto general en los puntajes de las niñas. En los Paneles B y C de la Tabla 4, vemos los efectos por género y la distribución inicial de los puntajes.

En primer lugar, no existe un impacto diferencial para las niñas en las calificaciones, ni en los subtemas ni a nivel general (Panel B). De acuerdo con la distribución de los puntajes iniciales, el impacto es estadísticamente casi igual a cero en toda la distribución. Para los niños (Panel C) existe un efecto positivo en el módulo “mundo físico” y los puntajes generales (Columnas 3 y 4). También notamos que la dimensión de los efectos

aumenta en los puntajes iniciales. En tanto la mitad inferior tuvo impactos (insignificantes), con desviaciones estándar que oscilan entre 0,1 y 0,17, la mitad superior tuvo impactos significativos con desviaciones típicas de entre 0,37 y 0,4 (Columna 3). Un patrón similar surge de los puntajes generales donde los efectos oscilan entre un impacto insignificante con desviaciones estándar de 0,07 en el cuartil inferior y un impacto significativo con desviaciones estándar de 0,24 en el cuartil superior (Columna 4).

Como observamos previamente que los efectos principales se presentaban en las áreas urbanas, repetimos el análisis de los efectos heterogéneos de acuerdo con las calificaciones iniciales y género solamente para escuelas urbanas.^{xiii} La Tabla 5 a continuación presenta los resultados.

Tabla 5: Resultados según puntaje en pruebas de línea base – Muestra urbana

Variables dependientes:	Ciencias y medio ambiente			
	Cuerpo humano	Medio ambiente	Mundo Físico	Generales
	(1)	(2)	(3)	(4)
Panel A: Efectos por cuartil ó Totalidad de la muestra urbana				
Cuartilo 1	-0,08 (0,10)	0,14 (0,09)	0,13 (0,12)	0,09 (0,11)
Cuartilo 2	0,01 (0,11)	0,19** (0,09)	0,12 (0,13)	0,13 (0,12)
Cuartilo 3	0,08 (0,08)	-0,01 (0,09)	0,29** (0,13)	0,17 (0,11)
Cuartilo 4	0,00 (0,06)	0,03 (0,09)	0,23** (0,12)	0,13 (0,09)
Observaciones	1718	1718	1718	1718
Panel B: Efectos por cuartil ó Urbano niñas				
Cuartilo 1	-0,02 (0,15)	0,08 (0,13)	0,14 (0,16)	0,09 (0,15)
Cuartilo 2	-0,07 (0,15)	0,04 (0,12)	0,09 (0,18)	0,04 (0,15)
Cuartilo 3	0,08 (0,10)	-0,03 (0,11)	0,19 (0,13)	0,12 (0,12)
Cuartilo 4	-0,10 (0,11)	0,07 (0,12)	0,10 (0,17)	0,04 (0,13)

Tabla 5: Resultados según puntaje en pruebas de línea base – Muestra urbana

Variables dependientes:	Ciencias y medio ambiente			
	Cuerpo humano	Medio ambiente	Mundo Físico	Generales
	(1)	(2)	(3)	(4)
Panel C: Efectos por cuartilo ó Urbano varones				
Cuartilo 1	-0,11 (0,13)	0,19* (0,11)	0,13 (0,14)	0,09 (0,13)
Cuartilo 2	0,07 (0,13)	0,31*** (0,10)	0,15 (0,12)	0,21* (0,11)
Cuartilo 3	0,08 (0,14)	-0,00 (0,13)	0,38** (0,16)	0,22 (0,15)
Cuartilo 4	0,10 (0,08)	-0,00 (0,12)	0,36*** (0,13)	0,22* (0,12)
Observaciones	1718	1718	1718	1718
Errores estándar estimados agrupados por escuela entre paréntesis. * Significativo en 10 %; ** significativo en 5 %; *** significativo en 1%.				

En general, observamos un patrón muy similar con respecto a la muestra completa. La Columna 3 del Panel A muestra impactos crecientes en toda la distribución inicial en el módulo “mundo físico”. Estos impactos son significativos una vez más para los alumnos con mejor desempeño, de la mitad superior. El Panel B muestra impactos nulos para las niñas, en tanto el Panel C documenta efectos positivos para los niños en la mitad superior de los puntajes iniciales con respecto al módulo “mundo físico” y la calificación general. Claramente, los resultados obtenidos para la muestra completa están influidos por los resultados logrados en las escuelas urbanas.

4.6. Efectos en otras materias

El nuevo método de enseñanza no tuvo como objetivo mejorar las calificaciones en otras áreas académicas (matemáticas y comprensión lectora). Sin embargo, se incorporó al programa piloto este posible efecto en los exámenes iniciales y finales de estos cursos.

Los efectos estimados, sin embargo, sugieren que no habrá impactos estadísticamente significativos en estos cursos.

Los resultados indican que, al menos en el corto plazo, el nuevo método pedagógico no tiene efecto sobre el aprendizaje de otros cursos. Sin embargo, queda abierto el interrogante de si el método podría impactar en otras áreas del aprendizaje a largo plazo.

4.7. Influencia del nivel de educación de los padres

La heterogeneidad de los resultados podría responder parcialmente a diferencias en las características del hogar (Freeman et al., 2010). Un aspecto clave dentro de estas características es el nivel de educación del jefe de familia. En la Tabla 6 se muestran los diferentes impactos por nivel de educación del jefe del hogar, es decir, sin nivel secundario o con nivel secundario o superior.

Tabla 6: Efectos del nivel educación del jefe de hogar		
Variables dependientes:	Calificaciones estandarizadas al final	
	Comparación	Calificación
	Media del grupo	Final
	(1)	(2)
Panel A: Efectos - Cuerpo humano		
Jefe hogar sin secundario	-0,20	0,13 (0,14)
Jefe hogar con secundario o superior	0,23	-0,04 (0,11)
Observaciones		1109
Panel B: Efectos - Medio ambiente		
Jefe hogar sin secundario	-0,11	0,08 (0,18)
Jefe hogar con secundario o superior	0,24	0,08 (0,13)
Observaciones		1109
Panel C: Efectos - Mundo físico		
Jefe hogar sin secundario	-0,14	0,30** (0,14)
Jefe hogar con secundario o superior	0,21	0,29* (0,15)
Observaciones		1109

Tabla 6: Efectos del nivel educación del jefe de hogar

Variables dependientes:	Calificaciones estandarizadas al final	
	Comparación	Calificación
	Media del grupo	Final
	(1)	(2)
Panel D: Efectos ó General ciencias y medio ambiente		
Jefe hogar sin secundario	-0,19	0,23 (0,17)
Jefe hogar con secundario o superior	0,28	0,16 (0,15)
Observaciones		1109
Panel E: Efectos - Matemática		
Jefe hogar sin secundario	-0,12	0,10 (0,14)
Jefe hogar con secundario o superior	0,17	0,22 (0,14)
Observaciones		1115
Panel F: Efectos - Lectura		
Jefe hogar sin secundario	0,02	0,03 (0,15)
Jefe hogar con secundario o superior	0,18	0,15 (0,11)
Observaciones		1107

Errores estándar estimados agrupados por escuela entre paréntesis. * Significativo en 10 %; ** significativo en 5 %; *** significativo en 1%.

La primera columna de la tabla muestra los puntajes promedio de las pruebas estandarizadas según nivel de educación del jefe del hogar, respecto del grupo de control al final de la implementación. Estos puntajes tienen en promedio un valor de cero y una desviación estándar de uno. Para todas las materias evaluadas, la calificación promedio es negativa para los estudiantes pertenecientes a hogares cuyos jefes no tienen educación secundaria; es decir estudiantes por debajo de la media de la distribución general. Por el contrario, los estudiantes pertenecientes a hogares donde el jefe de familia tiene estudios secundarios o superiores, presentan puntajes promedio positivos (es decir, por encima de la media). Consecuentemente, los estudiantes con padres más educados presentan calificaciones más altas.

En la Columna 2 de esta tabla, se muestran los impactos del programa diferenciados por nivel de educación del jefe del hogar. Sólo se observa un impacto positivo significativo en el módulo “mundo físico” (Panel C). No obstante, la dimensión del impacto es similar, con desviación estándar de 0,3 y 0,29, para los jefes de hogar sin y con educación secundaria, respectivamente. Por lo tanto, si bien existe una diferencia marcada en los puntajes para estudiantes según el nivel de educación del hogar, el impacto absoluto positivo del programa es independiente de estas diferencias educacionales.

4.8. Percepción de los docentes

El estudio piloto también tuvo como propósito determinar el impacto de la intervención en la percepción de los docentes respecto de su propio trabajo y de sus estudiantes. Las posibles respuestas a cada una de las preguntas de la encuesta se diseñaron en una escala de uno a cinco, donde (1) indicaba "muy de acuerdo" y (5) "muy en desacuerdo". Esta escala se estandarizó entre el intervalo [0; 1], de modo tal que si la percepción específica tenía una connotación negativa, entonces un valor cero (0) indicaba "muy de acuerdo", mientras que el valor (1) representaba "muy en desacuerdo". Si la percepción era sobre una característica positiva, entonces el valor cero (0) indicaba "muy en desacuerdo" y el uno (1) "muy de acuerdo".

Las distintas áreas de percepciones del docente comprendidas en el cuestionario fueron: trabajo en la escuela, conocimientos de ciencias y de medio ambiente; importancia de las ciencias y del medio ambiente; motivación en la enseñanza de ciencias y medio ambiente; y el proceso de enseñanza. En general, no hubo impactos de la intervención sobre las percepciones de los maestros con respecto a su trabajo. No obstante, se registró

una excepción significativa. Hubo un impacto negativo en la percepción respecto de la necesidad de enseñar primero teoría y sólo después la práctica. Si bien la conveniencia de la oportunidad entre teoría y práctica es discutible, esta pregunta estaba destinada a detectar si los docentes habían adquirido lo que se buscaba con la metodología: alejarse de las explicaciones en clase para pasar a ejemplos prácticos, utilizando los juegos LEGO. En consecuencia, este resultado muestra que los docentes en el grupo de tratamiento, al menos en la retórica, trabajaron con éxito en la transición del enfoque tradicional sobre la historia de las ciencias y la memorización de conceptos.

Además, el cuestionario estaba dirigido a las percepciones respecto de sus alumnos. En particular, evaluó la percepción de los docentes respecto de los estudiantes, en las siguientes áreas: motivación general, trabajo en equipo, relaciones domésticas (consideradas en términos de percepción respecto del apoyo y compromiso de las familias en el proceso de aprendizaje de los alumnos), capacidad intelectual y desempeño en ciencias y medio ambiente. Los resultados se presentan en la Tabla 7.

Tabla 7: Percepciones de los docentes respecto de alumnos

	Comparación Final Media grupo	Impacto programa	Número de Observaciones
	(1)	(2)	(3)
Panel A: Variables dependientes ó			
Motivación general			
Buena conducta	0,66	0,01 (0,05)	124
Muy interesados en aprender	0,77	-0,01 (0,04)	125
Muy buena tasa de asistencia	0,77	-0,01 (0,05)	124
Promedio por familia de logros (en desviaciones estándar)		-0,02 (0,12)	136

Tabla 7: Percepciones de los docentes respecto de alumnos

	Comparación Final Media grupo (1)	Impacto programa (2)	Número de Observaciones (3)
Panel B: Variables dependientes - Trabajo en equipo			
Saben trabajar en equipo	0,73	0,04 (0,05)	124
Saben escuchar y hablar en grupos	0,70	-0,03 (0,04)	124
Son autónomos	0,64	-0,05 (0,05)	124
Promedio por familia de logros (en desviaciones estándar)		0,07 (0,08)	136
Panel C: Variables dependientes - Relaciones domésticas			
Problemas en el hogar afectan su aprendizaje	0,49	0,02 (0,06)	124
Sus familias los apoyan	0,41	0,07 (0,05)	123
Promedio por familia de logros (en desviaciones estándar)		0,08 (0,14)	136
Panel D: Variables dependientes:			
Buena capacidad de expresión oral y escrita	0,58	0,04 (0,04)	125
Problemas de concentración afectan su aprendizaje	0,42	-0,03 (0,05)	124
Con capacidad de aprender cualquier concepto	0,60	0,02 (0,04)	123
Saben razonar y nunca estudian de memoria	0,57	-0,03 (0,04)	124
Promedio por familia de logros (en desviaciones estándar)		0,06 (0,11)	136
Panel E: Variables dependientes - Desempeño en ciencias y medio ambiente			
Muy interesados en aprender ciencias y medio ambiente	0,72	0,08* (0,04)	125
Buen desempeño en ciencias y medio ambiente	0,66	0,06** (0,03)	124
Promedio por familia de logros (en desviaciones estándar)		0,34** (0,14)	136

La columna (1) informa el logro promedio dentro del grupo de control al final. La columna (2) informa coeficientes a partir de una regresión, donde el indicador de la escuela de tratamiento se indica como variable a la derecha. La columna (3) informa el número de observaciones en cada regresión. Errores estándar estimados agrupados por escuela entre paréntesis. * Significativo en 10 %; ** significativo en 5 %; *** significativo en 1%.

El único efecto significativo se presentó en las percepciones del interés y desempeño en ciencias y medio ambiente (Panel E). Los maestros en las escuelas de tratamiento, en

comparación con las de control, percibieron que los estudiantes estuvieron interesados en aprender “ciencias” y “medio ambiente” y que tuvieron un buen desempeño en estos cursos. El impacto, medido según las desviaciones estándar del índice, fue de 0,34 unidades. En consecuencia, parece que los docentes perciben que sus estudiantes están más interesados y presentan un mejor desempeño en este curso. Cuando las percepciones de los docentes se correlacionan con el desempeño estudiantil, interpretamos que estos resultados corroboran las evidencias de los efectos positivos en los puntajes de las pruebas estandarizadas.

5) Conclusiones

A medida que el Perú se acerca a la universalización de la educación primaria, el país presta más atención a mejorar la calidad de la educación, en lugar de aumentar la cobertura. La calidad de la educación en el Perú es un problema importante, tal como lo demuestra el pobre desempeño del país en las pruebas estandarizadas nacionales, regionales e internacionales. La necesidad de contar con evidencia que demuestre qué mejora la calidad de la educación es crucial para las políticas.

En este documento, presentamos los resultados de una evaluación experimental sobre un nuevo enfoque pedagógico centrado en el alumno, para la enseñanza de ciencias y medio ambiente, para el tercer grado, en el Perú. El estudio se basa en los datos obtenidos a partir de una intervención diseñada en forma rigurosa, así como también de encuestas a directores, docentes, estudiantes y padres. A pesar de las complicaciones que ocasionó la implementación parcial del piloto, observamos mejoras positivas y significativas en las calificaciones de los estudiantes que aprendieron con este nuevo método. Estos efectos se

concentraron en las áreas geográficas donde el programa se implementó de manera más intensiva (escuelas urbanas) y fueron más fuertes en estudiantes que habían tenido mejores calificaciones iniciales. Además, aparentemente debido a que monopolizan el empleo de los materiales didácticos, los varones se beneficiaron más del programa.

Un reto para la posible ampliación del programa es asegurar que todos los estudiantes se beneficien de los nuevos métodos, particularmente las niñas y el alumnado rural. Esta es una cuestión fundamental para el diseño de políticas en un país con inequidades en la distribución de la calidad de la educación y en los logros de aprendizaje.

ⁱ Por ejemplo, entre 1998 y 2009 la matrícula preescolar aumentó del 53,4 al 66,3 %; por su parte la matriculación en la educación primaria pasó del 90,6 al 94,4 %; y en la secundaria del 59 al 76,5 % (fuente: Ministerio de Educación de Perú).

ⁱⁱ Evaluación Censal de Estudiantes, UMCE; MINEDU 2010.

ⁱⁱⁱ El salario mínimo en Perú durante el programa piloto era de S/. 550 por mes. Luego fue aumentado a S/. 700 en agosto de 2011.

^{iv} Obsérvese que expresamos la función en una especificación lineal, para fines de exposición. No obstante, la función de producción no conocida puede adoptar distintas formas no lineales.

^v Téngase en cuenta, no obstante, que existe una extensa literatura no experimental que emplea técnicas de simulación con el objetivo de estimar las funciones de producción educativas y la incidencia relativa de distintas variables cognitivas y no cognitivas (Cunha y Heckman, 2007).

^{vi} Además, este estudio emplea datos provenientes del censo nacional de educación de Perú de 2009, para el control de las variables en análisis.

^{vii} Obsérvese, no obstante, que si la elección de dos secciones, en las escuelas con tres o más secciones, no hubiese sido realmente aleatoria, y además, no hubiese habido un equilibrio entre escuelas de tratamiento y de control, el diseño de nuestra investigación habría estado sesgado. Sin embargo, tal como se muestra en la Tabla 1, las escuelas de ambos grupos están equilibradas, independientemente de la cantidad de secciones que tenga cada establecimiento.

^{viii} La validación del diseño de estas pruebas iniciales había sido comprobada previamente en dos escuelas de Lima que no son parte de la población muestral. Se tomaron exámenes de referencia a los alumnos de tercer grado a comienzos y a fines del ciclo lectivo. Si bien el programa piloto no incluyó actividades específicas para mejorar las destrezas en lectura y matemáticas, se consideró que era deseable evaluar también estas áreas. Las pruebas abarcaron las expectativas de aprendizaje para cada curso en esos dos momentos del año. Se estandarizaron en las áreas de conocimiento que se suponía serían adquiridas según el programa de estudios nacional. Recuérdese que nuestra metodología sólo modificó el modo de la enseñanza, y que los objetivos de aprendizaje fueron los mismos, tanto en la metodología tradicional como en la nueva metodología experimental que se pretendía analizar.

^{ix} Adviértase que todos los puntajes están expresados en desviaciones estándar con respecto al grupo de control. Es decir, las calificaciones en el grupo de control muestran una desviación media de cero y una desviación estándar de uno.

^x El número total de alumnos al inicio era 2.802; no obstante, de esta cifra se excluyeron los estudiantes que pertenecían al programa de inclusión (que integra alumnos con necesidades especiales de aprendizaje en aulas comunes), los que pasaban al cuarto grado durante el período escolar y los que no terminaron todas las pruebas, lo cual deja 2.771 observaciones efectivas. Al final, la cantidad total de evaluados fue 2.663; si se excluyen a los estudiantes del programa de inclusión, los que habían pasado al cuarto grado y los que no finalizaron todos los exámenes, la cifra desciende a 2.552. De este último grupo, solamente se evaluaron 2.373 alumnos al inicio.

^{xi} Los resultados de los exámenes fueron estandarizados de modo tal que el grupo de control mostró una desviación media de cero y una desviación estándar de uno. En consecuencia, las diferencias entre ambos grupos están expresadas en desviaciones estándar con respecto al grupo de control.

^{xii} Nuevamente, los resultados fueron estandarizados de modo tal que el grupo de control mostró una desviación media de cero y una desviación estándar de uno. En consecuencia, las diferencias entre los grupos están expresadas en desviaciones estándar con respecto al grupo de control.

^{xiii} También lo hicimos para las escuelas rurales. Sin embargo, no se encontraron impactos.

Referencias

Banerjee, A., Cole, S., Duflo, E. y Linden, L. (2007), Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics*, 122(3), pp. 1235-1264.

Barrera-Osorio, F. y Linden, L. (2009). El Uso y Mal Uso de Computadores en Educación: Evidencia de una Prueba de Evaluación Aleatoria en Colombia. *World Bank Policy Research Working Paper 4836*.

Benavides, M. y Mena, M. (2010). Informe de progreso educativo, Perú 2010. GRADE, Peru.

Bloom, H. y Michalopoulos, C. (2010). When is the story in the sub groups: strategies for interpreting and reporting effects for sub groups. *MDRC Working Paper 551*.

Bloom, H., Bos, J. y Lee, S. (1999). Using cluster random assignment to measure program impacts: statistical implications for the evaluation of education programs. *Evaluation Review*, 23 (4), pp. 445-469.

Bonwell, C. y Eison, J. (1991). Active Learning; Creating Excitement in the Classroom. *AEHE-ERIC Higher Education Report No. 1*. Washington DC.

Carrillo, P., Onofa, M. y Ponce, J. (2010). Information Technology and Student's Achievement: Evidence from a Randomized Experiment in Ecuador. Documento no publicado. Washington, DC: Inter-American Development Bank.

Colburn, A. (2000). An inquiry primer. *Science Scope*, 23 (6), pp. 42-44.

Cristia, J., Cueto, S., Ibarra, P., Severin, E., y Santiago, A. (2012). Technology and Child Development: Evidence from the One Laptop per Child Program. *IDB Working Paper No 273*.

Cunha, F., y Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, 97(2), pp 31-47.

Duflo, E., Dupas, P., y Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5), pp. 1739–1774.

Freeman R., Machin, S. y Viarengo, M. (2010). Variation in educational outcomes and policies across countries and schools within countries. *CEE Discussion Papers 0117*.

Furtak, E., Seidel, T., Iverson, H., and Briggs, D. (2012) Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis. *Review of Educational Research*, 82 (3), pp. 300-329.

Glewwe, P. y Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries, in: Hanushek E. y Welch, F. (eds.) *Handbook of the economics of education* (Elsevier), 2(2).

Glewwe, P., Hanushek, E., Humpage, S., y Ravina, R. (2011). School resources and educational outcomes in developing countries: A Review of the Literature from 1990 to 2010. *Manuscrito no publicado, University of Minnesota*.

Healy, J. (1990) *Endangered minds: why our children don't think and what we can do about it*. (New York: Simon and Schuster).

Iturrizaga, I. (2000). Study of the educational impact of the LEGO Dacta Materials-INFOESCUELA-MED. Ministerio de Educación del Perú, pp. 1-39.

Hussain S., Lindh, J. y Shukur, G. (2006). The effect of Lego training on student's school performance in mathematics problem solving ability and attitude: Swedish data. *Educational technology and society*, 9(3), pp. 182-194.

Lowery, L. (1998). *The biological basis of thinking and learning*. University of California, Berkeley.

Malamud, O. and Pop-Eleches, C. (2011) Home Computer Use and the Development of Human Capital. *Quarterly Journal of Economics*, 126 (2), pp. 987-1027.

Ministerio de Educación. (2008) *Diseño Curricular Nacional de Educación Básica Regular*, Lima - Peru.

Noble M. (2001) *The Educational Impact of Lego Dacta Materials*. Sheffield Hallam University.

Organization for Economic Cooperation and Development (OECD) (2011) *PISA 2009 Results: What Students Know and Can Do*. Paris: OECD

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2008) *First Report of the Second Regional Comparative Education Study*. Santiago de Chile, Chile.