



Conecta Ideas

Evaluación de Impacto - Reporte Final

Fecha de publicación: 5 de Junio, 2024

Mauricio Romero; J-PAL & ITAM; E-mail: mtromero@itam.mx

Paul Gertler; J-PAL & UC Berkeley; E-mail: gertler@berkeley.edu

ITAM

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

Berkeley
UNIVERSITY OF CALIFORNIA



J-PAL

ABDUL LATIF JAMEEL POVERTY ACTION LAB

LATIN AMERICA & CARIBBEAN

Este proyecto de investigación fue financiado por la Education Endowment Foundation, en colaboración con la Fundación BHP, como parte del proyecto "Construyendo un ecosistema de evidencia global para la enseñanza". Acuerdo de Subvención EEF -032022-SUMMA.

Las opiniones expresadas en este trabajo son las de los autores y no representan necesariamente las de la Education Endowment Foundation — EEF, BHP Foundation o sus respectivas Juntas Directivas.

Miembros del equipo:

SUMMA: Laboratorio de Investigación e Innovación Educativa para América Latina y el Caribe

- Rafael Carrasco, Subdirector de SUMMA
- Dante Castillo, Director de Innovación
- María José Sepúlveda, Directora de Evaluación

Equipo de Evaluación

- Mauricio Romero, Associate Professor, Instituto Tecnológico Autónomo de México (ITAM)
- Paul Gertler, Profesor Li Ka Shing, Escuela de Negocios Haas, UC Berkeley
- Cristian Muñoz, Gerente de Investigación, J-PAL
- Gabriel Durán, Gerente de Investigación, J-PAL
- Emiliano Ramírez, Investigador Asociado Senior, J-PAL
- Laura Torres, Investigadora Asociada, J-PAL
- Alondra Arellano, Asistente de Campo, J-PAL

Equipo de Implementación

- Roberto Araya, Profesor Titular. Universidad de Chile
- Julián Cristia, Economista Principal, Departamento de Investigación del Banco Interamericano de Desarrollo
- Julián Messina, Beatriz Galindo Investigadora Distinguida Senior, Universidad de Alicante

Cómo citar este documento: SUMMA, J-PAL (2023) Conecta Ideas — Evaluación de Impacto. Proyecto del Fondo de Impacto "Construcción de un ecosistema global de evidencia para mejorar la enseñanza".

Resumen ejecutivo

El proyecto

- El estudio evalúa el impacto de un software de aprendizaje ("Conecta Ideas") en los niveles de logro de los estudiantes en Chile.
- La intervención fue realizada por AutoMind (la empresa que desarrolló Conecta Ideas). Se llevó a cabo durante las clases regulares de matemáticas en el laboratorio de computación de la escuela y fue dirigido por coordinadores de laboratorio contratados por AutoMind. En una sesión típica de IC, los estudiantes trabajan para resolver de 20 a 30 ejercicios matemáticos. Los estudiantes reciben retroalimentación automática sobre si sus respuestas son correctas. Los coordinadores de laboratorio, contratados y supervisados por el equipo de Conecta Ideas, son responsables de llevar a cabo las sesiones de aprendizaje en el laboratorio de computación en colaboración con los maestros de las aulas regulares.
- El programa fue asignado aleatoriamente, lo que permite recuperar el efecto causal de reemplazar las clases regulares de matemáticas por sesiones de Conecta Ideas.
- Los colegios de la evaluación se encuentran todos en el área metropolitana de Santiago de Chile. Todos ellos expresaron su interés en Conecta Ideas. Los estudiantes de 4° grado son el objetivo de la intervención y la evaluación.
- Las escuelas que optaron por participar en el estudio tienen más probabilidades de: 1) ser públicas, 2) tener matrícula cero, 3) tener un acuerdo con el Programa de Integración Escolar (PIE) y 4) tener más estudiantes en riesgo de abandonar la escuela. Por lo tanto, los resultados del experimento pueden no extrapolarse al universo de las escuelas de Santiago.
- La intervención comenzó en junio de 2022 y finalizó en diciembre de 2022. Estimamos el impacto utilizando datos de pruebas estandarizadas del Sistema de Medición de Calidad de la Educación (SIMCE) para estudiantes de 4° grado. La prueba SIMCE se llevó a cabo en noviembre de 2022 y midió el logro de los estudiantes en matemáticas y lenguaje (español).
- Hay tres grupos de tratamiento diferentes (con 47 escuelas cada uno): uno con una versión "simple" o "plain vanilla" de Conecta Ideas, en el que el estudiante recibe puntos simbólicos por completar ejercicios en la plataforma. Un segundo grupo recibe incentivos individuales, en los que los puntos simbólicos sirven como boletos de rifa para un premio. Un tercer grupo recibe incentivos grupales, en los que se suman los puntos simbólicos de cada aula y estos sirven como boletos de rifa de un premio para toda la clase. Además, hay un grupo de control (con 94 escuelas) donde las clases de matemáticas se dan como de costumbre.
- En total, 4.053 estudiantes de escuelas control rindieron la prueba SIMCE 2022. De los grupos de tratamiento, 1.295 estudiantes del grupo de "vainilla simple", 1.291 del grupo de incentivos individuales y 1.208 del grupo de incentivos grupales tomaron el examen.
- Los estudiantes de todas las escuelas tratadas pasaron un tiempo promedio de 17,71 minutos por semana en la plataforma y realizaron 22 ejercicios semanales. Esto equivale a aproximadamente 642 ejercicios en el transcurso del experimento y 8,5 horas pasadas en la plataforma. En concreto, los alumnos de los centros educativos con la versión "simple" o "plain vanilla" de Conecta Idea pasaron una media de 15,65 minutos semanales en la plataforma y realizaron 20 ejercicios semanales. Los estudiantes de las escuelas de incentivo individuales pasaron un tiempo promedio de 18,38 minutos por semana en la plataforma y realizaron 24 ejercicios semanales. Los estudiantes de las escuelas de incentivo grupal pasaron un tiempo promedio de 17,5 minutos por semana en la plataforma y realizaron 23 ejercicios semanales.

Las principales conclusiones del estudio son:

- La sustitución de las clases regulares de matemáticas por sesiones de Conecta Ideas no mejoró el rendimiento de los estudiantes de 4° grado en matemáticas medidas por la prueba estandarizada SIMCE.

No hay diferencias estadísticamente significativas en el rendimiento de los estudiantes en las puntuaciones de matemáticas en ninguno de los grupos de tratamiento (en relación con el control).

- Conecta Ideas tuvo un efecto negativo, aunque estadísticamente insignificante, en las puntuaciones de matemáticas. Las estimaciones puntuales son negativas para los tres tratamientos, y podemos descartar que los efectos de cualquiera de las versiones del tratamiento estén por encima de $0,045\sigma$ al nivel del 95%. En la Tabla 1 se resumen los principales hallazgos.
- Proporcionar incentivos individuales o grupales no cambió significativamente el uso de la plataforma. Consistentemente, el impacto en los puntajes de matemáticas es similar si los estudiantes reciben incentivos individuales o grupales.
- Hay algunas pruebas sugestivas de efectos secundarios negativos de las ideas de Conecta en las puntuaciones lingüísticas. Los incentivos individuales tienen un efecto negativo ($-0,091\sigma$) y estadísticamente significativo (valor $p = 0,017$). No está claro por qué podría haber un efecto negativo en las puntuaciones lingüísticas. Hay un efecto negativo y estadísticamente significativo en las puntuaciones lingüísticas para el grupo con incentivos individuales (las estimaciones puntuales son negativas, pero no estadísticamente significativas para los otros dos grupos de tratamiento).
- La evaluación del proceso sugiere que hubo varios problemas con la implementación del programa, incluyendo retrasos en el inicio del programa, infraestructura inadecuada en las escuelas, pérdida de tiempo en la transición entre el aula regular y el laboratorio de computación, y premios inadecuados para los incentivos para motivar a los estudiantes. Además, los datos de la plataforma sugieren que los estudiantes están perdiendo efectivamente tiempo de instrucción en matemáticas debido al tratamiento. Estos problemas pueden ayudar a explicar la falta de efecto en las puntuaciones de matemáticas. No está claro por qué las puntuaciones lingüísticas pueden haber bajado.

Tabla 1: Resumen de las principales conclusiones

Grupo	Efecto en puntajes de matemáticas	Efecto en puntajes de lenguaje	Número de estudiantes (matemáticas)	p-valor (matemáticas)	p-valor (lenguaje)
Conecta Ideas Simple	-0.059σ	0.006σ	1,295	0.16	0.89
Incentivos Individuales	-0.037σ	-0.091σ	1,291	0.367	0.014
Incentivos Grupales	-0.075σ	-0.043σ	1,208	0.103	0.245

- Estos números corresponden a la Tabla 13. Consulte la sección 4 para obtener más detalles.

Hallazgos adicionales

- Conecta Ideas no tiene ningún efecto sobre la autopercepción de los estudiantes sobre su capacidad de aprendizaje o motivación para asistir a la escuela y aprender.
- Un análisis dosis-respuesta proporciona resultados contradictorios. El uso de la plataforma se correlaciona con los puntajes de las pruebas. Sin embargo, esto podría deberse a diferencias en las características de la escuela, los estudiantes y los maestros que se correlacionan tanto con los puntajes de las pruebas como con el uso (p. ej., el uso también se correlaciona con los puntajes previos al tratamiento). El uso de un enfoque de variables de instrumento (es decir, instrumentar el uso con la asignación aleatoria entre los grupos de tratamiento) sugiere que el uso de la plataforma no mejora significativamente las puntuaciones de matemáticas y lenguaje (incluso puede disminuir las puntuaciones).
- No hay evidencia de efectos heterogéneos del tratamiento en cuanto al sexo, el estado de NSE o el rendimiento inicial.

Costos

El costo por estudiante (por año) para 147 secciones tratadas es de USD 170,8. A gran escala (con más de 5.000 cursos), muchos de estos costes se reducen al ser principalmente fijos, y el coste por año (al año) pasa a ser de 91,05 dólares.

Contenido

<i>Antecedentes</i>	<i>6</i>
<i>Objetivos</i>	<i>10</i>
<i>Metodología</i>	<i>11</i>
<i>Resultados de la evaluación de impacto.....</i>	<i>33</i>
<i>Resultados de la evaluación de procesos.....</i>	<i>79</i>
<i>Costos</i>	<i>93</i>
<i>Conclusiones</i>	<i>94</i>
<i>Referencias</i>	<i>97</i>
<i>Equipo del proyecto</i>	<i>101</i>
<i>Tablas y figuras adicionales</i>	<i>103</i>

Antecedentes

Contexto y pertinencia de la evaluación

La pandemia de COVID-19 ha tenido un gran impacto en las oportunidades educativas en todo el mundo, y los informes de pérdidas de aprendizaje son asombrosos (véanse las revisiones de [Patrinos, Vegas, and Carter-Rau \(2022\)](#); [Moscoviz and Evans \(2022\)](#); [Betthausen, Bach-Mortensen, and Engzell \(2023\)](#)), así como la evidencia más reciente de México (Alasino, Ramírez, Romero, Schady, & Uribe, 2024) e India ([Singh, Romero, & Muralidharan, 2022](#)). América Latina tuvo el cierre de escuelas más largo debido a la pandemia de COVID-19, con 158 días de cierre total de escuelas en promedio, y Chile no fue la excepción, con la mayoría de las escuelas públicas cerradas durante 2020 y 2021 ([Unicef, 2021](#); [Schady, Holla, Sabarwal, Silva, & Yi Chang, 2023](#)). El panorama educativo para los próximos años es sombrío, especialmente entre las poblaciones vulnerables que ya carecían de oportunidades suficientes. Antes de la pandemia, Chile ocupaba el último lugar de la OCDE en las evaluaciones internacionales de matemáticas, mientras que las diferencias de rendimiento entre los estudiantes de nivel socioeconómico bajo y alto eran las más grandes ([Schleicher, 2019](#)). Debido a que los recursos que facilitan el aprendizaje en línea (por ejemplo, computadoras portátiles y tabletas, Internet de alta velocidad, tutorías, espacio adecuado dentro del hogar para el estudio y apoyo de los miembros de la familia) son más abundantes para los estudiantes de alto nivel socioeconómico, la pandemia seguramente aumentará las desigualdades de aprendizaje, incluso dentro de un aula ([Lai & Widmar, 2021](#)).

Los sistemas educativos de todo el mundo deben implementar políticas correctivas para garantizar que los estudiantes que se desvincularon durante la pandemia se pongan al día con los que obtuvieron mejores resultados con el aprendizaje a distancia. Esto presenta varios desafíos. Requerirá más gasto público en países con presupuestos extremadamente ajustados. También requerirá una motivación sustancial de los estudiantes para pasar más tiempo estudiando y ejercitando matemáticas y otras materias. Las plataformas tecnológicas que motivan a los estudiantes a aprender a través de juegos pueden ser rentables en comparación con otras políticas educativas (por ejemplo, transferencias adicionales a las escuelas, maestros adicionales o transferencias monetarias condicionadas; consulte [Kremer, Brannen, and Glennerster \(2013\)](#); [N. Angrist, Djankov, Goldberg, and Patrinos \(2021\)](#) para una revisión de otras alternativas de políticas y su costo-efectividad).

En este artículo estudiamos un experimento de campo a gran escala para evaluar Conecta Ideas. Conecta Ideas es una plataforma de aprendizaje diseñada en 2002 por AutoMind, un equipo multidisciplinario cuyo liderazgo se encuentra ahora en la Universidad de Chile. Desde 2004, la plataforma se ha puesto a prueba en varias escuelas para perfeccionarla. La plataforma se basa en la gamificación para motivar el aprendizaje en matemáticas, proporcionando pegatinas y premios virtuales a los estudiantes que lo hacen bien resolviendo los ejercicios de matemáticas en cada sesión y empleando la retroalimentación enfocada en abordar las áreas de oportunidad de los estudiantes para guiar el proceso de aprendizaje, realizar un seguimiento fácil del progreso del estudiante (en la sesión actual y en todas las sesiones) y ofrecer a cada estudiante informes personalizados en forma de gráficos donde se comparan con sus compañeros. El diseño de la evaluación nos permitirá probar el efecto de Conecta Ideas frente a un grupo control. Además, el diseño experimental contará con tres grupos de tratamiento (además del grupo control): Primero, un grupo donde los estudiantes obtendrán incentivos puramente simbólicos a nivel individual. Estos incentivos se otorgarán por el uso de la plataforma y por obtener respuestas correctas a los ejercicios. En segundo lugar, un grupo en el que estos incentivos simbólicos se traducen en premios tangibles (por ejemplo, libros) para los estudiantes. En tercer lugar, un grupo en el que los premios se otorgan en función del rendimiento en el aula. Los premios son los mismos que en el caso de los incentivos individuales, pero son recibidos por toda una clase, a diferencia de los estudiantes individuales dispersos por las aulas. Además de probar el efecto de Conecta Ideas, este diseño experimental nos permitirá probar: a) incentivos individuales vs grupales y b) incentivos tangibles, pero no monetarios vs. incentivos puramente simbólicos.

Si bien los ensayos controlados aleatorios anteriores habían encontrado un impacto positivo de Conecta Ideas, esta evaluación mejora los diseños de las evaluaciones anteriores al minimizar el papel de los efectos secundarios y el tiempo de instrucción adicional. Además, al tener múltiples grupos de tratamiento, podremos comprender si incentivar a los estudiantes tiene algún efecto adicional en los puntajes (y si los incentivos individuales o grupales funcionan mejor). Más explícitamente, el ensayo más reciente fue aleatorizado a nivel de aula dentro de las escuelas ([Araya, Ortiz, Botton, & Cristia, 2019](#)). Por lo tanto, es posible que el aula de control tuviera efectos secundarios del tratamiento, lo que hace que los efectos del tratamiento estén sesgados. Los efectos secundarios pueden ser positivos (p. ej., si los maestros concentran más su energía en las aulas de control) o negativos (p. ej., si los maestros en las aulas de control se sienten decepcionados de que sus estudiantes no sean tratados).

Además, todos los ensayos anteriores habían implementado Conecta Ideas fuera del horario escolar (al menos parcialmente), lo que efectivamente convirtió el tratamiento en un paquete de Conecta Ideas e instrucción adicional en matemáticas. Es posible, por lo tanto, que los efectos del tratamiento hayan sido impulsados por el tiempo adicional de matemáticas (independientemente de si se enseñó con Conecta Ideas o con un maestro tradicional). Además, como señalan Markopoulos, Fragkou, Kasidiaris, and Davim (2015), las evaluaciones de plataformas con enfoque de gamificación son relativamente escasas, por lo que probar la robustez de los resultados de Conecta Ideas contribuirá a construir evidencia robusta sobre gamificación. Por último, podremos comparar una versión "plain vanilla" de Conecta Ideas con versiones con incentivos adicionales para los estudiantes (ya sean individuales o grupales). En la Tabla 2 se presenta una comparación completa entre la evaluación experimental previa de Conecta Ideas realizada por Araya et al. (2019) y esta.

Tabla 2: Comparación entre las intervenciones de 2017 y 2022

Elemento	2017	2022
Población objetivo	Estudiantes de cuarto grado	Estudiantes de cuarto grado
Sesiones por semana	Dos sesiones de 90 minutos: una reemplazaba la clase de matemáticas tradicional y la otra era tiempo adicional de instrucción de matemáticas	Una sesión de 60-90 minutos
Método	Ensayo de campo aleatorizado	Ensayo de campo aleatorizado
Unidad de aleatorización	Clase	Escuela
Muestra	24 escuelas primarias públicas atendidas por estudiantes de bajos ingresos en Santiago, Chile	235 escuelas públicas y privadas-subsidiadas de la región metropolitana de Santiago (las escuelas que optaron por participar eran más propensas a ser públicas, sin matrícula, y con más estudiantes en riesgo de abandonar)
Grupos experimentales	Tratamiento y control. El tratamiento incluía cuatro estrategias diferentes de gamificación	Tres grupos de tratamiento y un grupo de control. Un tratamiento con la implementación tradicional de Conecta Ideas, uno con incentivos individuales para los estudiantes y uno con incentivos grupales para los estudiantes
Otros detalles de la aleatorización	Aleatorización a nivel de sección dentro de la escuela. No hay escuelas de control puro	Aleatorización en dos etapas (primero asignando escuelas a grupos de tratamiento, luego dentro de las escuelas tratadas asignando secciones al tratamiento)

Tiempo de exposición	Siete meses (de marzo a noviembre de 2017)	Cinco meses (de julio a noviembre de 2022)
Tiempo total promedio de uso de Conecta Ideas por estudiante	27 horas	8.5 horas
Uso en casa	2% del uso total	4% del uso total
Coordinadores de laboratorio para las sesiones de CI	Sí	Sí
Acceso a la plataforma de CI para las secciones del grupo de control	No	Las secciones de control de las escuelas tratadas podían usar la plataforma sin ayuda de los coordinadores de laboratorio
Efectos fijos	Efectos fijos de la escuela	Efectos fijos de estratos
Errores agrupados	A nivel de sección	A nivel de escuela

La literatura disponible sobre el uso de la tecnología en la educación revela que la mera expansión del acceso al hardware (por ejemplo, computadoras o tabletas) no tiene efectos positivos en el aprendizaje [Malamud and Pop-Eleches \(2011\)](#); [Fairlie and Robinson \(2013a\)](#); [Beuermann, Cristia, Cueto, Malamud, and Cruz-Aguayo \(2015\)](#); [Cristia, Ibarra, Cueto, Santiago, and Severin \(2017\)](#). Sin embargo, se ha demostrado que el software educativo tiene efectos positivos (por ejemplo, [Barrow, Markman, and Rouse \(2009\)](#); [Jackson and Makarin \(2018\)](#); [Naik, Chitre, Bhalla, and Rajan \(2020\)](#)), incluso después de períodos cortos (4-5 meses) de uso. La idea es que el software educativo permita un nivel de personalización de la enseñanza que sería imposible de lograr con un solo profesor para varios alumnos ([Banerjee, Cole, Duflo, & Linden, 2007](#); [Fairlie & Robinson, 2013b](#); [Kulik & Fletcher, 2016](#)). Además, cuando las habilidades de enseñanza son bajas (es decir, el conocimiento del contenido del profesor y las prácticas pedagógicas), el software educativo puede proporcionar una mejora con respecto al statu quo, que es el caso de las comunidades más pobres de todo el mundo. ([Bold et al., 2017](#); [Brunetti, Büchel, Jakob, Jann, & Steffen, 2023](#)).

Intervención

Conecta Ideas (CI) es una plataforma online que refuerza el aprendizaje matemático. En este caso, la intervención se centra en alumnos de 4° de primaria. Las sesiones de Conecta Ideas se llevan a cabo en los laboratorios de informática de los centros educativos en una sesión semanal de 90 minutos, con el apoyo de profesores y un coordinador externo.¹ En una sesión típica de IC, los estudiantes trabajan para resolver de 20 a 30 ejercicios que se les asignan, que están alineados con los temas cubiertos en la instrucción regular de matemáticas esa semana. Los estudiantes reciben retroalimentación automática sobre si sus respuestas son correctas. Más allá de una simple etiqueta correcta/incorrecta, el software explica por qué la respuesta es incorrecta y cómo llegar a la solución correcta. Los coordinadores de laboratorio, contratados y supervisados por el equipo de Conecta Ideas, son responsables de llevar a cabo las sesiones de aprendizaje en el laboratorio de computación en colaboración con los maestros de las aulas regulares. Los coordinadores son todos profesores de educación básica (aunque no

¹ Las clases, en teoría, eran de 90 minutos semanales, pero en algunos casos se acortaron a una hora lectiva (45 minutos), aunque hubo pocos casos. En algunas escuelas, hubo muchas clases perdidas debido a las vacaciones (hubo muchos lunes sin clases).

necesariamente eran profesores de matemáticas). Cada semana, los coordinadores de laboratorio, junto con los profesores, seleccionan los objetivos de aprendizaje y los ejercicios de la sesión para que estén alineados con lo que el profesor está cubriendo. Por lo tanto, cada escuela tratada cubrió diferentes temas y diferentes ejercicios cada semana. Por lo general, los ejercicios no se repiten en todas las sesiones a menos que los estudiantes progresen muy lentamente. Además, los coordinadores del laboratorio explicaron, al inicio de la clase, de qué se tratarían los ejercicios. Esto a veces no se hacía por falta de tiempo (los profesores de IC exigían que los niños realizaran un mínimo de 20 ejercicios por clase). Además, los maestros de escuela a veces se tomaban de 15 a 20 minutos adicionales al comienzo de las clases para dar una explicación tradicional. Los coordinadores de laboratorio recibieron un día de capacitación y supervisión continua por parte del equipo de implementación. Si bien los maestros no recibieron capacitación formal, el programa promueve el aprendizaje práctico.

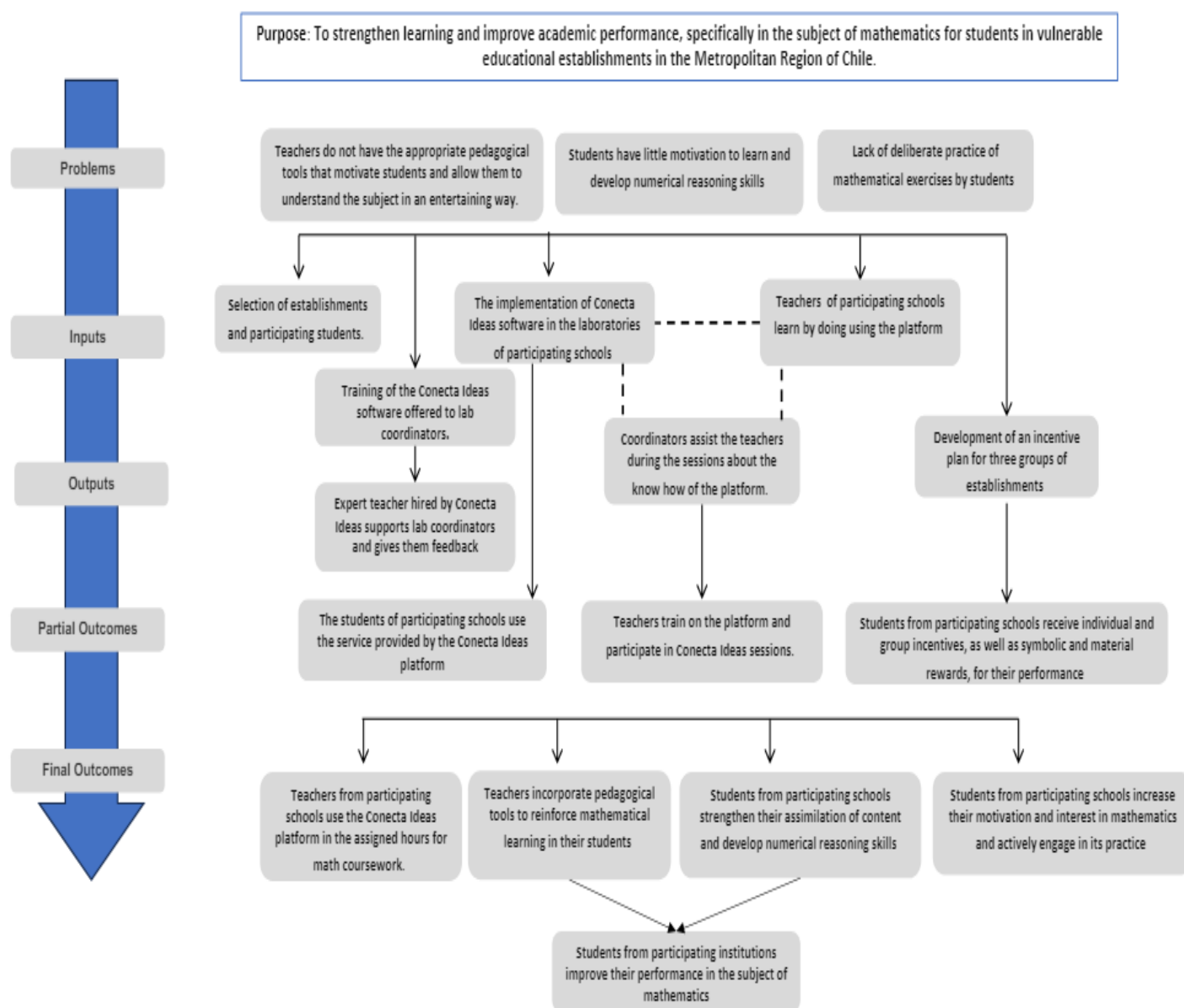
El software proporciona calcomanías virtuales y premios, incentivos simbólicos, a los estudiantes que obtienen buenos resultados en cada sesión de IC. También realiza un seguimiento del progreso de los estudiantes (en la sesión actual y en todas las sesiones) y ofrece a cada estudiante informes personalizados con gráficos fáciles de leer que muestran su progreso y cómo se comparan con sus compañeros.

El diseño de la evaluación nos permite probar el efecto de la implementación típica de Conecta Ideas (como se describió anteriormente) frente a un grupo de control. Además, nos permite medir el efecto de variar los detalles de implementación de Conecta Ideas (en particular, probando diferentes diseños de incentivos).

En algunas escuelas de tratamiento, además de la implementación típica de Conecta Ideas, a los estudiantes se les ofrecerán premios basados en el uso de su plataforma. En algunas escuelas, estos premios se basarán en el desempeño individual; en otros, en el desempeño de toda la clase, los estudiantes de la misma escuela compiten en grupo contra otras escuelas. Conecta Ideas no suele proporcionar estos incentivos tangibles, pero nos permite probar los efectos de proporcionar incentivos individuales tangibles (frente a los simbólicos) y de los incentivos individuales frente a los grupales. Los premios incluían cómics para colorear que explicaran STEM (ciencia, tecnología, ingeniería y matemáticas).² El diseño del incentivo individual será el siguiente: los incentivos simbólicos de los estudiantes servirán como "boletos de rifa". Por cada 100 preguntas correctas, los estudiantes ganaron un boleto. Cuantos más boletos, mayor será la posibilidad de ganar el sorteo. Por otro lado, cuantos más estudiantes tengan boletos en el aula, menor será la posibilidad de ganar el sorteo. Para el incentivo grupal, la lógica es la misma, pero las aulas ganan boletos de rifa proporcionales al número total de preguntas correctas en el aula. Los premios son los mismos, pero los gana toda una clase (compitiendo con las otras aulas en el sorteo).

² La evidencia cualitativa sugiere que estos premios no eran atractivos para los niños y los decepcionaban. Esto se discutirá más adelante en la Sección 5.

Figura 1: Teoría de cambio



Objetivos

- Evalúe el impacto de Conecta Ideas en estudiantes de 4° grado frente a un grupo de control (business as usual) en dos grupos principales de resultados:
 - **Resultados primarios:** Puntajes estandarizados de las pruebas estandarizadas de matemáticas y lenguaje de SIMCE de los estudiantes.
 - **Resultados secundarios:** Índices creados con herramientas estadísticas de análisis multicomponente y teoría de respuesta al ítem de valoraciones subjetivas de comportamientos relacionados con la escuela de estudiantes, maestros y tutores.

- Los índices de los estudiantes se refieren a la autopercepción en la escuela, la percepción que los profesores tienen de ellos en la escuela, la motivación en la escuela y la capacidad digital.
 - Los índices de los profesores se refieren a sus percepciones sobre el comportamiento en clase, su propia capacidad digital y la integridad de sus clases de matemáticas durante el año escolar.
 - Y el índice de tutores está relacionado con la ayuda que dan a sus hijos.
- Evalúe el impacto de dar incentivos tangibles no monetarios basados en el uso de Conecta Ideas (en relación con la implementación simple de Conecta Ideas que consiste en incentivos simbólicos).
 - Evaluar los incentivos individuales vs grupales (tangibles) en el Contexto de Conecta Ideas.

Metodología

Evaluación de impacto

Diseño metodológico

La evaluación de impacto utiliza un diseño de tratamiento experimental aleatorizado de tres brazos en el que los participantes se asignan a un grupo de control o de tratamiento. Por lo tanto, no hay sesgos sistemáticos entre el grupo de tratamiento y el grupo de control antes de comenzar el tratamiento. Por lo tanto, cualquier diferencia sistemática que observemos después de la intervención puede atribuirse al tratamiento.

De manera más explícita, entre las escuelas interesadas (Sección [3.1.2](#) sobre cómo se construyó la muestra experimental) aleatorizamos la asignación del tratamiento. La aleatorización se llevó a cabo en Stata 17 en un proceso de dos etapas. En primer lugar, aleatorizamos a nivel de escuela, estratificando según si las escuelas tienen una o más secciones en el 4º grado y el cuartil de un índice que resumía la puntuación media de la Teoría de Respuesta al Ítem (IRT) y la desviación estándar de las puntuaciones dentro de una escuela.³ La inclusión de los cuartiles de la distribución del índice IRT en la estratificación aumenta la potencia del experimento, ya que el rendimiento basal es altamente predictivo del rendimiento final. La probabilidad de ser asignado al control fue de 0,4, mientras que la probabilidad de ser asignado a cada uno de los tres tratamientos (Ideas Conecta de vainilla simple, IC con incentivos individuales e IC con incentivos grupales) fue de 0,2. Después de que las escuelas fueron asignadas a uno de los cuatro grupos experimentales, de cada escuela de tratamiento, seleccionamos al azar una sola sección (en caso de que hubiera varias secciones en el 4º grado) para el tratamiento.

Tabla 3: Diseño experimental

Diseño del estudio		Ensayo aleatorizado (conglomerado y estratificado) con cuatro grupos (3 tratamientos+control)
Unidad de aleatorización		Escuela. Dentro de las escuelas tratadas, se trata un aula de 4º grado seleccionada al azar (si hay varias aulas).
Variables de estratificación		Maniqué de la escuela que tiene varias secciones de 4º grado y el desempeño de los estudiantes en el examen de referencia, que fue la misma prueba para todos los estudiantes (promediada en toda la escuela). El rendimiento se midió utilizando el parámetro latente de un modelo de la Teoría de Respuesta al Ítem (TRI) que estimó simultáneamente la capacidad del examinado y la dificultad de las preguntas. Específicamente, el estrato fue el cuartil del desempeño promedio de los estudiantes en la escuela.
	Variable	Logros matemáticos

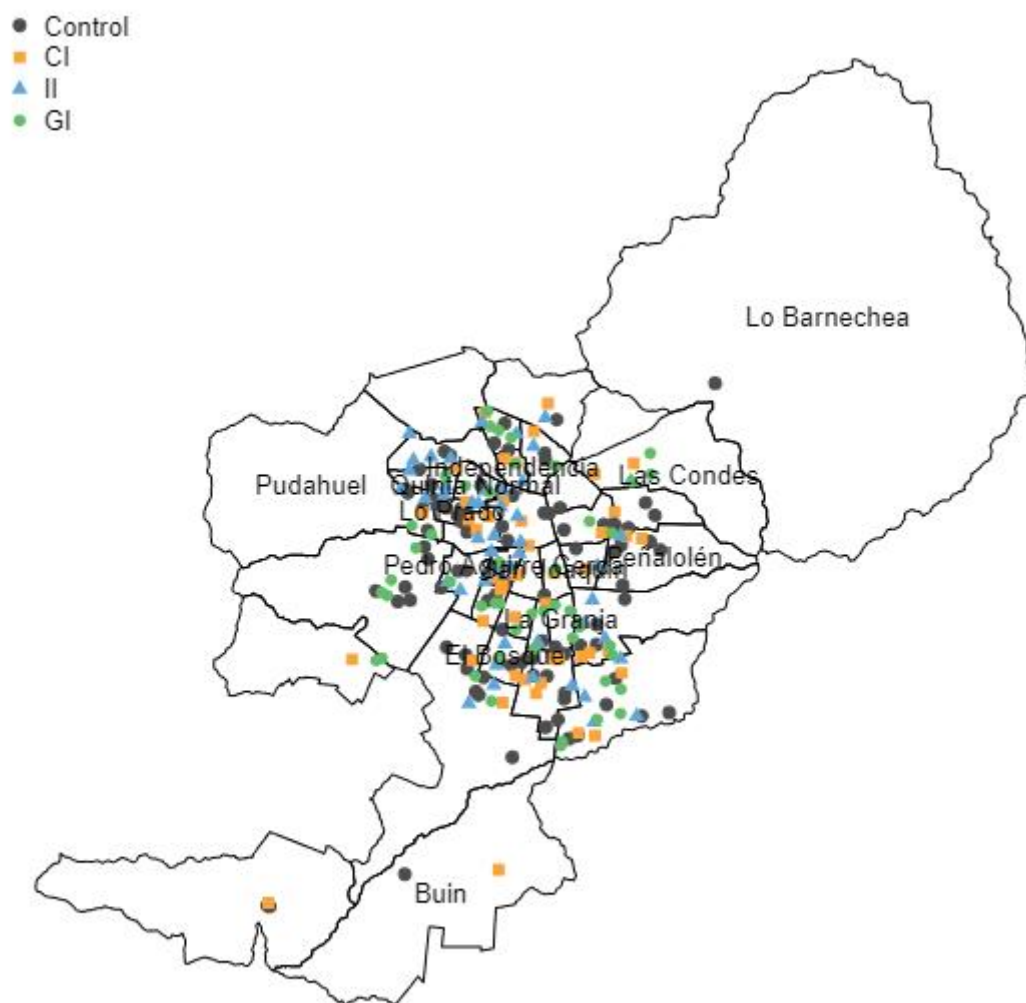
³ La teoría de respuesta al ítem (IRT) se refiere a una familia de modelos matemáticos que intentan explicar la relación entre los rasgos latentes (característica o atributo no observable) y sus manifestaciones (es decir, los resultados, las respuestas o el rendimiento observados). En este caso, lo usamos para estimar la "habilidad" o "nivel de conocimiento" del estudiante, que es el rasgo latente. Los resultados observados son las respuestas a los ítems de la prueba. (Columbia, 2024). Para obtener más información, consulte la sección [3.1.3](#).

Diseño del estudio		Ensayo aleatorizado (conglomerado y estratificado) con cuatro grupos (3 tratamientos+control)
Resultados primarios	Medición	Puntaje de la prueba estandarizada de Matemáticas del Sistema de Medicina de Calidad de la Educación (SIMCE). El SIMCE es una prueba estandarizada a nivel nacional realizada por el gobierno para medir la calidad de la educación. Esta puntuación se estandarizó para tener una media de cero y una desviación estándar de uno para el grupo de control. Vinculamos estos datos a nivel de estudiante con el puntaje de referencia utilizando el número de identificación nacional (RUT) del estudiante. Sin embargo, no todos los estudiantes que rindieron la prueba de referencia se encontraron en los datos del SIMCE. Por lo tanto, controlamos los puntajes promedio de las pruebas escolares durante la línea de base en nuestro análisis principal para evitar la caída de puntos de datos que no pudimos hacer coincidir entre los datos del SIMCE y la prueba de referencia. En una especificación alternativa, limitamos el análisis a las observaciones pareadas.
Resultados secundarios	Variable	Motivación y comportamiento de los estudiantes
	Medición	Esto se desprende de una encuesta que responden los estudiantes como parte del examen SIMCE
Línea de base para los resultados primarios	Variable	Logros matemáticos
	Medición	Puntaje de la prueba estandarizada de matemáticas, aplicado por la ONG Aptus
Valor basal para los resultados secundarios	Variable	Motivación de los estudiantes
	Medición	Encuesta aplicada por la ONG Aptus, que incluyó después de la prueba estandarizada de matemáticas. Creamos un índice (utilizando el Análisis de Componentes Principales (PCA)) que agrega las preguntas relacionadas con la motivación de los estudiantes en una sola medida.

Selección y características de los participantes

Nuestra muestra está formada por 235 centros educativos. Entre noviembre de 2021 y mediados de abril de 2022, AutoMind invitó a colegios del Área Metropolitana de Santiago a participar de la evaluación. Automind invitó a todas las escuelas públicas y privadas subvencionadas a menos de 40 minutos a pie de una estación de metro. En total, 1.211 escuelas recibieron una carta de invitación. AutoMind también intentó comunicarse con las escuelas por teléfono y correo electrónico (con una presentación y un video de YouTube (<https://youtu.be/suHjGu0Fis0>) que describe la intervención y el programa). Se obtuvo una carta de patrocinio del Ministerio de Educación para el proyecto y se mostró a las escuelas cuando se les invitó a participar en el proyecto. El equipo implementador también desarrolló un "Acuerdo de Participación para el Proyecto de Gamificación en Matemáticas con Conecta Ideas", que fue firmado digitalmente por cada director de escuela interesado que participa en el proyecto (ver https://docs.google.com/forms/d/e/1FAIpQLSe_6rd29PnvcIAKkEALfziDriCYs5i6TX9rAg6dW_MTmlMjMA/viewform). La Figura 2 muestra la distribución espacial de las escuelas en el experimento. En total, se asignaron 94 escuelas al grupo control, 47 al tratamiento Conecta Ideas (IC), 47 al tratamiento Incentivos Individuales (II) y 47 al tratamiento Incentivos Grupales (GI). En las escuelas de control, hay un total de 144 secciones; en las escuelas de IC hay 74 secciones, de las cuales 27 no son tratadas; en las escuelas II, hay 69 secciones, de las cuales 22 no están tratadas; y en las escuelas GI hay 81 secciones, de las cuales 34 no están tratadas.

Figura 2: Distribución escolar de tratamiento y control



Nota: Esta figura muestra la ubicación de las escuelas asignadas aleatoriamente a las Ideas Conectadas (IC), los Incentivos Individuales (II), los Incentivos Grupales (GI) o el grupo de control.

Al comparar los colegios del experimento con el resto de colegios públicos y privados subvencionados del Área Metropolitana de Santiago para participar en la evaluación, vemos que los colegios que optaron por participar en el estudio tienen más probabilidades de ser públicos, más propensos a tener matrícula cero, tener convenio con el Programa de Integración Escolar (PIE), y tener más estudiantes en riesgo de deserción escolar (Índice de Vulnerabilidad Escolar o IVE) (ver Tabla 2).⁴ Por lo tanto, los resultados del experimento pueden no extrapolarse al universo de las escuelas de Santiago

⁴ El Acuerdo PIE (Programa de Integración Escolar) es una estrategia inclusiva dentro del sistema educacional chileno. Su propósito es contribuir a la mejora continua de la calidad de la educación mediante la mejora de las experiencias de aprendizaje en el aula y la promoción de la participación de los estudiantes.

Tabla 4: Diferencias entre las escuelas del experimento y las que no participaron

Elemento	Otras escuelas (1)	En el experimento (2)	Diferencia (2)-(1)
Públicas	0.27 (0.44) [235]	0.52 (0.50) [235]	0.26*** (0.03)
PIE	[2,373] (0.49) [2,373]	0.81 (0.39) [235]	0.40*** (0.03)
Secular	0.67 (0.47) [2,373]	0.71 (0.45) [235]	0.05 (0.03)
Gratuitas	0.82 (0.38) [2,373]	0.89 (0.32) [235]	0.06*** (0.02)
Integración secundaria	0.32 (0.47) [2,373]	0.35 (0.48) [235]	0.03 (0.03)
Matrícula primaria	419.37 (296.57) [1,351]	424.17 (316.90) [235]	4.80 (22.16)
Estudiantes primarios IVE	0.80 (0.14) [1,351]	0.83 (0.12) [235]	0.03*** (0.01)
Comunidad IVE	0.79 (0.13) [1,351]	0.80 (0.12) [235]	0.01** (0.01)

Notas: La columna 1 muestra la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) para las escuelas que no están en la muestra experimental. La columna 2 muestra la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) para las escuelas de la muestra experimental. La columna 3 muestra la diferencia entre las escuelas del experimento y las que no lo están, así como el error estándar de la diferencia (entre paréntesis). El Acuerdo PIE (Programa de Integración Escolar) es una estrategia inclusiva dentro del sistema educacional chileno. Su propósito es contribuir a la mejora continua de la calidad de la educación mediante la mejora de las experiencias de aprendizaje en el aula y la promoción de la participación de los estudiantes. IVE son las siglas de 'Índice de Vulnerabilidad Escolar' (ver https://centroestudios.mineduc.cl/wp-content/uploads/sites/100/2019/12/DOC-DE-TRABAJO-19_Retenci%C3%B3n.pdf). Los errores estándar, agrupados en el nivel de sección, están entre paréntesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Medición de resultados

Referencia

La ONG Aptus realizó pruebas a los estudiantes antes de que comenzara la intervención de Conecta Ideas. Aplicaron la misma prueba a todos los estudiantes y escuelas. Los datos de referencia (recopilados por APTUS), combinados con los datos administrativos de la escuela, son la entrada para la asignación aleatoria estratificada (véase la Sección 3.1.1). APTUS recopiló datos entre el 28 de abril y el 17 de mayo de 2022. En total, los estudiantes fueron evaluados en las escuelas.⁵

En total, los estudiantes respondieron 35 preguntas de matemáticas. Además, los estudiantes respondieron 18 preguntas sobre su nivel socioeconómico, el acceso a las tecnologías de la información y la comunicación (por ejemplo, Internet y una computadora) y sus opiniones subjetivas sobre la escuela y las matemáticas. En el apéndice C se reproduce la prueba completa. Las respuestas a estas preguntas se utilizaron para comprobar el equilibrio previo al tratamiento en la asignación aleatoria.

Escalamos las puntuaciones de las pruebas utilizando métodos de la Teoría de Respuesta al Ítem (IRT). Los modelos IRT son el estándar en la literatura de evaluación. Estiman simultáneamente la capacidad del examinado y la dificultad de las preguntas, lo que permite que la contribución de las "respuestas correctas" a la medida de la capacidad varíe de una pregunta a otra.⁶ Estimamos un modelo logístico de tres parámetros (3PL) utilizando el paquete *mirt* en R (Chalmers, 2012), ya que la evaluación consta de preguntas de opción múltiple. Siguiendo la práctica estándar, normalizamos las puntuaciones de la IRT para que tengan una media de cero y una desviación estándar de uno. En la Figura 3 se presenta la distribución de las puntuaciones de la TRI al inicio del estudio.

Realizamos varias pruebas para comprobar la validez de la prueba. En primer lugar, el alfa de Cronbach (1951) (una medida de la correlación entre sí de los elementos de prueba) es 0.87. Se considera que una prueba tiene buena consistencia interna si el alfa está por encima de 0.9. Sin embargo, esta es una heurística, y dado que el examen no es una prueba alta para los estudiantes, un valor de es adecuado para nuestros propósitos (Nunnally, 1994; Bland & Altman, 1997; DeVellis, 2016).

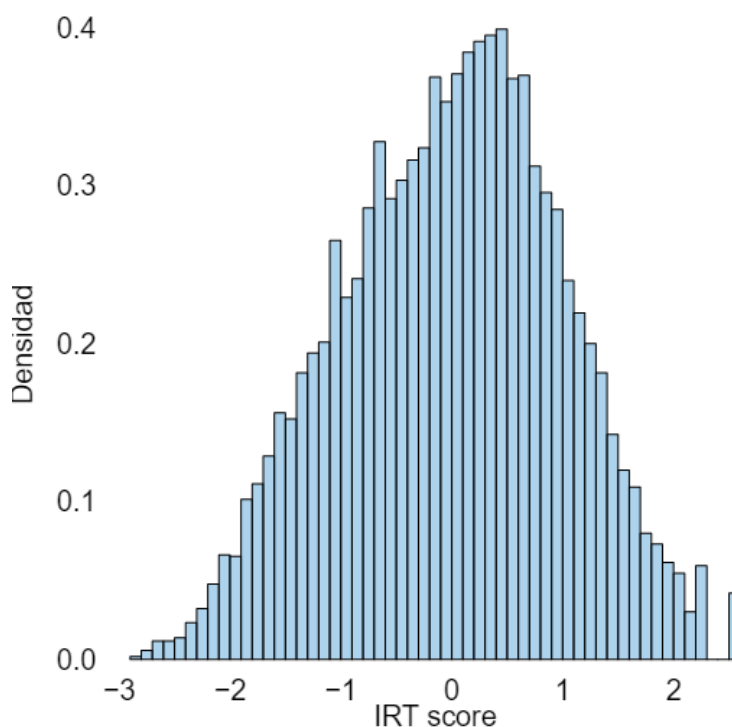
En segundo lugar, probamos las propiedades del modelo IRT. También hay un buen ajuste empírico con las curvas características de los ítems estimadas del modelo IRT (Figura C.8), y la función de información de la prueba alcanza su punto máximo alrededor de la puntuación mediana de la IRT (Figura C.9). Además, el % de las preguntas tienen un parámetro discriminatorio —que mide si una pregunta puede diferenciar a sujetos con alta y baja capacidad— considerado adecuado (De Ayala, 2013). Además, % tienen una capacidad alta o muy alta para

⁵ Además, hay datos de estudiantes que no respondieron ni una sola pregunta de la prueba. Omitimos estas observaciones de los datos. Una escuela (RBD 26110) fue excluida de los datos y del experimento (aunque una de sus secciones fue probada) ya que no nos permitió evaluar todas las secciones de 4º grado de la escuela.

⁶ Por ejemplo, los modelos IRT se utilizan para estimar la capacidad de los estudiantes en los Exámenes de Registro de Graduados (GRE), la Prueba de Evaluación Académica (SAT), el Programa para la Evaluación Internacional de Estudiantes (PISA), el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS) y las evaluaciones del Estudio Internacional de Progreso en Alfabetización Lectora (PIRLS). El uso de modelos IRT también es cada vez más común en estudios en países en desarrollo que se basan en evaluaciones de estudiantes desarrolladas por investigadores (p. ej. g. Das and Zajonc (2010); Bau and Das (2020); Muralidharan et al. (2019); Singh (2020); Mbiti, Muralidharan, et al. (2019); Mbiti, Romero, and Schipper (2019); Romero, Sandefur, and Sandholtz (2020); Romero and Sandefur (2021)). Das and Zajonc (2010)). proporciona una introducción a los modelos IRT, a la vez que van der Linden (2017) proporciona un tratamiento completo de los modelos IRT.

diferenciar entre sujetos con diferentes conocimientos (Baker & Kim, 2017). Todas estas medidas sugieren que la prueba es adecuada para nuestros propósitos y es capaz de medir las habilidades de los estudiantes. El Apéndice C contiene más detalles sobre las propiedades psicométricas del examen, así como una reproducción completa de las preguntas de la prueba.

Figura 3: Distribución de las puntuaciones de la IRT



Nota: Distribución de las puntuaciones de IRT al inicio del estudio.

Datos administrativos de la escuela

Las características administrativas de los colegios de nuestra muestra provienen de la información pública, publicada anualmente en el Directorio de Establecimientos Educacionales.⁷ Esta información se resume en la Tabla 5. Esta base de datos tiene información sobre si la escuela es pública, privada subvencionada o privada no subvencionada. También cuenta con información sobre el Programa de Integración Escolar (PIE). La mayoría de las escuelas de nuestra muestra cuentan con este programa, donde los estudiantes con discapacidad se integran al sistema educativo regular, compartiendo la misma aula con el resto de los estudiantes. También tenemos información sobre el Índice de Vulnerabilidad Escolar (IVE). Esta variable aproxima el riesgo de deserción de los estudiantes a través de una evaluación socioeconómica de los mismos. Los valores de IVE oscilan entre 0 y 1, reflejando la proporción de estudiantes en situación de vulnerabilidad. Este índice se construye para cada municipio, así como para cada escuela, distinguiendo entre estudiantes de primaria y secundaria. En nuestra muestra, más del 80% de los estudiantes son vulnerables, según esta medida.

Datos primarios de resultados: Puntajes de las pruebas SIMCE

Utilizamos una prueba estandarizada nacional para medir el dominio de los estudiantes en matemáticas y lenguaje. El Sistema de Medición de la Calidad Educativa (SIMCE) fue creado en 1988. Mide los resultados de aprendizaje de los estudiantes en diferentes grados y áreas como matemáticas, lenguaje y ciencias sociales y naturales, entre otras (Canales & Webb, 2018). Cada año, el SIMCE se aplica a todos los estudiantes de 4º grado. En 2020 y 2021, el SIMCE fue suspendido debido a la pandemia de COVID-19.

⁷ Directorio de Establecimientos Educacionales 2021. Centro de Estudios MINEDUC Chile. Obtenido de <https://datosabiertos.mineduc.cl/directorio-de-establecimientos-educacionales/>

En 2022, el SIMCE regresó y se aplicó a los alumnos de 4° básico de educación primaria. La prueba de 4° grado se llevó a cabo el 16 de noviembre y evaluó los objetivos de aprendizaje de 1° a 4° grado (Agencia de Calidad de la Educación, 2022). Tenía una duración máxima de 90 minutos y una mínima de 45 minutos.

Datos secundarios de resultados: encuesta SIMCE

Estudiantes, maestros, directores y padres de familia responden una encuesta para el examen SIMCE. Utilizamos estas respuestas a la encuesta para medir los resultados secundarios.

Para los estudiantes, construimos cuatro índices que capturan la autopercepción sobre su capacidad, la percepción del estudiante sobre la opinión del profesor sobre ellos, la motivación y la capacidad digital. Para los índices de autopercepción, motivación y habilidades digitales se construyeron dos tipos de índices. El conjunto completo de preguntas utilizadas para construir cada índice se enumera en la Sección 4.5. El primero se basó en el Análisis de Correspondencias Múltiples, que es una extensión del Análisis de Correspondencias para analizar el patrón de relaciones de varias variables dependientes categóricas. Estimamos estas puntuaciones utilizando la función MCA del paquete FactoMineR en R (Le, Josse, & Husson, 2008). El segundo tipo se realizó con IRT al igual que con la prueba basal APTUS. Estimamos un modelo logístico de tres parámetros (3PL) utilizando el paquete mirt en R (Chalmers, 2012). Para los otros dos índices, estimamos solo el índice de Análisis de Correspondencias Múltiples, ya que no hay suficientes componentes para estimar el índice IRT⁸.

En el caso de los docentes, se construyeron índices para tres categorías: comportamiento de los estudiantes, capacidad digital y objetivos de enseñanza de matemáticas establecidos por el Ministerio de Educación de Chile. Para esto último, creamos una variable que mide la proporción de estos objetivos de enseñanza de matemáticas que los profesores alcanzaron durante el año escolar. El conjunto completo de preguntas utilizadas para construir cada índice se enumera en la Sección 4.5. Para los objetivos de la enseñanza de la habilidad digital y las matemáticas, construimos los dos tipos de índices tal como lo hicimos con los índices de los estudiantes, es decir, con los métodos MCA e IRT. No podemos estimar un modelo IRT para la categoría de comportamiento del estudiante ya que no hay suficientes variables en el índice.

Para los padres, construimos índices para el apoyo de los padres a los estudiantes en tareas relacionadas con la escuela. El conjunto completo de preguntas utilizadas para construir cada índice se enumera en la Sección 4.5. No fue posible estimar un índice IRT para este índice.

Tamaño de la muestra

Como se mencionó anteriormente, hay escuelas en el experimento. Este número estuvo determinado principalmente por el número de escuelas que estaban interesadas en el programa. Sin embargo, son suficientes para tener la potencia adecuada para detectar los efectos del tratamiento en el experimento. Los cálculos de potencia ex ante fueron los siguientes: Sobre la base de los datos de APTUS de la línea de base, la correlación intraconglomerado para los estudiantes dentro de la misma escuela es .14. Para los estudiantes dentro de la misma escuela y sección, es .145. En total, inicialmente teníamos la intención de probar e incluir en el experimento 242 escuelas. Sin embargo, de estos, realizamos la prueba basal en las escuelas. Se aleatorizó el tratamiento entre estas escuelas. Por lo tanto, tuvimos una tasa de deserción a nivel escolar de aproximadamente el 2.9% de las escuelas. Suponemos que la misma proporción de escuelas se desgastará al final.

A nivel estudiantil, aproximadamente el 15% de los estudiantes matriculados en escuelas no fueron evaluados al inicio del estudio. También asumiremos que esta tendencia continúa y que además tendremos otro 15% de deserción por parte de los estudiantes. Dado que en la sección promedio evaluamos a los estudiantes al inicio, esperamos tener aproximadamente 24 estudiantes por sección al final.

Al agruparse a nivel de escuela, el efecto mínimo detectable (MDE) es igual a

$$MDE = (t_{\alpha/2} + t_{1-\beta}) \left(\frac{\sqrt{\rho + \frac{1-\rho}{n}}}{\sqrt{P(1-P)J}} \right) \sigma$$

⁸ En estos casos, hay 3 variables, lo que lleva a 9 patrones de respuesta diferentes ya que las variables son dicotómicas). Sin embargo, el modelo debe estimar 9 parámetros (una pendiente para cada pregunta, una intersección y el parámetro de adivinación). Esto lleva a un modelo no identificado (van der Linden, 2017).

donde α es el tamaño (o nivel de significación) de la prueba, β es la potencia, P es la proporción de escuelas tratadas, J es el número total de escuelas, ρ es la correlación intra-conglomerado y n es el número de estudiantes evaluados por escuela. Por último, está la varianza de la prueba σ (Duflo, Glennerster, & Kremer, 2007).

Contamos con tres grupos de tratamiento: Ideas Conecta (CI), Ideas Conecta con incentivos individuales (II) e Ideas Conecta con incentivos grupales (GI). Si asignamos 50 escuelas a cada uno de estos grupos, nos quedan escuelas que sirven como controles.

Basándonos en la prueba de referencia, asumimos que $\rho = 145$. Asumimos (de forma conservadora) que las covariables basales (p. ej., la puntuación de la prueba basal, la edad y el sexo) explican el 30% de la varianza y, por lo tanto, que $\sigma = 0.84 \approx \sqrt{1 - 0.3\rho}$. Por último, teniendo en cuenta la deserción a nivel de los estudiantes, asumimos $n = 24$.

Por lo tanto, podemos detectar un tamaño del efecto de al menos (MDE) $0,21\sigma$ con una potencia del 90% (β) utilizando una prueba del 5% de nivel (α) al comparar cualquiera de los tratamientos con el control. Al comparar dos grupos de tratamiento cualesquiera, tenemos un MDE de 0.23σ ⁹.

Ex-post, estamos capacitados para detectar efectos de tratamiento de $0,149\sigma$ (McKenzie & Ozier, 2019). Esto se estima utilizando el error estándar de la regresión principal (Tabla 13) y asumiendo una potencia del 90% y un nivel del 5%. Por lo tanto, los cálculos de poder ex-ante muestran que tenemos más poder del que pensábamos antes de la intervención. Esto se debe en parte a que ninguna escuela se desgastó a partir de los datos y tuvimos más estudiantes de los que esperábamos por escuela. Como referencia, la mediana del efecto de las intervenciones educativas es de 0,1 sobre el aprendizaje (Evans & Yuan, 2022). Sin embargo, la comparación de los efectos del tratamiento entre los estudios en las desviaciones tiene algunos escollos, ya que el tamaño de los efectos depende en gran medida del diseño de la prueba. (Singh, 2015).

Aleatorización

La aleatorización se realizó en Stata 17 MP. Específicamente, entre todas las escuelas en el experimento, estimamos el número de secciones en el Grado 4 y creamos un índice PCA basado en el promedio, la mediana y la desviación estándar de los puntajes de IRT al inicio del estudio. A continuación, agrupamos las escuelas en ocho grupos en función del cuartil del índice PCA: 4 grupos para las escuelas con una sola sección y 4 grupos para las escuelas con varias secciones. Esto definió los estratos de aleatorización (es decir, hay ocho estratos de aleatorización). A continuación, se utilizó el comando "randtreat" (Carril, 2015) para asignar escuelas a los cuatro grupos experimentales (los tres tratamientos y el control). La probabilidad de ser asignado a cada grupo de tratamiento fue de 1/5, mientras que la probabilidad de asignación al control fue de 2/5. Usamos como semilla "28022022".¹⁰ Para tratar los estratos desiguales (que no son divisibles por 4) se utilizó la opción wstrata, que asigna aleatoriamente los inadaptados de forma independiente en todos los estratos (utilizando la probabilidad de asignación especificada para cada tratamiento).

Después de las escuelas, asignamos aleatoriamente a los grupos de tratamiento, luego asignamos aleatoriamente una sección en cada escuela tratada al tratamiento. Específicamente, usamos el comando "gsample" (nuevamente con la semilla establecida en "28022022") para seleccionar aleatoriamente una sola sección de cada escuela: si la escuela tiene una sola sección, esa sección se elige de manera determinista mediante el comando "gsample".

En el apéndice E se reproduce el código de aleatorización.

Equilibrar

Las características observables de los estudiantes (Tabla 3) y de las escuelas (Tabla 5) son prácticamente las mismas en los cuatro grupos experimentales (los tres tratamientos y el control). Cerca de la mitad de los estudiantes son niñas, aunque los grupos presentan un ligero desequilibrio, y la edad promedio es de 9,3 años. Más del 80% de los estudiantes de nuestra muestra tienen una madre que al menos ha completado la educación secundaria, alrededor del 70% tiene una computadora en casa y más del 90% tiene internet en casa.

⁹ Con una potencia del 80%, el MDE entre cualquier tratamiento y el control es y entre dos tratamientos cualesquiera. $\sigma\sigma$

¹⁰ La fecha de nacimiento del hijo de uno de los autores. Las computadoras generan "aleatoriedad" utilizando generadores de números cuasialeatorios que necesitan una semilla para iniciar el proceso. Especificar la semilla nos permite replicar el proceso.

Tabla 5: Balance a nivel estudiante

Característica	Control	Conecta Ideas (CI)	Incentivos Individuales (II)	Incentivos Grupales (GI)	p-valor (igualdad)
Puntaje base (IRT)	-0.00	-0.06	-0.10	-0.04	0.840
(Desviación estándar)	(0.99)	(0.97)	(1.02)	(0.98)	
[Número de observaciones]	[3,979]	[1,217]	[1,295]	[1,251]	
Mujeres (%)	0.51	0.45	0.45	0.49	0.080*
(Desviación estándar)	(0.50)	(0.50)	(0.50)	(0.50)	
[Número de observaciones]	[3,745]	[1,138]	[1,214]	[1,169]	
Edad (01/01/2022)	9.30	9.32	9.31	9.33	0.730
(Desviación estándar)	(0.45)	(0.55)	(0.49)	(0.56)	
[Número de observaciones]	[3,381]	[1,016]	[1,084]	[1,041]	
Madre con secundaria completa	0.86	0.85	0.83	0.85	0.524
(Desviación estándar)	(0.34)	(0.36)	(0.38)	(0.36)	
[Número de observaciones]	[1,961]	[627]	[658]	[559]	
Computadora en casa	0.72	0.68	0.69	0.70	0.903
(Desviación estándar)	(0.45)	(0.47)	(0.46)	(0.46)	
[Número de observaciones]	[3,727]	[1,124]	[1,210]	[1,164]	
Internet en casa	0.93	0.91	0.93	0.93	0.486
(Desviación estándar)	(0.25)	(0.28)	(0.26)	(0.26)	

[Número de observaciones] [3,723] [1,124] [1,226] [1,178]

Notas* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Las columnas 1 a 4 muestran la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) para cada uno de los siguientes grupos: Control, estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta (IC), estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos individuales (II) y estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos grupales (GI). En el grupo de tratamiento, solo se incluyen los datos de las secciones tratadas. Las columnas 5 muestran el valor p de la prueba de si la media es la misma en todos los grupos. La prueba de igualdad considera el diseño de aleatorización (es decir, incluye efectos fijos de estratos). Los errores estándar, agrupados a nivel de sección escolar, están entre paréntesis. En el cuadro B.1 se muestran datos más detallados sobre las diferencias entre los grupos.

El cuestionario de referencia también incluyó algunas preguntas sobre las autopercepciones y aspiraciones de los estudiantes (véase la Tabla 4). Más del 40% de los estudiantes aspiran a la educación terciaria (y el resto aspira a graduarse de la escuela secundaria o con metas poco claras). En particular, la mayoría de los estudiantes disfrutan aprendiendo matemáticas (~90%) o piensan que aprender matemáticas es fácil (~75%); Sin embargo, la mayoría de ellos se preocupan por tener malas calificaciones en matemáticas (~84%) e informan que las pruebas de matemáticas los ponen nerviosos (~60%). Pocos estudiantes (menos del 7%) reportan faltar a clases o que su comportamiento molestó al maestro.¹¹

Tabla 6: Percepciones basales de los estudiantes

Percepción	Control	CI	II	GI	Valor p (igualdad)
Aspira a educación terciaria	0.45	0.45	0.44	0.43	0.457
Desviación estándar	(0.50)	(0.50)	(0.50)	(0.50)	
N	[3,913]	[1,193]	[1,275]	[1,228]	
Capaz de aprender	0.85	0.84	0.83	0.84	0.424
Desviación estándar	(0.35)	(0.37)	(0.38)	(0.36)	
N	[3,928]	[1,193]	[1,272]	[1,231]	
Aprende rápido	0.80	0.77	0.81	0.79	0.016**
Desviación estándar	(0.40)	(0.42)	(0.39)	(0.41)	
N	[3,933]	[1,191]	[1,276]	[1,225]	
Esfuerzo = Éxito	0.92	0.91	0.93	0.92	0.449

¹¹ Hay algunas diferencias significativas en las características de los estudiantes entre los grupos. Por ejemplo, hay una diferencia significativa (~2%) entre los grupos en los estudiantes de aprendizaje rápido. Además, existe una diferencia (~4%) entre los grupos en la proporción de estudiantes que consideran que las matemáticas son un tema fácil. Hay una diferencia de (~3%) entre los grupos en la proporción de estudiantes que se ponen nerviosos en las pruebas de matemáticas. Y, por último, la mayor diferencia en la proporción de estudiantes que disfrutan aprendiendo matemáticas es del 2%. Sin embargo, no podemos rechazar la hipótesis nula en una prueba conjunta de que los grupos están globalmente equilibrados en todas las características.

Desviación estándar	(0.27)	(0.29)	(0.26)	(0.27)	
N	[3,915]	[1,194]	[1,270]	[1,240]	
Matemáticas es fácil	0.78	0.74	0.76	0.78	0.037**
Desviación estándar	(0.41)	(0.44)	(0.43)	(0.42)	
N	[3,924]	[1,197]	[1,270]	[1,232]	
Disfruta aprender matemáticas	0.90	0.90	0.92	0.92	0.062*
Desviación estándar	(0.30)	(0.30)	(0.27)	(0.28)	
N	[3,933]	[1,195]	[1,276]	[1,230]	
Las pruebas de matemáticas me ponen nervioso	0.63	0.61	0.65	0.66	0.045**
Desviación estándar	(0.48)	(0.49)	(0.48)	(0.47)	
N	[3,927]	[1,196]	[1,271]	[1,220]	
Se preocupa por las malas notas en matemáticas	0.84	0.84	0.84	0.86	0.482
Desviación estándar	(0.37)	(0.37)	(0.37)	(0.35)	
N	[3,930]	[1,196]	[1,277]	[1,232]	
Disfruta ir a la escuela	0.83	0.82	0.82	0.82	0.410
Desviación estándar	(0.37)	(0.38)	(0.38)	(0.39)	
N	[3,953]	[1,202]	[1,288]	[1,241]	
Esfuerzo en las tareas	0.88	0.87	0.87	0.88	0.427
Desviación estándar	(0.32)	(0.34)	(0.34)	(0.33)	
N	[3,940]	[1,199]	[1,284]	[1,232]	

Faltas a clase	0.06	0.06	0.07	0.07	0.770
Desviación estándar	(0.25)	(0.24)	(0.26)	(0.25)	
N	[3,908]	[1,195]	[1,278]	[1,233]	
Molesta al maestro	0.07	0.09	0.08	0.09	0.310
Desviación estándar	(0.25)	(0.28)	(0.28)	(0.28)	
N	[3,921]	[1,203]	[1,282]	[1,234]	

Notas: Las columnas 1 a 4 muestran la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) para cada uno de los siguientes grupos: Control, estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta (IC), estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos individuales (II) y estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos grupales (GI). En el grupo de tratamiento, solo se incluyen los datos de las secciones tratadas. La columna 5 muestra el valor p de la prueba de si la media es la misma en todos los grupos. Las preguntas originales de la encuesta se reproducen en el Apéndice C. Capaz de aprender es igual a 1 si el estudiante está de acuerdo o completamente de acuerdo en que es capaz de aprender lo que estudia, incluso las materias más difíciles. Aprendiz rápido es igual a 1 si el estudiante está de acuerdo o completamente de acuerdo en que aprende rápidamente la mayoría de las materias. Esfuerzo = Éxito es igual a 1 si el estudiante está de acuerdo o completamente de acuerdo, si se esfuerza, tendrá éxito en la escuela. Las matemáticas son fáciles es igual a 1 si el estudiante está de acuerdo o completamente de acuerdo en que las matemáticas son fáciles para él. Disfruta aprendiendo matemáticas es igual a 1 si el estudiante está de acuerdo o completamente de acuerdo en que disfruta aprendiendo matemáticas. Las pruebas de matemáticas me ponen nervioso es igual a 1 si el estudiante está de acuerdo o completamente de acuerdo con que las pruebas de matemáticas lo ponen nervioso. La preocupación por las malas calificaciones en matemáticas es igual a 1 si el estudiante está de acuerdo o completamente de acuerdo en que se preocupa por tener malas calificaciones en matemáticas. Disfruta ir a la escuela es igual a 1 si el estudiante siempre o casi siempre disfruta ir a la escuela. El esfuerzo en la tarea es igual a 1 si el estudiante siempre o casi siempre hace mucho esfuerzo para hacer su tarea. Faltar a clase es igual a 1 si el estudiante siempre o casi siempre falta a clase. Molesta al profesor es igual a 1 si el alumno siempre o casi siempre siente que su comportamiento molesto a su profesor. La prueba de igualdad considera el diseño de aleatorización (es decir, incluye efectos fijos de estratos). Los errores estándar, agrupados a nivel de sección escolar, están entre paréntesis. En el cuadro B.1 se muestran datos más detallados sobre las diferencias entre los grupos. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

La mayoría de las escuelas de nuestra muestra son públicas (55%), no cobran tasas de matrícula (89%) y no son administradas por instituciones religiosas (70%) (ver Tabla 5).

Tabla 7: Balance de nivel escuela

Característica	Control	CI	II	GI	Valor p
Secciones	1.53	1.57	1.47	1.72	0.124
Desviación estándar	(0.73)	(0.90)	(0.58)	(1.19)	
N	[94]	[47]	[47]	[47]	0.790
Pública	0.54	0.55	0.47	0.51	
Desviación estándar	(0.50)	(0.50)	(0.50)	(0.51)	
N	[94]	[47]	[47]	[47]	0.704
PIE	0.81	0.83	0.77	0.85	
Desviación estándar	(0.40)	(0.38)	(0.43)	(0.36)	
N	[94]	[47]	[47]	[47]	0.302
Laica	0.76	0.72	0.60	0.72	
Desviación estándar	(0.43)	(0.45)	(0.50)	(0.45)	
N	[94]	[47]	[47]	[47]	0.800
Gratuita	0.88	0.91	0.89	0.85	
Desviación estándar	(0.32)	(0.28)	(0.31)	(0.36)	
N	[94]	[47]	[47]	[47]	0.446
Educación secundaria integrada	0.30	0.43	0.38	0.34	
Desviación estándar	(0.46)	(0.50)	(0.49)	(0.48)	
N	[94]	[47]	[47]	[47]	0.389
Inscripción primaria	411.80	432.45	394.40	470.38	
Desviación estándar	(306.57)	(343.50)	(199.05)	(399.68)	
N	[94]	[47]	[47]	[47]	0.543
Estudiantes primarios IVE	0.84	0.82	0.83	0.82	
Desviación estándar	(0.12)	(0.13)	(0.11)	(0.11)	
N	[94]	[47]	[47]	[47]	0.311
Comunidad IVE	0.81	0.80	0.81	0.78	
Desviación estándar	(0.08)	(0.08)	(0.07)	(0.08)	

N [94] [47] [47] [47]

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Las columnas 1 a 4 muestran la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) para cada uno de los siguientes grupos: Control, estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta (IC), estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos individuales (II) y estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos grupales (GI). En el grupo de tratamiento, solo se incluyen los datos de las secciones tratadas. Las columnas 5 muestran el valor p de la prueba de si la media es la misma en todos los grupos. La prueba de igualdad considera el diseño de aleatorización (es decir, incluye efectos fijos de estratos). Los errores estándar, agrupados a nivel de escuela, están entre paréntesis. . En el cuadro B.2 se muestran datos más detallados sobre las diferencias entre los grupos.

Al seleccionar aleatoriamente una sección para ser tratada en escuelas con múltiples secciones, también verificamos que las características de los estudiantes en las secciones tratadas son similares a las de las secciones no tratadas en las escuelas de tratamiento (Tabla 6), y que las características de los estudiantes en las secciones no tratadas en las escuelas de tratamiento son similares a las de los estudiantes de control (Tabla 7).

Tabla 8: Estudiantes no tratados vs tratados en escuelas tratadas

Característica	Estudiantes no tratados	Diferencia sección tratada
Puntaje base (IRT)	0.10	-0.02
(Desviación estándar)	(1.02)	(0.03)
[Número de observaciones]	[2,505]	[6,268]
Mujeres	0.48	-0.02
(Desviación estándar)	(0.50)	(0.02)
[Número de observaciones]	[2,354]	[5,875]
Edad (01/01/2022)	9.27	0.02*
(Desviación estándar)	(0.45)	(0.01)
[Número de observaciones]	[2,160]	[5,300]
Madre con secundaria completa	0.86	0.00
(Desviación estándar)	(0.34)	(0.01)

[Número de observaciones]	[1,257]	[3,101]
Computadora en casa	0.75	0.02*
(Desviación estándar)	(0.43)	(0.01)
[Número de observaciones]	[2,348]	[5,846]
Internet en casa	0.94	-0.00
(Desviación estándar)	(0.23)	(0.01)
[Número de observaciones]	[2,394]	[5,922]

Notas: La columna 1 muestra la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) de los estudiantes no tratados en las escuelas tratadas. La columna 2 muestra la diferencia entre los estudiantes tratados y los no tratados en las escuelas tratadas, así como el error estándar de la diferencia (entre paréntesis) y el número de observaciones utilizadas para la estimación (entre corchetes). Las diferencias entre los grupos (Columna 2) controlan los efectos fijos de la escuela. Los errores estándar, agrupados a nivel de sección escolar, están entre paréntesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tabla 9: Verificación de balance para estudiantes no tratados en escuelas tratadas

Característica	Control	Conecta Ideas (CI)	Incentivos Individuales (II)	Incentivos Grupales (GI)	p-valor (igualdad)
Puntaje base (IRT)	-0.00	-0.01	0.09	0.20	0.507
(Desviación estándar)	(0.99)	(1.01)	(1.05)	(1.01)	
[Número de observaciones]	[3,979]	[815]	[661]	[1,029]	
Mujeres	0.51	0.48	0.51	0.47	0.099*
(Desviación estándar)	(0.50)	(0.50)	(0.50)	(0.50)	
[Número de observaciones]	[3,745]	[764]	[633]	[957]	
Edad (01/01/2022)	9.30	9.25	9.29	9.27	0.238
(Desviación estándar)	(0.45)	(0.41)	(0.45)	(0.48)	

[Número de observaciones]	[3,381]	[686]	[563]	[911]	
Madre con secundaria completa	0.86	0.83	0.87	0.88	0.412
(Desviación estándar)	(0.34)	(0.37)	(0.34)	(0.32)	
[Número de observaciones]	[1,961]	[413]	[333]	[511]	
Computadora en casa	0.72	0.70	0.75	0.80	0.224
(Desviación estándar)	(0.45)	(0.46)	(0.44)	(0.40)	
[Número de observaciones]	[3,727]	[757]	[617]	[974]	
Internet en casa	0.93	0.93	0.94	0.95	0.759
(Desviación estándar)	(0.25)	(0.25)	(0.24)	(0.21)	
[Número de observaciones]	[3,723]	[778]	[622]	[994]	

Notas: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Las columnas 1 a 4 muestran la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) para cada uno de los siguientes grupos: Control, estudiantes en secciones no tratadas en escuelas tratadas con Ideas Conecta (IC), estudiantes en secciones no tratadas en escuelas tratadas con Ideas Conecta + incentivos individuales (II) y estudiantes en secciones no tratadas en escuelas tratadas con Ideas Conecta + incentivos grupales (GI). En el grupo de tratamiento, solo se incluyen los datos de las secciones no tratadas. La columna 5 muestra el valor p de la prueba de si la media es la misma en todos los grupos. La prueba de igualdad considera el diseño de aleatorización (es decir, incluye efectos fijos de estratos). Los errores estándar, agrupados a nivel de sección escolar, están entre paréntesis. En el cuadro B.3 se muestran datos más detallados sobre las diferencias entre los grupos.

Análisis estadístico

Análisis primario

Estimaremos el efecto de los tres tratamientos diferentes frente al control por mínimos cuadrados ordinarios de la siguiente manera:

$$Y_{it} = \beta_1 CIs + \alpha_1 CIs \times TreatedSection_{it} + \beta_2 IIs + \alpha_2 IIs \times TreatedSection_{it} + \beta_3 GIs + \alpha_3 GIs \times TreatedSection_{it} + \gamma_s + \lambda X_i + \phi X_s + \varepsilon_{it}$$

En donde Y_{it} representa el resultado de los estudiantes en las escuelas, γ_s son los efectos fijos de los estratos de aleatorización (es decir, si las escuelas tienen una o más secciones en el 4° grado y el cuartil de un índice que resume el puntaje promedio de IRT y la desviación estándar de los puntajes dentro de una escuela), X_i son características individuales (p. ej., puntajes iniciales, sexo y edad) y X_s son características invariantes en el tiempo a nivel escolar (p. ej., desempeño promedio de la escuela en el SIMCE en el pasado, nivel socioeconómico e instalaciones). CIs indica si a la escuela se le ofreció la implementación simple, IIs si se le ofrecieron incentivos a nivel individual y GIs si se le ofrecieron incentivos a nivel grupal. $TreatedSection_{it}$ indica

si la sección en la que está matriculado el estudiante está tratada (en las escuelas tratadas). Agrupamos los errores estándar a nivel escolar (que es el nivel en el que se asignó el tratamiento).

Nuestros coeficientes de interés son, α_1 , α_2 , y α_3 , que nos permiten probar el efecto de diferentes implementaciones de Conecta Ideas frente al grupo control, así como β_1 - β_3 , que miden cualquier efecto derrame sobre la sección no tratada en las escuelas tratadas. Tenga en cuenta que estas son estimaciones por intención de tratar (ITT, por sus siglas en inglés), ya que no todas las escuelas a las que se les ofreció el tratamiento lo toman.

Además, podemos estimar el efecto adicional (por encima del efecto de la Idea Conecta simple) de los incentivos individuales (estimando la diferencia $\alpha_2 - \alpha_1$), el efecto adicional de los incentivos grupales (estimando la diferencia $\alpha_3 - \alpha_1$). Por último, podemos comparar el efecto de los incentivos grupales frente a los individuales (estimando la diferencia $\alpha_3 - \alpha_2$).

Las variables de control que utilizamos son:

- El puntaje promedio estandarizado de matemáticas de la escuela de la prueba SIMCE de referencia de 2018 (nivel escolar).
- El puntaje promedio estandarizado de la escuela a partir de la prueba SIMCE de referencia de 2018 (nivel escolar).
- La puntuación media del IRT de la escuela a partir de la prueba APTUS basal realizada antes de la intervención (nivel escolar).
- Variable dicotómica que indica con uno si la escuela no rindió la prueba SIMCE 2018 y con cero en caso contrario (nivel escolar).
- Número de estudiantes de la escuela que rindieron la prueba estandarizada de matemáticas SIMCE 2022.
- Variable dicotómica que indica con uno si se trata de una escuela rural y con cero en caso contrario (nivel escolar).
- Variable dicotómica que indica con uno si se trata de un colegio público y con cero en caso contrario (nivel escolar).
- Variable dicotómica que indica con uno si se trata de un colegio subvencionado en particular y con cero en caso contrario (nivel escolar).
- Variable dicotómica que indica con uno si la escuela es de NSE bajo y con cero en caso contrario (nivel escolar).
- Edad del estudiante cuando se realizó el examen.
- Género del estudiante (dummy con uno que indica estudiante mujer, cero en caso contrario).

Análisis secundario

Utilizaremos la misma ecuación de estimación que para los puntajes del SIMCE para estudiar el efecto del tratamiento sobre los datos de la encuesta SIMCE. La única diferencia es que los resultados (Y_{is}) provendrán de las respuestas a la encuesta de los niños, los maestros y los padres.

Cumplimiento imperfecto

Dado que algunas escuelas (4) firmaron el experimento y se les ofreció el tratamiento, pero optaron por no recibirlo, estimamos el efecto promedio local del tratamiento (LATE) en los cumplidores (es decir, aquellos que aceptan el tratamiento cuando se ofrece) utilizando mínimos cuadrados de dos etapas (2SLS). Consideramos que una pareja escuela-sección recibió el tratamiento si tuvo al menos una semana de sesiones de laboratorio de Conecta Ideas (es decir, apareció en los datos del panel de Conecta Ideas). Específicamente, instrumentamos si la escuela es tratada con un maniquí igual a uno si se ofrece el tratamiento. Nótese que hay tres variables endógenas (que indican cuál de los tratamientos recibió realmente la escuela) y tres instrumentos (la asignación aleatoria en los tres grupos).

Datos faltantes y desgaste

Un posible efecto del tratamiento es que más estudiantes terminen el año escolar y rindan el examen SIMCE. Esto daría como resultado que más estudiantes en los grupos de tratamiento (en relación con el control) tomaran la prueba SIMCE.¹² Esto sesgaría nuestras estimaciones, ya que la muestra de estudiantes de las escuelas de tratamiento y control ya no sería comparable. Para probar esto, estimamos si el número de estudiantes evaluados está equilibrado entre los grupos experimentales. Además, comprobamos si las características observables de los grupos están equilibradas. Como se muestra en la sección 4.7, no hay razones para sospechar que la composición de los estudiantes sea diferente entre los grupos (Tabla 11).

Heterogeneidad

Centraremos el análisis de heterogeneidad en las secciones que fueron tratadas en las escuelas tratadas (es decir, en los efectos directos). En general, la idea es observar los efectos de los subgrupos de estudiantes por separado para ver si algunos tienen efectos de tratamiento más altos o bajos. Por lo tanto, estimaremos la siguiente ecuación (utilizando datos de las secciones tratadas en las escuelas tratadas y de todas las secciones en las escuelas control):

$$Y_{is} = \beta_1 CIs + \alpha_1 CIs \times H_i + \beta_2 IIs + \alpha_2 IIs \times H_i + \beta_3 GIs + \alpha_3 GIs \times H_i + \gamma_s + \lambda X_i + \phi X_s + \varepsilon_{is}$$

donde todo está en la ecuación 2 y H_i es una característica del estudiante o de la escuela. En este contexto, α_1, α_2 y α_3 mide si hay algún efecto heterogéneo a lo largo de H_i . Haremos este análisis para el género, el nivel socioeconómico (SES) y el desempeño de referencia, es decir, H_i serán cada una de estas variables en cada uno de los análisis de heterogeneidad.¹³

Análisis adicional

No podemos controlar el rendimiento inicial de los estudiantes en nuestros análisis principales. En su lugar, controlamos los agregados a nivel escolar. Sin embargo, solicitamos a la Agencia de Calidad de la Educación que fusionara los datos de APTUS con los datos de SIMCE. Para proteger la identidad de los examinados, se estimó el decil de la capacidad de los estudiantes al inicio del estudio, el cual se fusionó con los datos del SIMCE.

La Agencia de Calidad de la Educación devolvió esta base de datos fusionada con los estudiantes que pudieron cotejar en los datos del SIMCE 2022 con base en la cédula de identidad nacional. Pudieron emparejar a 7.474 estudiantes. En concreto, de los 2.034 estudiantes que no pudieron emparejar, 1.868 estudiantes no fueron encontrados. El resto de los estudiantes no fueron emparejados debido a "inconsistencias en la clave de identificación, la clave de identificación de la escuela o por cuestiones de seguridad". Esta base de datos representa un subconjunto del total de estudiantes que componen nuestra muestra principal, pero con la característica adicional de que tenemos el decil de desempeño en la línea de base. Dado que el emparejamiento podría introducir un desgaste diferencial, comprobamos si la probabilidad de emparejamiento es diferente entre los distintos grupos. En general, no encontramos evidencia de deserción diferencial. Luego, probamos si esta tasa de coincidencia está correlacionada con el tratamiento (no lo está). Finalmente, estimamos el efecto de Conecta Ideas sobre esta submuestra, controlando el decil de habilidad basal del estudiante.

Tamaños de efecto

Se estandarizaron los puntajes del SIMCE para tener una media de 0 y una desviación estándar de 1 en el grupo control. Por lo tanto, los efectos que estimamos ya están estandarizados y pueden interpretarse como cambios en la media del grupo de tratamiento, medidos en términos de la desviación estándar en el grupo de control. Por ejemplo, un efecto de tratamiento de las puntuaciones de los exámenes se interpreta como que el estudiante medio se mueve 0,2 desviaciones estándar hacia la derecha en la distribución de las puntuaciones. 0.2σ

Los índices basados en las encuestas a estudiantes, maestros y padres también están estandarizados de manera similar. Al estudiar los componentes individuales de estos índices, no estandarizamos, sino que presentamos la media del grupo de control para que los tamaños del efecto puedan interpretarse como cambios en relación con la media de control.

¹² Los estudiantes sin puntajes SIMCE no se incluyen en las estimaciones principales.

¹³ Originalmente, teníamos la intención de hacer heterogeneidad según la motivación inicial de los estudiantes, pero no pudimos vincular estas mediciones de la línea de base con la prueba SIMCE 2022 debido a preocupaciones de privacidad.

Correlación intraclúster

Dado que el tratamiento fue aleatorizado primero a nivel escolar y luego en diferentes secciones de la misma escuela, nos agrupamos a nivel de sección escolar (Abadie, Athey, Imbens, & Wooldridge, 2023).

Al inicio del estudio, la correlación intraconglomerado para los estudiantes de la misma escuela es .14. Para los estudiantes dentro de la misma escuela y sección, es .145. Esto se estima usando el comando "loneway" en Stata 17.

Para la prueba SIMCE, la correlación intraconglomerado para los estudiantes dentro de la misma escuela es .164 en matemáticas y en lenguaje. Para los estudiantes dentro de la misma escuela y sección, es .17 en matemáticas y en lenguaje. Esto se estima usando el comando "loneway" en Stata 17.

Evaluación de Procesos

Objetivo

Esta evaluación del proceso tiene como objetivo determinar si el programa Conecta Ideas se alinea con su propósito y plan previsto e identificar los factores que facilitan o dificultan su ejecución. Acompaña a una evaluación de impacto aleatoria para evaluar los efectos del programa en el rendimiento en matemáticas de cuarto grado. Este doble enfoque mejora la comprensión de los mecanismos y condiciones de implementación del programa, proporcionando contexto para las interpretaciones de la evaluación de impacto y contribuyendo al aprendizaje continuo y la escalabilidad de Conecta Ideas.

La evaluación analiza las actividades y los resultados del programa utilizando Saunders, Evans, and Joshi (2005) siete criterios analíticos: reclutamiento, fidelidad, dosis administrada, satisfacción, alcance y contexto. Estos criterios ofrecen una visión integral de los procesos operativos y de seguimiento del programa, asegurando que la ejecución cumpla con los estándares mínimos para el logro de sus objetivos.

El criterio de contratación evalúa la idoneidad y la calidad de los procedimientos de contratación de participantes. Examina la efectividad de la estrategia de comunicación del programa, la capacidad del equipo de AutoMind para proporcionar retroalimentación, cómo los maestros reciben y entienden el programa, la claridad de la difusión de información a los maestros, el manejo de la deserción escolar, la transparencia de los criterios de selección y la claridad de la prueba diagnóstica APTUS para directores y maestros.

El criterio de fidelidad evalúa si la intervención se implementa según lo planeado, centrándose en las actividades, los insumos y los productos. Verifica si los establecimientos seleccionados tienen laboratorios de computación adecuados, las pruebas de funcionalidad de la plataforma, las ventajas y desventajas de tener un coordinador de laboratorio y las percepciones de los maestros sobre la plataforma.

El criterio de dosis administrada mide si los componentes de la intervención se entregan según lo planeado, analizando los objetivos relacionados con las escuelas y los estudiantes participantes, las razones de las discrepancias, la contratación del coordinador de laboratorio, los números de sesión, las cantidades de ejercicio y los minutos de participación de los estudiantes en la plataforma Conecta Ideas.

El criterio de dosis recibida evalúa la receptividad de los participantes a la intervención y su uso de los productos del programa. Se centra en el grado de motivación y atención de los participantes durante los ejercicios, su disposición a buscar ayuda y su compromiso con la plataforma.

El criterio de satisfacción determina la satisfacción de los participantes con los productos del programa. Explora la satisfacción general, las fortalezas y debilidades del programa, la satisfacción con la duración de la sesión, la plataforma, los ejercicios, el nivel de dificultad, la colaboración entre los coordinadores de laboratorio y los profesores, el papel de los monitores estudiantiles, las percepciones de los estudiantes sobre los incentivos y la interacción entre los participantes que reciben incentivos grupales.

El criterio de alcance evalúa la proporción de la audiencia a la que va dirigida que participa en la intervención. Examina la tasa de participación de las escuelas y los estudiantes, las medidas para prevenir la deserción escolar y las principales razones del abandono de los programas por parte de las escuelas.

El criterio de contexto refleja los aspectos ambientales que influyen en la implementación del programa. Busca comprender las características socioeconómicas de las escuelas y los participantes y los factores políticos, económicos, de salud y sociales que afectan el programa.

Metodología

Muestreo

Empleamos un enfoque de métodos mixtos para cumplir con los objetivos de la evaluación, combinando análisis cualitativo y cuantitativo. Esto implicó la triangulación de varias herramientas de recolección de datos. Profundizamos cualitativamente en los mecanismos de implementación del programa Conecta Ideas 2022. Cuantitativamente, medimos las percepciones de estudiantes, profesores y equipos directivos sobre el programa.

Diseñamos distintos métodos de muestreo para cada enfoque. Para la parte cualitativa, se seleccionó una muestra intencional de 15 escuelas, el ~10% de las que recibieron el tratamiento, a través de la aleatorización estratificada.¹⁴ Esta estratificación consideró tres factores: los puntajes de referencia de la escuela en matemáticas de APTUS (por encima o por debajo de la mediana), el número de secciones de cuarto grado y la variante del tratamiento recibido.

Las escuelas elegidas están distribuidas en 12 municipios de la Región Metropolitana, divididas en partes iguales entre los brazos de tratamiento de Conecta Ideas. De estas escuelas se recolectaron datos basados en tres metodologías: observación no participante, entrevistas semiestructuradas y encuestas autoadministradas. El objetivo principal de la recolección de datos fue comprender la implementación del programa en las aulas, su distinción de las clases regulares de matemáticas y las percepciones y satisfacción de las partes interesadas con el programa.

Observación no participante

Asistimos a clases de Conecta Ideas y clases regulares de matemáticas, observando sin alterar las conductas rutinarias (Kawulich, 2005). Para minimizar el sesgo del observador, nos familiarizamos con el equipo de implementación y los coordinadores de laboratorio, comunicamos los objetivos del estudio y aclaramos que la evaluación se centró en el programa, no en el desempeño individual. Antes de las observaciones, visitamos las 15 escuelas seleccionadas, explicamos el alcance de la evaluación a los profesores y a la dirección, y acordamos métodos de observación no intrusivos. Se realizaron 38 observaciones no participantes, centrándonos en las sesiones de Conecta Ideas. También observamos clases regulares de matemáticas para las estrategias de enseñanza y las similitudes/diferencias con las sesiones de la plataforma.

Entrevistas semiestructuradas

Se realizaron entrevistas para comprender las percepciones de los actores clave e identificar los procesos que no pudieron ser observados durante las observaciones no participantes. Guiados por un cuestionario estandarizado con respuestas abiertas, entrevistamos a docentes, directores, coordinadores de laboratorio y al equipo de implementación (en el Anexo 11 se reproducen los cuestionarios completos). Nuestras estrategias para garantizar la apertura incluyeron comunicar el propósito de las entrevistas, garantizar la confidencialidad y permitir la retirada. Elegimos lugares accesibles y privados para las entrevistas. Realizamos 44 entrevistas en profundidad que abarcaron el funcionamiento del programa, las necesidades de implementación escolar, las expectativas, la satisfacción, los procesos de toma de decisiones y el contexto.

Encuestas

Además, implementamos encuestas autoadministradas en línea para estudiantes, profesores de matemáticas y directores de escuelas (ver Tabla 10). La encuesta a los estudiantes abordó las percepciones de la plataforma Conecta Ideas, las clases de matemáticas y la motivación. La encuesta a maestros y directores se centró en las expectativas, la satisfacción, la percepción general del programa, las actividades de implementación y las relaciones entre estudiantes y matemáticas.

¹⁴ Tenga en cuenta que la evaluación de impacto incluye todas las escuelas de tratamiento y control.

Tabla 10: Características de la encuesta

Encuesta	Fechas	Largura	Tasa de respuesta	de	Plataforma
Estudiantes	20/10/2022 21/11/2022	a 5 minutos	85%		SurveyMonkey durante una sesión de Conecta Ideas
Profesorado	25/10/2022 20/11/2022	a 9 minutos	83%		SurveyMonkey enviado por correo electrónico
Directores	25/10/2022 22/11/2022	a 5 minutos	77%		SurveyMonkey enviado por correo electrónico

Conecta Ideas Data

Además de los instrumentos de investigación, la plataforma Conecta Ideas registró todas las actividades de los estudiantes durante las sesiones. Estos datos ayudaron a responder preguntas sobre las características de la sesión, especialmente para los criterios de dosis recibida y fidelidad.

Plan de análisis

El análisis cualitativo se basó en el análisis de contenido temático, utilizando los criterios de **Saunders et al. (2005)**. Se emplearon herramientas analíticas inductivas para incluir posibles omisiones de categorías de análisis preestablecidas. Esto implicó un proceso de codificación en tres etapas: codificación abierta, axial y selectiva, incorporada inicialmente por la Teoría Fundamentada y utilizada en nuestro análisis temático.

Los datos cualitativos se procesaron a partir de observaciones y entrevistas utilizando el software ATLAS.ti. Esta herramienta analiza eficazmente el texto a través de la clasificación y la descomposición, lo que nos permite compilar notas de campo, transcripciones de entrevistas y documentación relevante en un solo espacio de trabajo.

La codificación abierta se refiere al enfoque inductivo inicial de los datos, creando categorías analíticas espontáneas y descubriendo ideas, conceptos o pensamientos a través de la generación de citas y la categorización en códigos, que pueden no necesariamente alinearse con el marco teórico del analista. La codificación axial implica analizar las relaciones entre categorías, dimensiones y subdimensiones, generar diagramas para ilustrar estas relaciones y distinguir problemas y niveles de relación. El análisis selectivo se centra en las ideas centrales, guiando nuestros análisis, discusiones y conclusiones.

También se realizó un análisis cuantitativo descriptivo de las encuestas a estudiantes, docentes y directores. Sin embargo, estos datos son representativos solo de la muestra encuestada. Cabe destacar que los resultados del análisis descriptivo de la encuesta complementan el análisis cualitativo, informando particularmente el criterio de satisfacción definido por **Saunders et al. (2005)**.

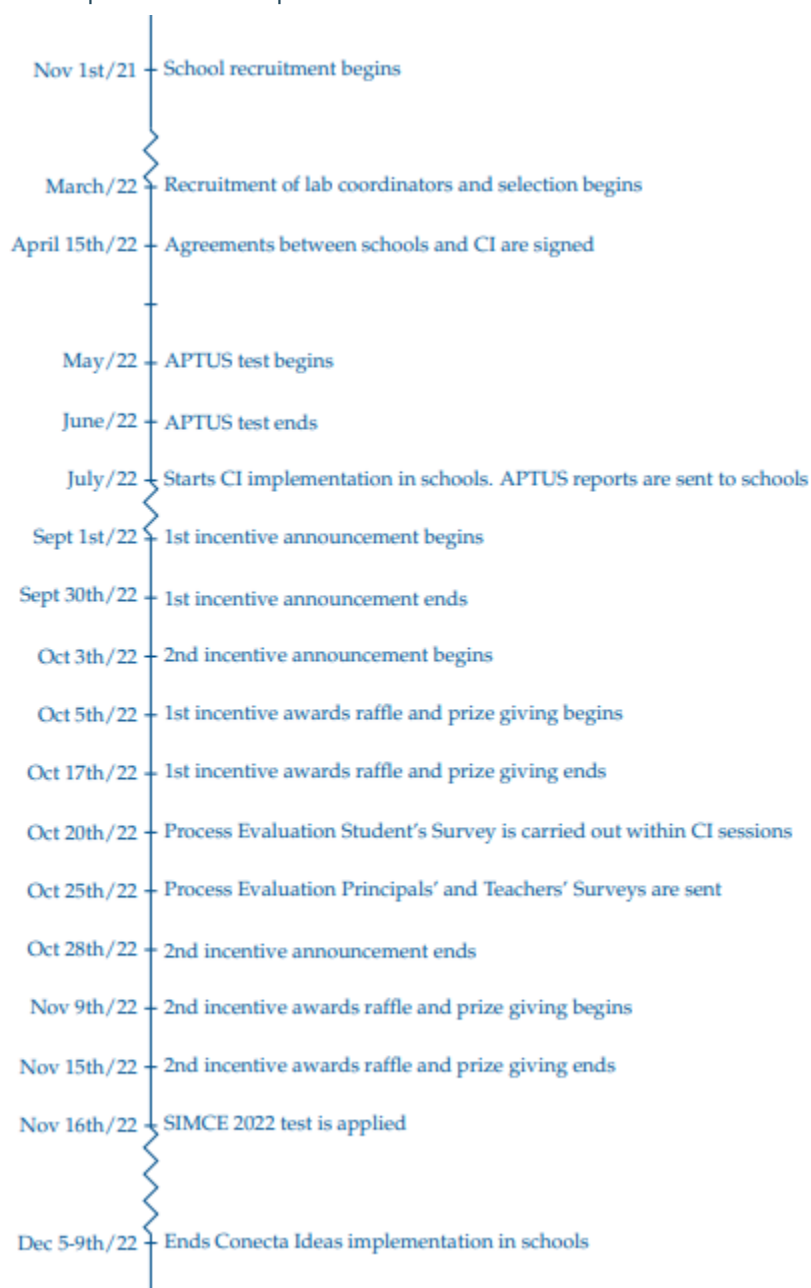
Sobre la base de los resultados de la evaluación del proceso, realizamos un análisis adicional. En concreto, se realizó un análisis dosis-respuesta y algunos análisis de heterogeneidad. Las conclusiones incorporan los hallazgos tanto de la evaluación de impacto como de la evaluación del proceso.

Datos de costos

Los datos de costos se solicitaron a AutoMind y son autoinformados.

Línea de tiempo

Figura 4: Cronología de eventos importantes en la implementación de Conecta Ideas

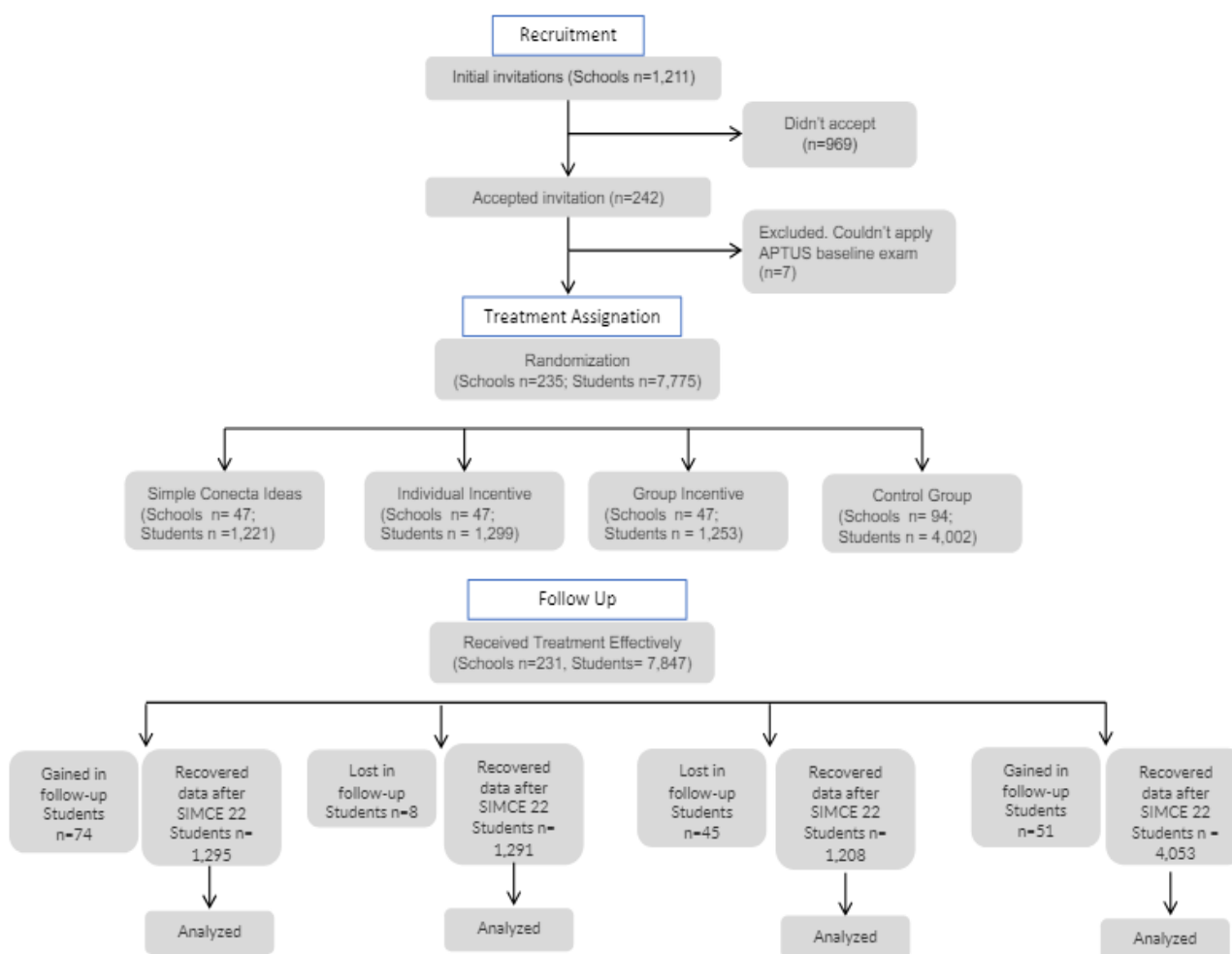


Resultados de la evaluación de impacto

Participantes

Las secciones 3.1.2 y 3.1.6 tienen una visión general de las características de los participantes en nuestro experimento. En la figura 5 se resumen los tamaños de muestra al inicio y al final. La sección 4.2 tiene más detalles sobre el número de estudiantes en el conjunto de datos del SIMCE 2022. En el Cuadro 11 se resumen los cálculos de potencia de la sección 3.1.4.

Figura 5: Diagrama de flujo de conteo de estudiantes y escuelas



Nota: El número de estudiantes en la asignación del tratamiento se estima a partir de los datos de la prueba que los estudiantes tomaron al inicio del estudio, que fue organizada por APTUS. Estos corresponden a la matrícula en ese momento, no al número de estudiantes que rindieron la prueba. Los datos del seguimiento provienen del SIMCE 2022. Esto incluye a todos los estudiantes, independientemente de si pudimos hacer coincidir los datos de APTUS con los datos del SIMCE 2022 (para ver un diagrama de flujo con los estudiantes emparejados, ver Figura B.4). El número de estudiantes "perdidos (ganados) en el seguimiento" es la diferencia entre los matriculados al inicio (según los datos de APTUS) y los que rindieron la prueba SIMCE 2022. Los niños de las secciones no tratadas en las escuelas tratadas están excluidos de este diagrama de flujo.

Tabla 11: Parámetros para el cálculo de poder estadístico

	Ex-ante	Ex-Post
Efecto mínimo detectable	0.21σ	0.149σ
Panel A: Correlación pre/post		
Nivel 1 (estudiante)	30%	55%
Nivel 2 (aula)	NA	63%
Nivel 3 (escuela)	NA	59%
Panel B: Correlación intracúmulo (ICC)		
Nivel 2 (aula)	.145	.17
Nivel 3 (escuela)	.14	.164
Panel C: Tamaño y potencia		
Tamaño	0.05	0.05
Poder	0.9	0.9
Panel D: Prueba de un lado o de dos lados		
	Bilateral	Bilateral
Panel E: Características del tamaño de la muestra		
Tamaño medio del clúster	28	33
Número de escuelas de Simple Conecta Ideas	50	47
Número de escuelas de incentivos individuales	50	47
Número de escuelas de incentivos grupales	50	47
Número de escuelas de control	85	94
Número total de escuelas	235	235

Número de alumnos de Simple Conecta Ideas	1,221	1,295
Número de estudiantes de Incentivo Individual	1,299	1,291
Número de estudiantes de Incentivo Grupal	1,253	1,208
Número de alumnos de Control	4,002	4,053
Número total de estudiantes	7,775	7,847

Notas: La correlación entre los puntajes de referencia de APTUS y el puntaje SIMCE (la correlación pre/post ex-post) se estimó ejecutando una regresión del puntaje de SIMCE 2022 en los deciles de desempeño en la prueba APTUS. Este análisis solo utiliza a los estudiantes que fueron emparejados en ambos conjuntos de datos. No se tenía un previo para la correlación entre los resultados previos y posteriores en el aula o en la escuela. Para las correlaciones entre el resultado de línea de base y el resultado final a nivel de aula y escuela, estimamos la correlación entre el SIMCE 2018 y el SIMCE 2022.

Desgaste

Resultado principal

La principal base de datos que utilizamos para estudiar el impacto del programa son los datos del SIMCE 2022 proporcionados por el Ministerio de Educación de Chile. Estos datos contienen información sobre 12.703 estudiantes de las 235 escuelas del experimento. Sin embargo, nuestra muestra final es más pequeña.

Hay 2,228 estudiantes a los que les faltan puntajes en matemáticas y lenguaje. Una vez eliminadas estas observaciones, tenemos una muestra de 10.475 estudiantes. Para recuperar la asignación de tratamiento de cada alumno, necesitamos identificar la sección de 4° de primaria a la que pertenecen.

De estos 10.475 estudiantes, 69 estudiantes no tienen datos sobre la sección a la que pertenecen. De ellos, 28 de estos alumnos pertenecen a escuelas del grupo control. La sección de la que provienen es irrelevante, por lo que asignamos uno al azar. En el caso de los otros 41 estudiantes, los eliminamos de los datos. Así, al final, tuvimos 10.434 alumnos.

En total, hay 4.053 alumnos en las escuelas de control. En las secciones tratadas en las escuelas tratadas con la versión simple de Ideas Conecta, hay 1.295 estudiantes. En las secciones tratadas de las escuelas individuales de incentivos, hay 1.291 estudiantes; en las secciones tratadas de las escuelas de Incentivo Grupal, hay 1.208 estudiantes. Los 2.587 estudiantes restantes se encuentran en secciones no tratadas en escuelas tratadas. Hay 429 estudiantes que tienen puntajes en exámenes de matemáticas, pero no un idioma.

En la Tabla 12 se compara el número de estudiantes evaluados al inicio (con la prueba APTUS) y al final (la prueba estandarizada SIMCE).

Tabla 12: Deserción de estudiantes en el análisis primario

Grupo	Número de estudiantes aleatorizados	Número de estudiantes analizados	Deserción (Número)	Deserción (Porcentaje)	Ganancia (Número)	Ganancia (Porcentaje)
Control	4,002	4,053	-	-	51	1.27%
Conecta Ideas Simple	1,221	1,295	-	-	74	6.06%
Incentivos Individuales	1,299	1,291	8	0.61%	-	-
Incentivos Grupales	1,253	1,208	45	3.60%	-	-
Total	7,775	7,847	53	-	125	-

Notas: El número de estudiantes de la fila "Aleatorizada" se estima a partir de los datos de la prueba que los estudiantes tomaron al inicio del estudio, que fue organizada por APTUS. Estos corresponden a la matrícula en ese momento, no al número de estudiantes que rindieron la prueba. Los datos de la fila "Analizados" provienen del SIMCE 2022. Esto incluye a todos los estudiantes, independientemente de si pudimos hacer coincidir los datos de APTUS con los datos de SIMCE 2022. Los niños de las secciones no tratadas en las escuelas tratadas están excluidos de este diagrama de flujo. Consulte el apéndice para los estudiantes emparejados con la versión B.4 de los datos de APTUS.

Resultados secundarios

En total, 10.785 estudiantes respondieron a la encuesta estudiantil. Sin embargo, algunos estudiantes se saltaron algunas preguntas. De estos 10,785 estudiantes, 4,184 pertenecen al grupo de control, 1,348 estudiantes a secciones tratadas en escuelas de Conecta Ideas (versión simple), 1,343 a secciones tratadas en escuelas de incentivos individuales y 1,259 a secciones tratadas en escuelas de incentivos grupales. Los 2.651 restantes pertenecen a secciones no tratadas en escuelas tratadas.

Además, 144 profesores de las escuelas participantes en el experimento respondieron a la encuesta. De estos 144 maestros, 58 pertenecen al grupo de control, 15 maestros para tratar secciones en escuelas tratadas sin incentivos, 12 para tratar secciones en escuelas de incentivos individuales y 13 maestros para tratar secciones en escuelas de incentivos grupales. Los 46 restantes pertenecen a secciones no tratadas en escuelas tratadas.

Finalmente, 9.559 padres respondieron a la encuesta: 3.752 en el grupo de control, 1.209 de escuelas tratadas sin los incentivos, 1.115 de las escuelas de incentivos individuales y 1.090 de escuelas de incentivos grupales.

Características del estudiante y de la escuela

Véase la Sección 3.1.6 para una descripción detallada de los alumnos y las escuelas del experimento. Véase la Sección 3.1.2 para una descripción detallada de cómo se comparan las escuelas del experimento con otras escuelas del país.

Análisis primario

El análisis principal consiste en evaluar el efecto de cada tratamiento en los puntajes de Matemática y Lenguaje de la prueba estandarizada SIMCE 2022 que los estudiantes de 4° grado realizan al final del año escolar. Para estimar el efecto de las intervenciones, comparamos a los estudiantes matriculados en las secciones tratadas en las escuelas tratadas con los estudiantes en las escuelas de control (ver Ecuación 2). En general, las estimaciones apuntan a un efecto negativo de Conecta Ideas en el rendimiento de los estudiantes.

Específicamente, no hay diferencias estadísticamente significativas en el desempeño de los estudiantes en las puntuaciones de matemáticas en ninguno de los grupos de tratamiento (en relación con el control). Las estimaciones puntuales son negativas para los tres tratamientos, y podemos descartar que los efectos de cualquiera de las versiones del tratamiento estén por encima de $0,045 \sigma$ al nivel del 95%. Las diferencias en los efectos del tratamiento entre los grupos son insignificantes. En cuanto al lenguaje, también encontramos que los incentivos individuales tienen un efecto negativo ($-0,090 \sigma$) y estadísticamente significativo (p-valor 0,014).

Tabla 13: Efectos del tratamiento en puntajes de matemáticas y lenguaje SIMCE 2022

		Promedio Puntaje Matemáticas (1)	Promedio Puntaje Matemáticas (2)	Promedio Puntaje Lenguaje (4)	Promedio Puntaje Lenguaje (5)
Conecta Simple	Ideas	-0.057	-0.058	-0.011	0.006
Incentivos Individuales		-0.063	-0.035	-0.120**	-0.090**
Incentivos Grupales		-0.061	-0.074	-0.037	-0.044
N		7,848	7,848	7,533	7,533
Promedio grupo ref.		-0.05	-0.05	-0.07	-0.07
Dif. Simple vs Indiv.		0.005 (0.91)	0.024 (0.6)	0.109 (0.06)	0.096 (0.07)
Dif. Simple vs Grupo		0.004 (0.94)	0.015 (0.76)	0.027 (0.65)	0.051 (0.32)
Dif. Indiv. vs Grupo		0.001 (0.98)	0.039 (0.43)	0.082 (0.13)	0.045 (0.32)
Cualquier Tratamiento		-0.06 (0.13)	-0.055 (0.09)	-0.056 (0.1)	-0.042 (0.13)
Controles	No	Sí	Sí	No	Sí
Nivel Reg.	Estudiante	Estudiante	Estudiante	Estudiante	Estudiante
FE: estrato	X	X	X	X	X
Errores Agrupados	por: rbd	por: rbd	por: rbd	por: rbd	por: rbd

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Las columnas (3) y (4) tienen menos observaciones que las columnas (1) y (2) ya que no todos los estudiantes que tomaron la prueba de Matemáticas tomaron la prueba de Lenguaje. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Análisis secundario

A continuación, estudiamos si el comportamiento del estudiante, profesor o padre de familia cambia en respuesta al tratamiento utilizando datos de las encuestas que respondieron como parte del SIMCE.

Cuestionario a nivel de estudiante

No todos los estudiantes que rindieron la prueba estandarizada SIMCE 2022 respondieron la encuesta. Sin embargo, no hay deserción diferencial entre los grupos experimentales (ver Sección 4.7).

El tratamiento no afecta al índice de autopercepción (Tabla 14) ni a ninguno de sus componentes (Tabla B.5). En todo caso, parece que los tratamientos tuvieron un efecto negativo en la percepción de los estudiantes sobre cuánto los valora su profesor (Tabla 15). Estos efectos negativos parecen deberse a que los estudiantes de todos los grupos de tratamiento tienen menos probabilidades de sentir que su profesor considera que son capaces de aprender y que son importantes (Tabla B.6). Las diferencias en los efectos del tratamiento entre los grupos son insignificantes.

No encontramos ninguna diferencia sistemática en la motivación de los estudiantes para asistir y aprender en la escuela entre los grupos (véanse las Tablas 16 y B.7). Sorprendentemente, tampoco hay ningún efecto en la familiaridad de los estudiantes con las tecnologías de la información y la comunicación (TIC) (véase la Tabla 17). Específicamente, no hay ningún efecto sobre la probabilidad de que sepan cómo enviar correos electrónicos, hacer presentaciones de PowerPoint o que usen Internet para buscar información (Tabla B.8). Las diferencias en los efectos del tratamiento entre los grupos son insignificantes.

Tabla 14: Estimaciones de impacto en puntajes de matemáticas y lenguaje SIMCE 2022

	IRT Autopercepción (1)	IRT Autopercepción (2)	MCA Autopercepción (1)	MCA Autopercepción (2)
CI simple	0.010	0.002	0.052	0.048
Desviación estándar	(0.041)	(0.042)	(0.035)	(0.033)
Incentivo individual	-0.030	-0.028	0.048	0.035
Desviación estándar	(0.043)	(0.043)	(0.034)	(0.033)
Incentivo grupal	-0.046	-0.054	0.010	0.016
Desviación estándar	(0.044)	(0.044)	(0.029)	(0.028)
N	[8,098]	[8,081]	[8,098]	[8,081]
Promedio grupo de referencia	-0.02	-0.02	0	0
Diferencia Simple vs Indiv.	0.04 (0.44)	0.03 (0.56)	0 (0.94)	0.01 (0.77)
Diferencia Simple vs Grupal	0.06 (0.27)	0.06 (0.28)	0.04 (0.3)	0.03 (0.42)

Diferencia vs Grupal	Indiv.	0.02 (0.76)	0.03 (0.63)	0.04 (0.32)	0.02 (0.61)
Cualquier tratamiento		-0.021 (0.48)	-0.026 (0.4)	0.037 (0.1)	0.037 (0.1)
Controles	No		Sí	No	Sí
Efectos fijos por estrato	X		X	X	X
Errores agrupados por	rbd		rbd	rbd	rbd
Nivel de regresión	Estudiante		Estudiante	Estudiante	Estudiante

Notas: Las diferencias en el número de observaciones se deben a que los estudiantes no responden a todas las preguntas que componen el índice. La tabla presenta dos versiones del índice: la versión IRT, que se construye con herramientas de la Teoría de Respuesta al Ítem, y la versión MCA, que se construye con las herramientas de Análisis Multicomponente. El índice de autopercepción de los estudiantes se compone de seis componentes (véase la Tabla B.5 del Apéndice para las regresiones por componentes del índice): El primer componente es "Buena tarea", que se relaciona con la percepción del estudiante sobre la calidad de sus tareas. "Aprende fácil" se relaciona con la percepción que tiene el estudiante de sus capacidades de aprendizaje. "Buenas calificaciones" se relaciona con la creencia del estudiante en su capacidad para obtener buenas calificaciones. "Buen estudiante" se relaciona con el estudiante que sabe que es un buen alumno. "Buenas calificaciones de fin de año" se refiere a la creencia del estudiante de terminar el año escolar con una evaluación positiva. "Aprende más" se refiere a la confianza de los estudiantes en aprender más temas matemáticos de los que ven en el aula. Todos los componentes siguen la misma dirección, es decir, un aumento en el valor del componente significa un aumento en el rasgo deseado. Por lo tanto, contribuye en la misma dirección para el índice. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** p

Tabla 15: Efectos del tratamiento en las creencias de los estudiantes sobre la percepción de los profesores sobre ellos

	MCA Percepción Profesor (1)	MCA Percepción Profesor (2)
Simple Conecta Ideas	-0.023 (0.031)	-0.034 (0.030)
Incentivos Individuales	-0.067** (0.030)	-0.073** (0.031)
Incentivos Grupales	-0.065* (0.034)	-0.072** (0.032)
N	8,074	8,057
Promedio grupo ref.	-0.02	-0.02
Dif. Simple vs Indiv.	0.04 (0.2)	0.04 (0.26)
Dif. Simple vs Grupo	0.04 (0.27)	0.04 (0.3)
Dif. Indiv. vs Grupo	0 (0.95)	0 (0.98)
Cualquier Tratamiento	0.051 (0.03)	0.059 (0.01)
Controles	No	Sí
Nivel Reg.	Estudiante	Estudiante
FE: estrato	X	X
Errores Agrupados por: rbd	X	X

Notas: Las diferencias en el número de observaciones se deben a que los estudiantes no responden a todas las preguntas que componen el índice. La tabla presenta la versión MCA, que se construye con las herramientas de la Teoría de Respuesta al Ítem. El índice de percepción del profesor sobre los estudiantes se relaciona con la percepción que los estudiantes creen que su maestro tiene de ellos. Se compone de tres componentes (véase el Cuadro B.6 del Apéndice para las regresiones por componentes del índice): "Buen estudiante" se refiere a la sensación del estudiante de que es un buen estudiante para el profesor. "Capaz de aprender" se relaciona con la percepción que tiene el profesor de las capacidades de aprendizaje del alumno. Y "soy importante" se refiere a la capacidad de los maestros de hacer que los estudiantes se sientan importantes para su escuela. Todos los componentes siguen la misma dirección, es decir, un aumento en el valor del componente significa un aumento en el rasgo deseado. Por lo tanto, contribuye en la misma dirección para el índice. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE

2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tabla 16: Efectos del tratamiento en la motivación

		IRT Motivación (1)	IRT Motivación (2)	MCA (3)	Motivación	MCA (4)	Motivación
Simple Ideas	Conecta	-0.037	-0.050	0.021		0.016	
		(0.041)	(0.041)	(0.032)		(0.031)	
Incentivos Individuales		-0.039	-0.040	0.033		0.023	
		(0.043)	(0.042)	(0.027)		(0.027)	
Incentivos Grupales		-0.068	-0.075*	-0.022		-0.027	
		(0.045)	(0.044)	(0.029)		(0.029)	
N		8,094	8,077	8,094		8,077	
Promedio grupo ref.		0.02	0.02	0.02		0.02	
Dif. Simple vs Indiv.		0 (0.96)	0.01 (0.83)	0.01 (0.74)		0.01 (0.85)	
Dif. Simple vs Grupo		0.03 (0.53)	0.02 (0.62)	0.04 (0.26)		0.04 (0.26)	
Dif. Indiv. vs Grupo		0.03 (0.56)	0.04 (0.5)	0.05 (0.11)		0.05 (0.15)	
Cualquier Tratamiento		-0.048 (0.14)	-0.055 (0.08)	0.011 (0.59)		0.005 (0.82)	
Controles		No	Sí	No		Sí	
Nivel Reg.		Estudiante	Estudiante	Estudiante		Estudiante	
FE: estrato		X	X	X		X	
Errores por: rbd	Agrupados	Sí	Sí	Sí		Sí	

Notas: Las diferencias en el número de observaciones se deben a que los estudiantes no responden a todas las preguntas que componen el índice. La tabla presenta dos versiones del índice: la versión IRT, que se construye con herramientas de la Teoría de Respuesta al Ítem, y la versión MCA, que se construye con las herramientas de Análisis Multicomponente. El índice de motivación consta de siete componentes (véase el cuadro B.7 del apéndice para las regresiones por componentes del índice): "Le gusta estudiar" se refiere a la satisfacción de los estudiantes cuando estudian. "Pone esfuerzo" se refiere al esfuerzo que un estudiante invierte en ser un buen estudiante. "Le gusta la escuela" se relaciona con cuánto le gusta al estudiante la

escuela a la que asiste. "Le gusta enseñar" tiene que ver con la afinidad que el estudiante siente hacia los modales de enseñanza de la escuela. "Asiste. Importa" se relaciona con la creencia del estudiante sobre la importancia de asistir a clases cuando es niño. "Esfuerzo en clase" se refiere al esfuerzo que un estudiante cree que está poniendo en las clases del maestro. "Preparación para exámenes" se relaciona con la frecuencia con la que el estudiante se prepara para sus exámenes. Todos los componentes siguen la misma dirección, es decir, un aumento en el valor del componente significa un aumento en el rasgo deseado. Por lo tanto, contribuye en la misma dirección para el índice. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores p . $p < 0.10$ $p < 0.05$ $p < 0.01$

Tabla 17: Efectos del tratamiento en el índice de habilidad digital de los estudiantes

		IRT Digital	Habilidad	IRT Digital	Habilidad	MCA Digital	Habilidad	MCA Digital	Habilidad
		(1)		(2)		(3)		(4)	
Simple Ideas	Conecta	0.018 (0.039)		0.013 (0.039)		0.029 (0.028)		0.030 (0.028)	
Incentivo Individual		-0.029 (0.040)		-0.031 (0.040)		0.017 (0.031)		0.006 (0.031)	
Incentivo de Grupo		-0.013 (0.040)		-0.024 (0.039)		0.008 (0.028)		0.014 (0.027)	
N		7,351		7,335		7,351		7,335	
Media grupo de referencia		0		0		0		0	
Dif. Simple vs. Indiv.		0.05 (0.34)		0.04 (0.37)		0.01 (0.74)		0.02 (0.5)	
Dif. Simple vs. Grupo		0.03 (0.52)		0.04 (0.44)		0.02 (0.53)		0.02 (0.64)	
Dif. Indiv. vs. Grupo		0.02 (0.76)		0.01 (0.88)		0.01 (0.81)		0.01 (0.81)	
Cualquier Tratamiento		-0.008 (0.78)		-0.014 (0.62)		0.018 (0.39)		0.017 (0.4)	
Controles		No		Sí		No		Sí	
Nivel de Regresión		Estudiante		Estudiante		Estudiante		Estudiante	
FE: estratos		X		X		X		X	

Errores agrupados rbd rbd rbd rbd
por:

Notas: Las diferencias en el número de observaciones se deben a que los estudiantes no responden a todas las preguntas que componen el índice. La tabla presenta dos versiones del índice: la versión IRT, que se construye con herramientas de la Teoría de Respuesta al Ítem, y la versión MCA, que se construye con las herramientas de Análisis Multicomponente. El índice de Habilidades Digitales se compone de 6 habilidades diferentes relacionadas con el uso de programas informáticos básicos (véase la Tabla B.8 del Apéndice para las regresiones por componentes del índice): "Texto" se refiere a la capacidad del estudiante para escribir documentos. "Enviar correos electrónicos" se relaciona con la capacidad del niño para escribir y enviar un correo electrónico. "PPTX Presen." se refiere a la capacidad de crear una presentación con diapositivas y archivos adjuntos multimedia. "Información de búsqueda" se refiere a la capacidad del estudiante para buscar la información necesaria para realizar su tarea. "Información. Discusión." se relaciona con la capacidad de discernir información útil e inútil de la web para la tarea. "Veracidad de la información" se refiere a la capacidad del estudiante para contrastar información para evaluar su veracidad. Todos los componentes siguen la misma dirección, es decir, un aumento en el valor del componente significa un aumento en el rasgo deseado. Por lo tanto, contribuye en la misma dirección para el índice. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Cuestionario a nivel docente

Al igual que con los estudiantes, no todos los maestros de 4° grado respondieron la encuesta. Sin embargo, no hay deserción diferencial entre los grupos experimentales (ver Sección 4.7).

El tratamiento parece mejorar la percepción de los maestros sobre el comportamiento de los estudiantes (ver Tablas 18). Las diferencias en los efectos del tratamiento entre los grupos son insignificantes. Si bien los profesores afirman que los estudiantes tienen la misma probabilidad de respetar las normas del aula y prestar atención en los cuatro grupos experimentales, afirman que es menos probable que los estudiantes las interrumpen en clase (véase la Tabla B.9).

Por otro lado, tampoco hay diferencias en la probabilidad de que los docentes utilicen las Tecnologías de la Información y la Comunicación (TIC) en el aula (véanse las Tablas 19 y B.10), ni en la probabilidad de que los docentes completen los objetivos de enseñanza establecidos por el Ministerio de Educación (véase la Tabla 20). Las diferencias en los efectos del tratamiento entre los grupos son insignificantes.

Tabla 18: Efectos del tratamiento en el índice de comportamiento en clase

		Índice de comportamiento en clase (MCA)	Índice de comportamiento en clase (MCA)
		(1)	(2)
Simple Ideas	Conecta	-0.190	-0.015
		(0.320)	(0.221)
Incentivo Individual		0.391**	0.384**
		(0.136)	(0.151)
Incentivo de Grupo		0.375**	0.325**
		(0.138)	(0.147)
N		98	98
Promedio ref.	grupo	0.2	0.2
Diferencia vs Indiv.	Simple	0.58 (0.05)	0.4 (0.08)
Diferencia vs Grupo	Simple	0.57 (0.06)	0.34 (0.11)
Diferencia Indiv. vs Grupo		0.02 (0.88)	0.06 (0.73)
Cualquier Tratamiento		-0.161 (0.38)	-0.229 (0.09)
Controles		No	Sí
Nivel de Reg.		Profesor	Profesor
FE: estrato		X	X
Errores agrupados por:		rbd	rbd

Notas: Las diferencias en el número de observaciones se deben a que los profesores no responden a todas las preguntas que componen el índice. La tabla presenta el índice IRT, que se construye con las herramientas de la Teoría de Respuesta al Ítem. El índice de comportamiento de clase captura el constructo de la percepción del maestro sobre la disciplina de los estudiantes. Se compone de tres componentes (véase la Tabla B.9 del Apéndice para las regresiones por componentes del índice): "Normas de respeto" tiene que ver con el respeto que los estudiantes dan a las normas impuestas por el profesor dentro del aula. "Presta atención" se relaciona con la atención de los estudiantes a la conferencia del maestro. Y "No interrumpir la clase" se refiere a la frecuencia con la que los estudiantes interrumpen la clase de manera indisciplinara. Todos los componentes

siguen la misma dirección, es decir, un aumento en el valor del componente significa un aumento en el rasgo deseado. Por lo tanto, contribuye en la misma dirección para el índice. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tabla 19: Efectos del tratamiento sobre la habilidad digital del profesor

		IRT Digital	Habilidad	IRT Digital	Habilidad	MCA Digital	Habilidad	MCA Digital	Habilidad
		(1)		(2)		(3)		(4)	
Conecta Simple	Ideas	0.001 (0.347)		-0.124 (0.340)		0.014 (0.257)		-0.064 (0.246)	
Incentivo Individual		-0.246 (0.310)		-0.224 (0.296)		-0.170 (0.230)		-0.130 (0.222)	
Incentivo Grupal		-0.232 (0.366)		-0.492 (0.383)		-0.154 (0.264)		-0.348 (0.273)	
N		98		98		98		98	
Media grupo de ref.		0.04		0.04		0.03		0.03	
Diferencia Simple vs Individual		0.25 (0.53)		0.1 (0.78)		0.18 (0.52)		0.07 (0.8)	
Diferencia Simple vs Grupal		0.23 (0.59)		0.37 (0.4)		0.17 (0.59)		0.28 (0.36)	
Diferencia Individual vs Grupal		0.01 (0.97)		0.27 (0.51)		0.02 (0.96)		0.22 (0.46)	
Cualquier Tratamiento		-0.146 (0.55)		-0.27 (0.28)		-0.094 (0.61)		-0.173 (0.34)	
Controles	No			Sí		No		Sí	
FE: estratos	X			X		X		X	
Errores agrupados por: rbd	X			X		X		X	

Nivel de Regresión: Profesor	Profesor	Profesor	Profesor	Profesor
---------------------------------	----------	----------	----------	----------

Notas: Las diferencias en el número de observaciones se deben a que los profesores no responden a todas las preguntas que componen el índice. La tabla presenta dos versiones del índice: la versión IRT, que se construye con las herramientas de la Teoría de Respuesta al Ítem, y la versión MCA, que se construye con las herramientas de Análisis Multicomponente. El índice de Habilidad Digital del Profesor captura el constructo del conocimiento del profesor en el uso de herramientas digitales y software computacional básico. Se compone de cinco componentes (consulte el Apéndice Tabla B.10 para ver las regresiones por componentes del índice): "Cargar documentos" se refiere a la capacidad del profesor para cargar cualquier documento en cualquier plataforma. "PPTX Presen." se refiere a la capacidad de crear una presentación con diapositivas y archivos adjuntos multimedia. "Ajustes de sintonía" se refiere a la capacidad del profesor para configurar sus ordenadores para mejorar su función. "Instalación de aplicaciones" se refiere al conocimiento del maestro sobre la instalación de cualquier aplicación en sus dispositivos portátiles e informáticos. "Excavación de clase. Plataformas" tiene que ver con la capacidad de utilizar plataformas digitales para apoyar las conferencias que imparten a los estudiantes. Todos los componentes siguen la misma dirección, es decir, un aumento en el valor del componente significa un aumento en el rasgo deseado. Por lo tanto, contribuye en la misma dirección para el índice. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tabla 20: Efectos del tratamiento en el índice de enseñanza de matemáticas

	IRT Enseñanza Matemáticas	IRT Enseñanza Matemáticas	MCA Enseñanza Matemáticas	MCA Enseñanza Matemáticas	% Enseñanza Matemáticas T.O.	% Enseñanza Matemáticas T.O.
	(1)	(2)	(3)	(4)	(5)	(6)
Conecta Ideas Simple	-0.193 (0.201)	-0.148 (0.192)	-0.031 (0.034)	0.004 (0.036)	-0.044 (0.038)	-0.048 (0.041)
Incentivo Individual	-0.154 (0.256)	-0.247 (0.229)	0.014 (0.057)	0.025 (0.056)	-0.069 (0.060)	-0.084 (0.056)
Incentivo Grupal	0.084 (0.326)	0.073 (0.337)	0.049 (0.073)	0.071 (0.074)	-0.065 (0.062)	-0.081 (0.065)
N	98	98	98	98	90	90
Media grupo de ref.	-0.05	-0.05	-0.05	-0.05	0.79	0.79
Diferencia Simple vs Individual	0.04 (0.89)	0.1 (0.71)	0.04 (0.43)	0.02 (0.72)	0.03 (0.67)	0.04 (0.58)
Diferencia Simple vs Grupal	0.28 (0.42)	0.22 (0.53)	0.08 (0.3)	0.07 (0.37)	0.02 (0.75)	0.03 (0.63)
Diferencia Individual vs Grupal	0.24 (0.52)	0.32 (0.39)	0.03 (0.69)	0.05 (0.6)	0 (0.95)	0 (0.97)
Cualquier Tratamiento	-0.1 (0.58)	-0.117 (0.49)	0.007 (0.85)	0.031 (0.41)	-0.057 (0.11)	-0.069 (0.05)
Controles FE: estratos	No X	Sí X	No X	Sí X	No X	Sí X
Errores agrupados por: rbd	X	X	X	X	X	X
Nivel de Regresión: Profesor	Profesor	Profesor	Profesor	Profesor	Profesor	Profesor

Notas: Las diferencias en el número de observaciones se deben a que los profesores no responden a todas las preguntas que componen el índice. Existen 27 Objetivos Didácticos (T.O.) en el plan de estudios oficial de 4° grado establecido por el Ministerio de Educación que deben ser cubiertos durante el año escolar. Los profesores respondieron al cuestionario sobre si sus clases habían completado cada una de las 27 T.O. Los índices se construyeron con las 27 respuestas dadas por los docentes. La tabla presenta dos versiones del

índice: la versión IRT, que se construye con herramientas de la Teoría de Respuesta al Ítem, y la versión MCA, que se construye con las herramientas de Análisis de Correspondencias Múltiples. Las columnas (1)-(4) son índices basados en IRT y MCA. Por otro lado, las columnas (5) y (6) son los porcentajes de Objetivos de Enseñanza de Matemáticas (T.O.) completados. Todos los componentes siguen la misma dirección, es decir, un aumento en el valor del componente significa un aumento en el rasgo deseado. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Cuestionario a nivel de padres

Al igual que con los estudiantes y docentes, no todos los padres de todos los estudiantes que rindieron la prueba SIMCE 2022 respondieron la encuesta. Sin embargo, no hay deserción diferencial entre los grupos experimentales (ver Sección 4.7).

En general, no encontramos ninguna evidencia de que los padres apoyen más a sus hijos en las tareas relacionadas con la escuela en respuesta al tratamiento (Tabla 21). En particular, no es más probable que ayuden con las actividades relacionadas con la escuela o la formación de un hábito de estudio (Cuadro B.11). Las diferencias en los efectos del tratamiento entre los grupos son insignificantes.

Tabla 21: Efectos del tratamiento en el índice de tutoría de los padres

		IRT Ayuda al Niño	IRT Ayuda al Niño	MCA Ayuda al Niño	MCA Ayuda al Niño
		(1)	(2)	(3)	(4)
Conecta Simple	Ideas	-0.057 (0.037)	-0.056 (0.037)	0.062* (0.033)	0.053* (0.032)
Incentivo Individual		-0.003 (0.039)	0.000 (0.039)	0.031 (0.030)	0.025 (0.029)
Incentivo Grupal		0.014 (0.039)	0.014 (0.039)	0.007 (0.030)	0.013 (0.030)
N		7,097	7,075	7,097	7,075
Media grupo de ref.		-0.04	-0.04	0.02	0.02
Diferencia Simple vs Individual		0.053 (0.25)	0.056 (0.23)	0.031 (0.43)	0.028 (0.46)
Diferencia Simple vs Grupal		0.071 (0.13)	0.07 (0.14)	0.056 (0.16)	0.04 (0.29)

Diferencia Individual vs Grupal	0.017 (0.72)	0.013 (0.79)	0.024 (0.52)	0.012 (0.73)
Cualquier Tratamiento	-0.017 (0.52)	-0.016 (0.56)	0.035 (0.11)	0.031 (0.14)
Controles	No	Sí	No	Sí
FE: estratos	X	X	X	X
Errores agrupados por: rbd	X	X	X	X
Nivel de Regresión: Tutor	Tutor	Tutor	Tutor	Tutor

Notas: Las diferencias en el número de observaciones se deben a que los tutores no responden a todas las preguntas que componen el índice. La tabla presenta dos versiones del índice: la versión IRT, que se construye con las herramientas de la Teoría de Respuesta al Ítem, y la versión MCA, que se construye con las herramientas de Análisis Multicomponente. El "Niño Ayuda" captura la ayuda que los tutores dan a sus hijos con respecto a las actividades escolares. Consta de cuatro componentes (véase el cuadro B.11 del apéndice para las regresiones por componentes del índice): "Ayuda a w. "Tecnología" se refiere a la ayuda del tutor con respecto a los problemas tecnológicos del niño que utiliza dispositivos informáticos. "Ayuda a Órgano" tiene que ver con el apoyo que el tutor da a la organización de las sesiones de estudio y a la designación de espacios de estudio. "Ayuda a estudiar Habs." se refiere a la ayuda que el tutor brinda para que el niño adquiera hábitos de estudio saludables. Y "Ayuda a la actividad" se relaciona con el apoyo que brinda el tutor en actividades escolares como tareas y preparación de exámenes. Todos los componentes siguen la misma dirección, es decir, un aumento en el valor del componente significa un aumento en el rasgo deseado. Por lo tanto, contribuye en la misma dirección para el índice. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Análisis de Incumplimiento

Hubo cuatro escuelas que inicialmente fueron asignadas a un grupo de tratamiento, pero decidieron no recibir el tratamiento. Se realizó un análisis de variables instrumentales para tener en cuenta en las estimaciones el incumplimiento de estas escuelas con su estado de tratamiento. Los resultados son muy similares, ya que el número de escuelas que no cumplen con las normas es relativamente pequeño. Por ejemplo, los principales resultados de las puntuaciones de las pruebas son casi idénticos (véase el cuadro 22), incluidos los efectos negativos en las puntuaciones lingüísticas.

Tabla 22: Efectos instrumentados del tratamiento efectivo en los puntajes del SIMCE 2022

		Promedio Puntaje Matemáticas	Promedio Puntaje Matemáticas	Promedio Puntaje Lenguaje	Promedio Puntaje Lenguaje
		(1)	(2)	(3)	(4)
Conecta Ideas Simple		-0.055 (0.050)	-0.053 (0.043)	-0.016 (0.052)	0.005 (0.046)
Incentivo Individual		-0.057 (0.048)	-0.029 (0.043)	-0.117** (0.047)	-0.085** (0.040)
Incentivo Grupal		-0.061 (0.059)	-0.073 (0.047)	-0.042 (0.046)	-0.048 (0.037)
N		7,848	7,848	7,533	7,533
Media grupo ref.	de	-0.05	-0.05	-0.07	-0.07
Diferencia Simple Indiv.	vs	0.002 (0.97)	0.024 (0.6)	0.101 (0.09)	0.09 (0.09)
Diferencia Simple Grupal	vs	0.006 (0.92)	0.02 (0.69)	0.025 (0.67)	0.053 (0.3)
Diferencia Indiv. Grupal	vs	0.004 (0.95)	0.044 (0.38)	0.075 (0.18)	0.037 (0.44)
Cualquier Tratamiento		-0.057 (0.16)	-0.052 (0.12)	-0.058 (0.1)	-0.041 (0.16)
Controles FE: estratos	No X	Sí X	No X	Sí X	
Errores agrupados por: rbd	X	X	X	X	
Nivel de Regresión: Clase	Clase	Clase	Clase	Clase	

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Las columnas (3) y (4) tienen menos observaciones que las columnas (1) y (2) ya que no todos los estudiantes que tomaron la prueba de Matemáticas tomaron la prueba de Lenguaje. Instrumentamos si la escuela fue realmente tratada con la asignación de tratamiento aleatorio. Cada columna representa la estimación instrumentada. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje

de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Análisis de deserción

Resultado primario

En general, no hay diferencia en el número de estudiantes que rindieron la prueba SIMCE 2022 entre los grupos (ver Tabla 10). Además, las características de los estudiantes que rindieron el examen SIMCE son similares en todos los grupos (ver Tabla 11). Por lo tanto, no hay evidencia de que el grupo de tratamiento influyera en la probabilidad de que los estudiantes tomaran la prueba SIMCE o que cambiara la composición de quienes la tomaron.

Tabla 23: Balance de la deserción en el SIMCE 2022 (Media y Desviación)

	Control	Conecta Ideas (CI)	Incentivo Individual (II)	Incentivo Grupal (GI)	p-valor (igualdad)
Estudiantes con puntaje en Matemáticas	28.15 (10.64) [144]	27.53 (10.50) [47]	27.47 (6.85) [47]	25.60 (8.04) [47]	0.571
Estudiantes con puntaje en Lenguaje	27.01 (10.42) [144]	26.38 (10.32) [47]	26.49 (6.81) [47]	24.51 (7.66) [47]	0.512
Estudiantes con Cuestionario	29.06 (10.08) [144]	28.68 (10.74) [47]	28.57 (6.94) [47]	26.79 (7.81) [47]	0.607
Profesores con Cuestionario	0.40 (0.51) [144]	0.32 (0.52) [47]	0.26 (0.44) [47]	0.28 (0.50) [47]	0.858
Tutores con Cuestionario	26.06 (9.39) [144]	25.72 (10.61) [47]	24.57 (7.58) [47]	23.19 (7.14) [47]	0.493

Notas: Las columnas 1 a 4 muestran la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) para cada uno de los siguientes grupos: Control, estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta (IC), estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos individuales (II) y estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos grupales (GI). En el grupo de tratamiento, solo se incluyen los datos de las secciones tratadas, es decir, se excluyen los estudiantes de las secciones control de las escuelas tratadas. Los errores estándar, agrupados a nivel de escuela, están entre paréntesis. Las columnas 5 muestran el valor p de la prueba

de si la media es la misma en todos los grupos. La prueba de igualdad considera el diseño de aleatorización (es decir, incluye efectos fijos de estratos). Los datos están a nivel de sección., * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tabla 24: Balance características de los estudiantes SIMCE 2022

	Control	Conecta Ideas (CI)	Incentivo Individual (II)	Incentivo Grupal (GI)	p-valor (igualdad)
Femenino	0.51 (0.50) [4,053]	0.46 (0.50) [1,295]	0.47 (0.50) [1,291]	0.50 (0.50) [1,208]	0.316
Edad	9.65 (1.62) [4,053]	9.61 (1.24) [1,295]	9.64 (1.08) [1,291]	9.70 (2.58) [1,208]	0.742
Madre con Educación Profesional	0.40 (0.49) [3,550]	0.40 (0.49) [1,134]	0.37 (0.48) [1,078]	0.39 (0.49) [1,007]	0.951
Padre con Educación Profesional	0.34 (0.47) [3,487]	0.34 (0.47) [1,124]	0.31 (0.46) [1,061]	0.35 (0.48) [995]	0.691
Tutores de Bajos Ingresos	0.44 (0.50) [4,053]	0.48 (0.50) [1,295]	0.43 (0.49) [1,291]	0.44 (0.50) [1,208]	0.320
Escuela Rural	0.00 (0.00) [4,053]	0.00 (0.00) [1,295]	0.01 (0.11) [1,291]	0.00 (0.00) [1,208]	0.795
Escuela Pública	0.35 (0.48) [4,053]	0.47 (0.50) [1,295]	0.37 (0.48) [1,291]	0.37 (0.48) [1,208]	0.621
Escuela Privada Subvencionada	0.57 (0.49) [4,053]	0.46 (0.50) [1,295]	0.57 (0.50) [1,291]	0.55 (0.50) [1,208]	0.663

Notas: Las columnas 1 a 4 muestran la media, la desviación estándar (entre paréntesis) y el número de observaciones (entre corchetes) para cada uno de los siguientes grupos: Control, estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta (IC), estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos individuales (II) y estudiantes en secciones tratadas en escuelas tratadas con Ideas Conecta + incentivos grupales (GI). En el grupo de tratamiento, solo se incluyen los datos de las secciones

tratadas, es decir, se excluyen los estudiantes de las secciones control de las escuelas tratadas. Los errores estándar, agrupados a nivel de escuela, están entre paréntesis. Las columnas 5 muestran el valor p de la prueba de si la media es la misma en todos los grupos. La prueba de igualdad considera el diseño de aleatorización (es decir, incluye efectos fijos de estratos). Los datos están a nivel de sección., * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Resultados secundarios

Además, el número de estudiantes, maestros y padres de familia que respondieron las encuestas asociadas con la prueba SIMCE también está equilibrado entre los grupos (ver Tabla 10), lo que sugiere que la deserción diferencial tampoco es un problema para los resultados secundarios.

Deserción igualada SIMCE 2022

La probabilidad de que los estudiantes que rindieron la prueba SIMCE 2022 puedan ser emparejados con la prueba APTUS desde la línea de base está equilibrada entre los grupos (ver Tabla 25). Es decir, la proporción de estudiantes que podemos emparejar es similar en todos los grupos.

Tabla 25: Efectos del tratamiento sobre la deserción (submuestra emparejada)

	Emparejado (1)	Emparejado (2)
Conecta Ideas Simple	-0.040 (0.033)	-0.034 (0.028)
Incentivo Individual	0.026 (0.028)	0.026 (0.026)
Incentivo Grupal	0.008 (0.026)	0.001 (0.023)
N	7,848	7,848
Media grupo de referencia	0.7	0.7
Diferencia Simple vs Individual	0.066 (0.07)	0.066 (0.07)
Diferencia Simple vs Grupal	0.048 (0.17)	0.048 (0.17)
Diferencia Individual vs Grupal	0.018 (0.56)	0.018 (0.56)
Cualquier Tratamiento	-0.002 (0.91)	-0.003 (0.88)
Controles	No	Sí
FE: estratos	X	X

Errores agrupados por: rbd	X	X
Nivel de Regresión	Estudiante	Estudiante

Notas: Las estimaciones se realizan con el universo de estudiantes emparejados que nos dio la Agencia de la Calidad de la Educación. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Análisis de subgrupos

Para estimar los efectos heterogéneos del tratamiento en todos los subgrupos, estimamos la ecuación [eq:het], interactuando las variables ficticias de tratamiento con un indicador de NSE alto (por encima de la mediana) y de género. No hay heterogeneidad en los efectos del tratamiento según el estado del NSE en las puntuaciones de matemáticas, sin embargo, el efecto negativo en las puntuaciones de lenguaje parece concentrarse en los estudiantes de NSE bajo (véase la Tabla 26). No hay heterogeneidad en los efectos del tratamiento según el sexo (ver Tabla 27).

Tabla 26: Heterogeneidad de los efectos del tratamiento por nivel socioeconómico (SES)

	Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje
	(1)	(2)
Conecta Simple	-0.077 (0.053)	-0.020 (0.059)
Incentivo Individual	-0.038 (0.051)	-0.133** (0.049)
Incentivo Grupal	-0.093 (0.058)	-0.108** (0.049)
SES Alto	0.213*** (0.037)	0.185*** (0.041)
Conecta Simple x SES Alto	0.038 (0.063)	0.033 (0.074)

Incentivo Individual x SES Alto	-0.024 (0.072)	0.095 (0.071)
Incentivo Grupal x SES Alto	0.049 (0.078)	0.125* (0.071)
N	6,832	6,739
Media grupo de ref.	-0.05	-0.07
Diferencia Simple vs Indiv. x SES Alto	0.062 (0.44)	0.062 (0.48)
Diferencia Simple Grupal vs Indiv. x SES Alto	0.011 (0.9)	0.092 (0.29)
Diferencia Indiv. vs Grupal x SES Alto	0.073 (0.43)	0.03 (0.72)
Cualquier Tratamiento	-0.069 (0.09)	-0.086 (0.02)
Controles	Sí	Sí
FE: estratos	X	X
Errores agrupados por: rbd	X	X
Heterogeneidad	Ingresos	Ingresos
Nivel de Regresión	Estudiante	Estudiante

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Un nivel socioeconómico alto es igual a uno si el estudiante está por encima de la mediana de un indicador de riqueza. El proxy patrimonial que utilizamos se construye utilizando la metodología MCA (Abdie & Valentin, 2007). Los componentes que constituyen el índice son dummies del nivel educativo de la madre y del padre y una variable categórica de distintos niveles de ingresos reportados por el tutor. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el

cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tabla 27: Heterogeneidad de los efectos del tratamiento por género

		Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje
		(1)	(2)
Conecta Simple	Ideas	-0.039 (0.048)	0.046 (0.053)
Incentivo Individual		-0.034 (0.052)	-0.100** (0.049)
Incentivo Grupal		-0.079 (0.052)	-0.043 (0.047)
Femenino		-0.213*** (0.032)	0.125*** (0.030)
Conecta Simple Femenino	Ideas x	-0.041 (0.057)	-0.085 (0.053)
Incentivo Individual Femenino	x	0.000 (0.064)	0.024 (0.057)
Incentivo Grupal Femenino	x	0.012 (0.063)	-0.003 (0.061)
N		7,848	7,533
Media grupo de referencia		0.05	-0.14
Diferencia Simple vs Indiv. x Femenino		0.04 (0.58)	0.109 (0.09)
Diferencia Simple Grupal Femenino		vs x 0.052 (0.47)	0.083 (0.22)

Diferencia Indiv. vs Grupal x Femenino	0.012 (0.88)	0.027 (0.71)
Cualquier Tratamiento	-0.05 (0.17)	-0.031 (0.37)
Controles	Sí	Sí
FE: estratos	X	X
Errores agrupados por: rbd	X	X
Heterogeneidad	Género	Género
Nivel Regresión	de Estudiante	Estudiante

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A continuación, se estudia si existe heterogeneidad en los efectos del tratamiento por el desempeño histórico de la escuela. Específicamente, clasificamos las escuelas en función de sus puntajes SIMCE 2018 en quintiles y estudiamos la heterogeneidad en los efectos del tratamiento por estos quintiles. No hay una tendencia clara. En algunos casos, los efectos son mayores para los quintiles superiores (p. ej., incentivos grupales en matemáticas) y en otros, son peores (p. ej., incentivos individuales en lenguaje). Sin embargo, ningún grupo tiene un efecto positivo y significativo del tratamiento (Tabla 28). Estos patrones son similares si utilizamos la prueba de referencia APTUS para clasificar las escuelas en quintiles (Tabla 29).

Tabla 28: Heterogeneidad de los efectos del tratamiento por quintiles del SIMCE 2018

	Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje
	(1)	(2)
Conecta Ideas Simple x q1	-0.006 (0.077)	0.032 (0.081)
Conecta Ideas Simple x q2	0.006 (0.081)	-0.056 (0.072)
Conecta Ideas Simple x q3	-0.154 (0.103)	0.002 (0.112)
Conecta Ideas Simple x q4	-0.042 (0.082)	0.049 (0.086)
Conecta Ideas Simple x q5	-0.036 (0.089)	0.043 (0.142)
Incentivo Individual x q1	-0.048 (0.088)	-0.145* (0.074)
Incentivo Individual x q2	-0.061 (0.088)	-0.153* (0.079)
Incentivo Individual x q3	0.007 (0.074)	-0.032 (0.073)
Incentivo Individual x q4	-0.103 (0.109)	-0.058 (0.100)
Incentivo Individual x q5	0.054	-0.109**

	(0.226)	(0.039)
Incentivo Grupal x q1	-0.079	-0.077
	(0.137)	(0.090)
Incentivo Grupal x q2	-0.203**	-0.135*
	(0.084)	(0.079)
Incentivo Grupal x q3	0.037	0.048
	(0.087)	(0.087)
Incentivo Grupal x q4	-0.171*	-0.056
	(0.091)	(0.069)
Incentivo Grupal x q5	0.117	0.007
	(0.118)	(0.081)
N	7,827	7,511
Media grupo de referencia	-0.05	-0.07
Controles	Sí	Sí
FE: quintil SIMCE 2018	X	X
Errores agrupados por: rbd	X	X
Heterogeneidad	SIMCE 2018	SIMCE 2018
Nivel de Regresión	Estudiante	Estudiante

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Cada uno representa el quintil de la distribución de puntajes escolares del SIMCE 2018. Se controla el promedio de matemáticas y lenguaje, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tenía la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, el tipo de escuela (municipal o subsidiada) y el nivel socioeconómico bajo. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tabla 29: Heterogeneidad por quintil de la prueba APTUS

	Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje
	(1)	(2)
Conecta Ideas Simple x q1	-0.187 (0.133)	-0.142 (0.104)
Conecta Ideas Simple x q2	-0.018 (0.081)	0.037 (0.056)
Conecta Ideas Simple x q3	-0.072 (0.106)	-0.075 (0.094)
Conecta Ideas Simple x q4	0.065 (0.083)	0.105 (0.111)
Conecta Ideas Simple x q5	-0.181** (0.079)	-0.011 (0.107)
Incentivo Individual x q1	-0.182* (0.102)	-0.202* (0.111)
Incentivo Individual x q2	-0.031 (0.096)	-0.125* (0.075)
Incentivo Individual x q3	-0.064 (0.084)	-0.077 (0.087)
Incentivo Individual x q4	0.022 (0.102)	-0.058 (0.086)
Incentivo Individual x q5	-0.027	-0.081

	(0.118)	(0.083)
Incentivo Grupal x q1	-0.198*	-0.047
	(0.102)	(0.100)
Incentivo Grupal x q2	-0.088	-0.013
	(0.072)	(0.114)
Incentivo Grupal x q3	-0.189*	-0.136*
	(0.096)	(0.076)
Incentivo Grupal x q4	0.068	0.014
	(0.117)	(0.071)
Incentivo Grupal x q5	-0.071	-0.092
	(0.124)	(0.075)
N	7,849	7,532
Media grupo de referencia	-0.05	-0.07
Controles	Sí	Sí
FE: quintil APTUS	X	X
Errores agrupados por: rbd	X	X
Heterogeneidad	Escuela APTUS	Escuela APTUS
Nivel de Regresión	Estudiante	Estudiante

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Cada uno representa el quintil de la distribución a nivel de escuela de la puntuación APTUS IRT utilizada para la estratificación. Controlamos los siguientes aspectos: los puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tenía la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, el tipo de escuela (municipal o subvencionada) y el nivel socioeconómico bajo. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finalmente, utilizando solo los datos de los estudiantes que fueron emparejados entre la línea de base del APTUS y la prueba SIMCE 2022, estudiamos la heterogeneidad según el desempeño de la línea de base (ver Tabla 30). Esto sugiere un efecto positivo y significativo del tratamiento para los estudiantes en el decil superior para el tratamiento de Ideas Simples Conecta en lenguaje y para el tratamiento de incentivos individuales en Matemáticas. Sin embargo, estos efectos positivos deben tomarse con cautela debido a las pruebas de hipótesis múltiples y al hecho de que muchos otros grupos tuvieron efectos negativos del tratamiento que son tan grandes o mayores.

Tabla 30: Heterogeneidad por decil de rendimiento en la prueba APTUS

		Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje
		(1)	(2)
Conecta Simple	Ideas x q1	0.000 (0.106)	0.083 (0.118)
Conecta Simple	Ideas x q2	-0.063 (0.082)	0.047 (0.113)
Conecta Simple	Ideas x q3	0.150* (0.085)	0.109 (0.109)
Conecta Simple	Ideas x q4	-0.020 (0.075)	-0.055 (0.096)
Conecta Simple	Ideas x q5	-0.075 (0.078)	0.077 (0.099)
Conecta Simple	Ideas x q6	-0.140* (0.080)	0.024 (0.092)
Conecta Simple	Ideas x q7	0.024 (0.081)	0.065 (0.103)
Conecta Simple	Ideas x q8	0.055 (0.082)	0.068 (0.106)
Conecta Simple	Ideas x q9	-0.016 (0.079)	0.027 (0.097)

Conecta Ideas Simple x q10	0.094 (0.079)	0.204* (0.107)
Incentivo Individual x q1	0.094 (0.102)	-0.076 (0.115)
Incentivo Individual x q2	-0.075 (0.063)	-0.068 (0.083)
Incentivo Individual x q3	0.069 (0.072)	-0.075 (0.090)
Incentivo Individual x q4	0.050 (0.080)	-0.056 (0.091)
Incentivo Individual x q5	-0.120 (0.087)	-0.280** (0.113)
Incentivo Individual x q6	-0.072 (0.090)	-0.126 (0.090)
Incentivo Individual x q7	-0.044 (0.093)	-0.028 (0.099)
Incentivo Individual x q8	-0.137* (0.081)	-0.146 (0.129)
Incentivo Individual x q9	-0.059 (0.085)	-0.034 (0.101)
Incentivo Individual x q10	0.212** (0.094)	0.079 (0.086)

Incentivo Grupal x q1	0.083 (0.099)	0.197* (0.107)
Incentivo Grupal x q2	-0.030 (0.081)	0.038 (0.102)
Incentivo Grupal x q3	-0.018 (0.074)	0.048 (0.073)
Incentivo Grupal x q4	-0.103 (0.109)	-0.058 (0.100)
Incentivo Grupal x q5	0.054 (0.226)	-0.109** (0.039)
Incentivo Grupal x q6	-0.072 (0.090)	-0.126 (0.090)
Incentivo Grupal x q7	-0.044 (0.093)	-0.028 (0.099)
Incentivo Grupal x q8	-0.137* (0.081)	-0.146 (0.129)
Incentivo Grupal x q9	-0.059 (0.085)	-0.034 (0.101)
Incentivo Grupal x q10	0.212** (0.094)	0.079 (0.086)
N	7,849	7,532
FE: quintil APTUS	X	X
Errores agrupados por: rbd	X	X
Media grupo de referencia	-0.05	-0.07
Controles	Sí	Sí
Heterogeneidad	Escuela APTUS	Escuela APTUS
Nivel de Regresión	Estudiante	Estudiante

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Cada uno representa el decil de la puntuación del IRT de APTUS a nivel de estudiante. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

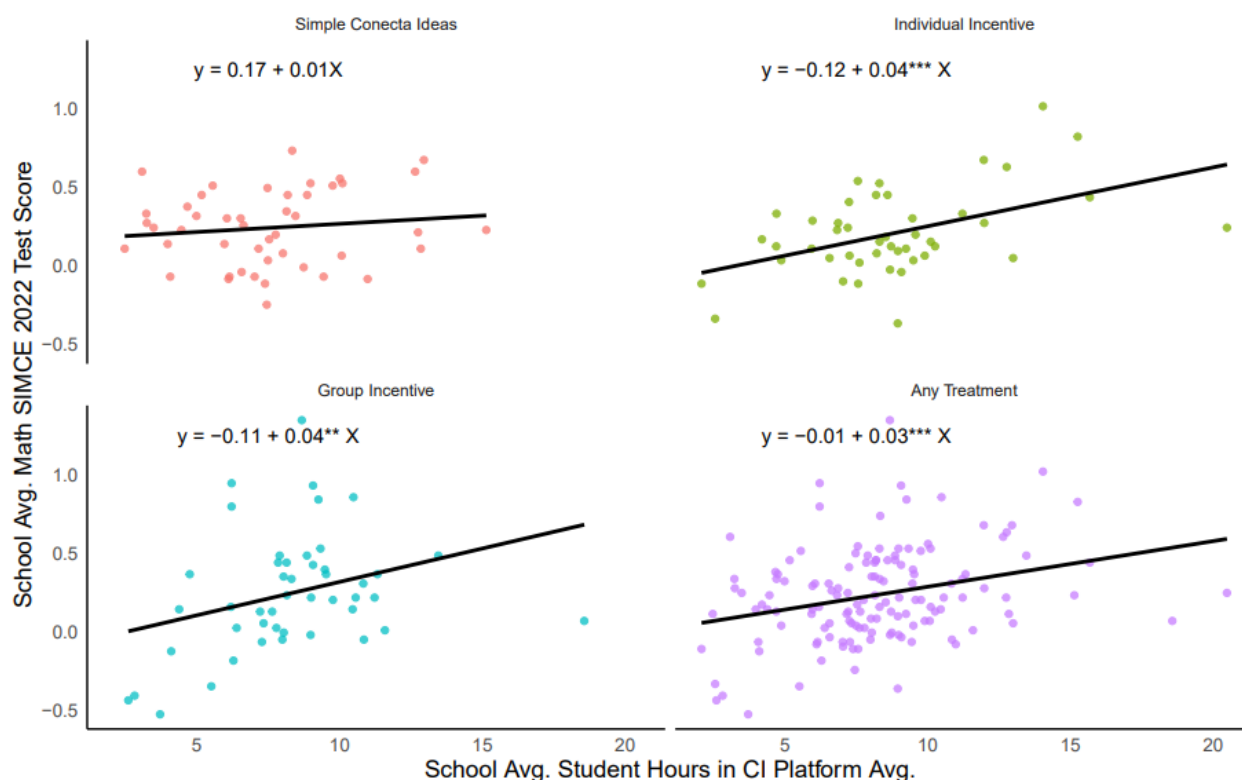
Análisis de Dosis-Respuesta

En este apartado se explora si existe relación entre la intensidad del tratamiento y el desempeño en la prueba SIMCE 2022. Definimos la intensidad del tratamiento de dos maneras: el promedio de la escuela de las horas promedio de los estudiantes que pasan en la plataforma Conecta Ideas y el promedio de la escuela del total promedio de preguntas intentadas de los estudiantes en la plataforma Conecta Ideas.¹⁵ Este análisis no es causal ya que existe endogeneidad en las variables de intensidad (es decir, la intensidad del tratamiento no se asigna exógenamente y puede correlacionarse con variables no observables que tienen un efecto en los puntajes del SIMCE).

En primer lugar, estudiamos cómo el desempeño del SIMCE 2022 se correlaciona con la intensidad del tratamiento (ver Figuras 6 y 7). Vemos una correlación positiva entre la intensidad de uso y los puntajes promedio de las pruebas estandarizadas de la escuela en 2022. Estos resultados son compatibles con la idea de que un mayor uso de la plataforma Conecta Ideas aumenta el rendimiento en las pruebas SIMCE.

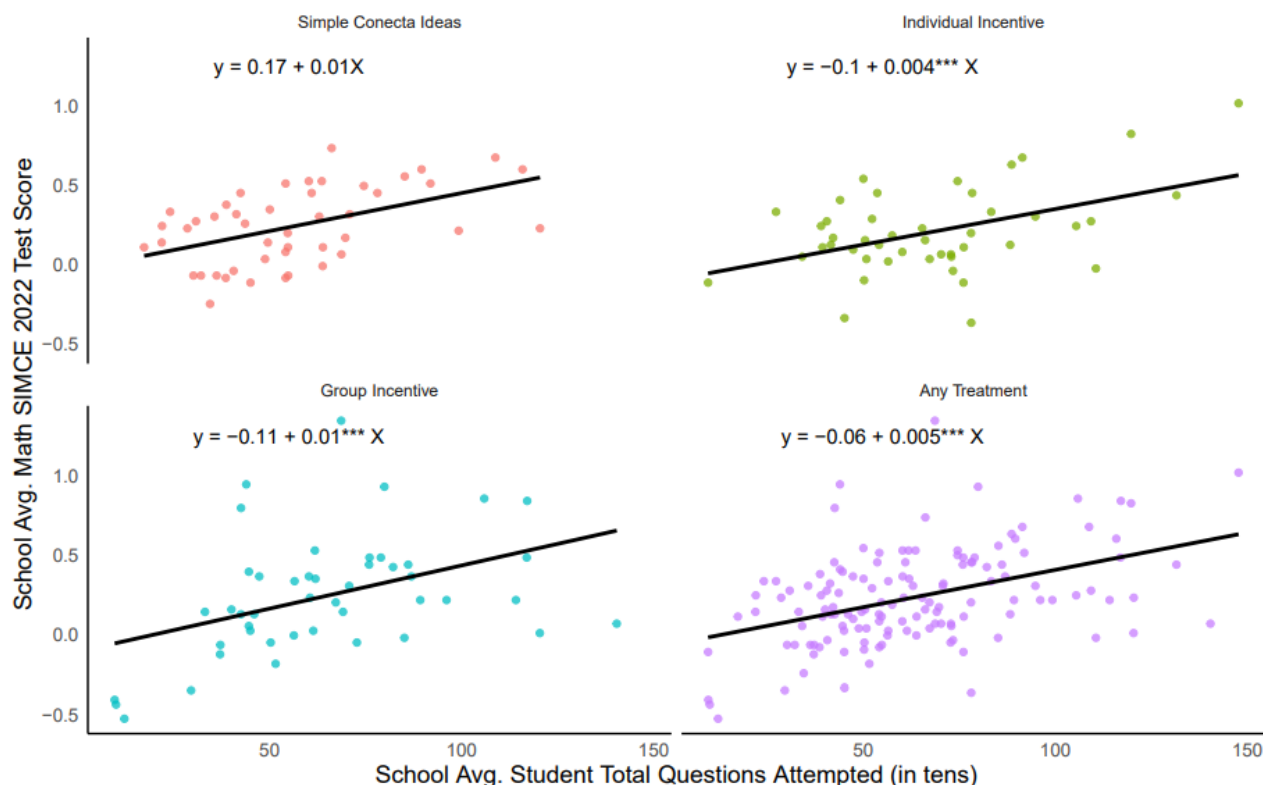
¹⁵ Dado que los datos del SIMCE 2022 no tienen identificadores de estudiantes, no podemos vincular los datos de uso a nivel de estudiante con los datos del SIMCE 2022 a nivel de estudiante. Por lo tanto, solo podemos realizar un análisis dosis-respuesta utilizando datos de uso a nivel escolar.

Figura 6: Intensidad del tratamiento: Correlación entre el tiempo en la plataforma y los puntajes del test de matemáticas SIMCE 2022



Nota: Los puntajes de las pruebas SIMCE están estandarizados en todas las escuelas. El promedio de horas totales de los estudiantes de la escuela se calculó como la media del total de horas dedicadas a la plataforma Conecta Ideas por los estudiantes. La estimación del modelo en cada panel de figuras es una regresión lineal simple. La variable dependiente es el puntaje de la prueba estandarizada de matemáticas SIMCE 2022 de la escuela y la variable independiente es el promedio de horas totales de los estudiantes de la escuela. Son respectivamente los siguientes: 0,01, 0,22 y 0,1. R^2

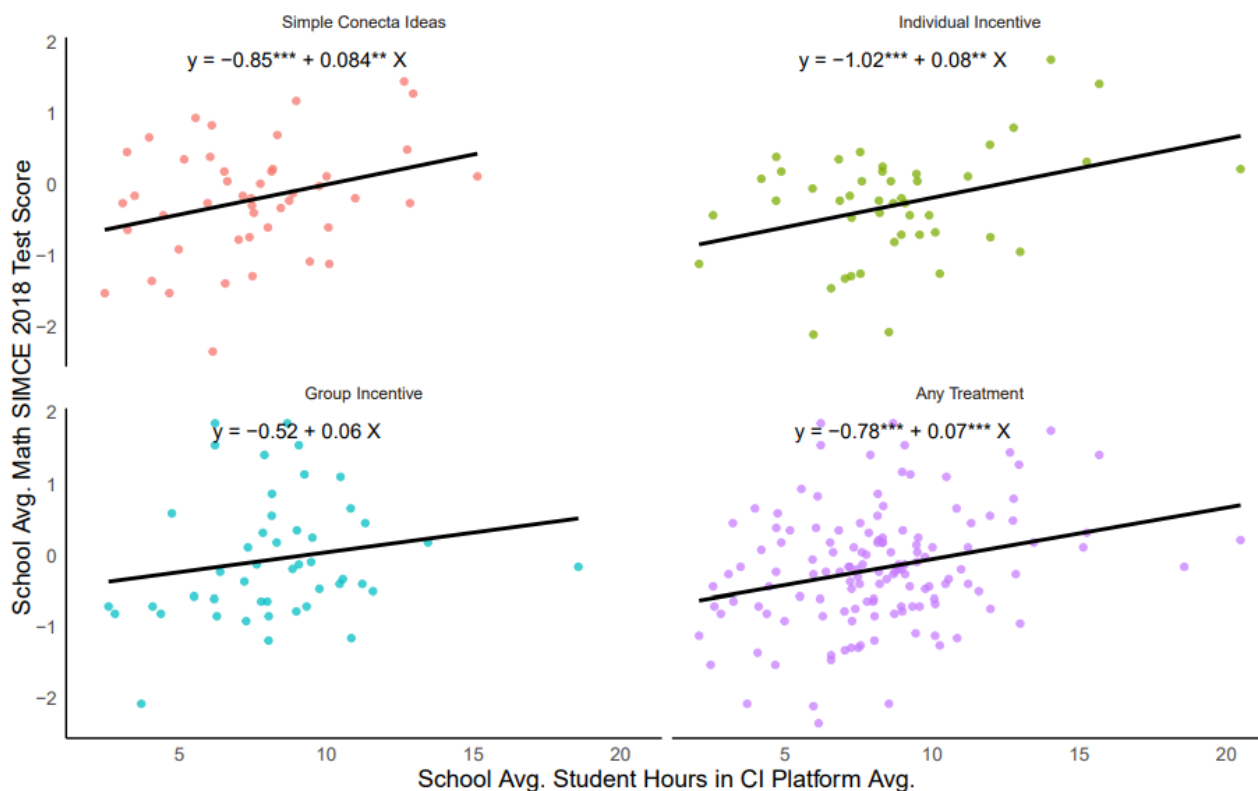
Figura 7: Intensidad del tratamiento: correlación entre las preguntas intentadas y los puntajes de las pruebas de matemáticas SIMCE 2022



Notas: Los puntajes de las pruebas SIMCE están estandarizados en todas las escuelas. El promedio de preguntas intentadas por los estudiantes de la escuela se calculó como la media del total de preguntas intentadas por cada uno de sus estudiantes. El total de preguntas intentadas se dividió por 10 para facilitar la lectura de la figura. La estimación del modelo en cada panel de figuras es una regresión lineal donde la variable dependiente es el puntaje promedio estandarizado de la prueba de matemáticas SIMCE 2022 de la escuela (usando solo las secciones tratadas) y la variable independiente es el total de preguntas intentadas por el promedio de estudiantes de la escuela. Son respectivamente los siguientes: 0,24, 0,22 y 0,18. R^2

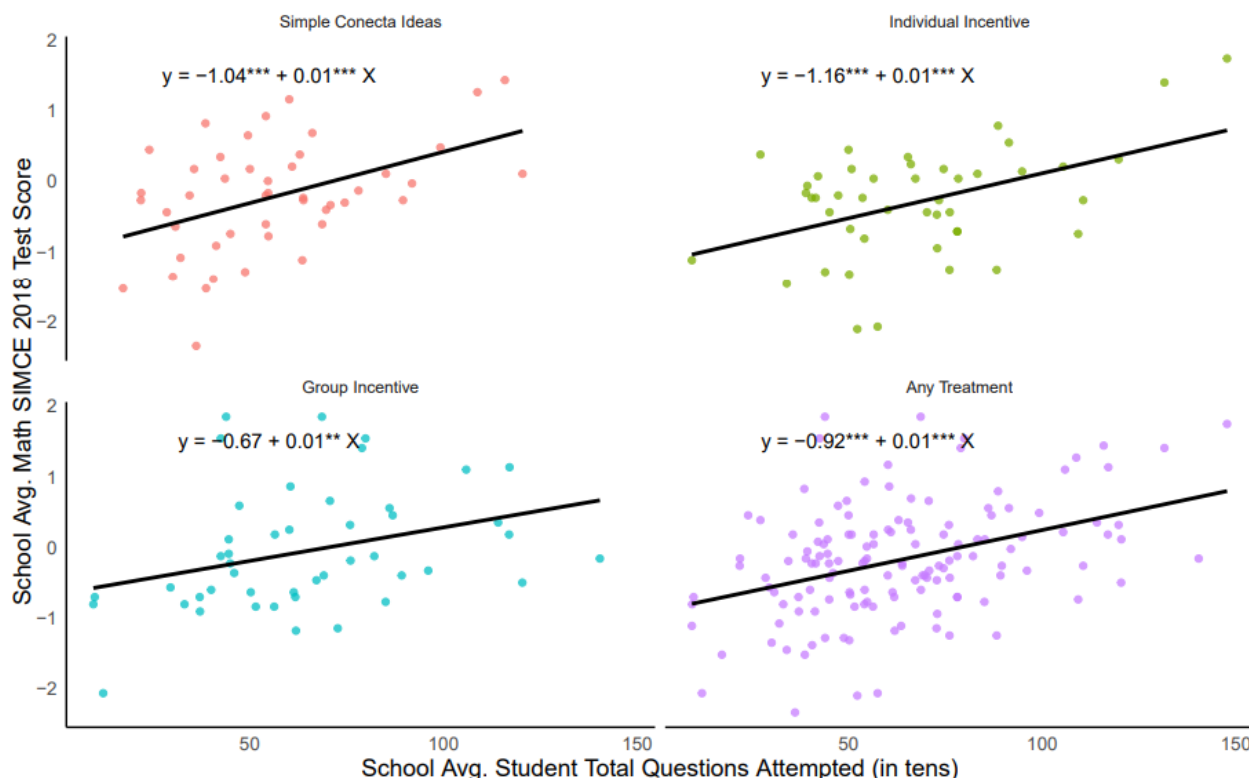
Sin embargo, esta correlación positiva podría deberse a factores no observables que impulsan tanto el aumento de Ideas Conecta como el desempeño en el examen SIMCE. Por ejemplo, las escuelas bien administradas o las escuelas con estudiantes más adinerados pueden usar más la plataforma y tener mejores puntajes SIMCE. Por lo tanto, incluso si la plataforma Conecta Ideas no tiene ningún efecto en las puntuaciones cognitivas de los estudiantes, es posible tener una correlación positiva entre las puntuaciones y el uso de la plataforma. Por ejemplo, realizamos el mismo análisis, pero con los puntajes de las pruebas SIMCE 2018 (es decir, la correlación entre el uso de las ideas de Conecta en 2022 y los puntajes de SIMCE 2018). En general, también existe una correlación positiva (véanse las Figuras 8 y 9), lo que sugiere que las características de los estudiantes y las escuelas podrían estar impulsando tanto las puntuaciones más altas (en 2018 y 2022) como el mayor uso de la plataforma (en lugar de un mayor uso que impulse las puntuaciones más altas en 2022).

Figura 8: Intensidad del tratamiento: Correlación entre el tiempo en la plataforma y los puntajes del test de matemáticas SIMCE 2018



Notas: Los puntajes de las pruebas SIMCE están estandarizados en todas las escuelas. El promedio de horas totales de los estudiantes de la escuela se calculó como la media del total de horas dedicadas a los estudiantes de la plataforma Conecta Ideas. La estimación del modelo en cada panel de figuras es una regresión lineal simple. La variable dependiente es el puntaje de la prueba estandarizada de matemáticas SIMCE 2018 de la escuela y la variable independiente es el promedio de horas totales de los estudiantes de la escuela. Son respectivamente los siguientes: 0,09, 0,13 y 0,03. Consulte la Figura R^226 del Apéndice para obtener una versión con los puntajes de las pruebas APTUS.

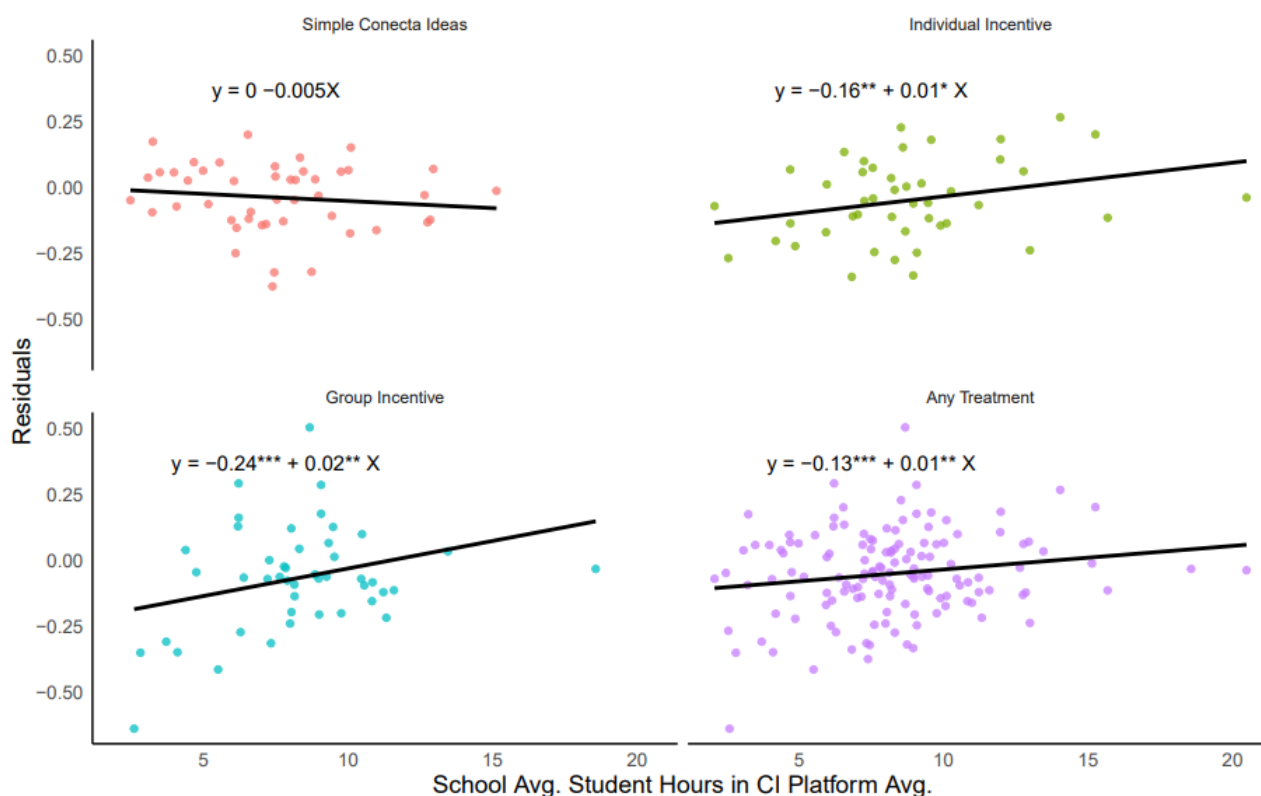
Figura 9: Intensidad del tratamiento: Correlación entre el tiempo en la plataforma y los puntajes de la prueba de matemáticas SIMCE 2018



Notas: Los puntajes de las pruebas SIMCE están estandarizados en todas las escuelas. El promedio de preguntas intentadas por los estudiantes de la escuela se calculó como la media del total de preguntas intentadas por cada uno de sus estudiantes. El total de preguntas intentadas se dividió en 10 para facilitar la lectura de la figura. La estimación del modelo en cada panel de figuras es una regresión lineal donde la variable dependiente es el puntaje promedio estandarizado de la prueba de matemáticas SIMCE 2018 de la escuela y la variable independiente es el total de preguntas intentadas por el promedio de estudiantes de la escuela. Son respectivamente los siguientes: 0,20, 0,21 y 0,10. Consulte la Figura R²²⁷ del Apéndice para obtener una versión con los puntajes de las pruebas APTUS

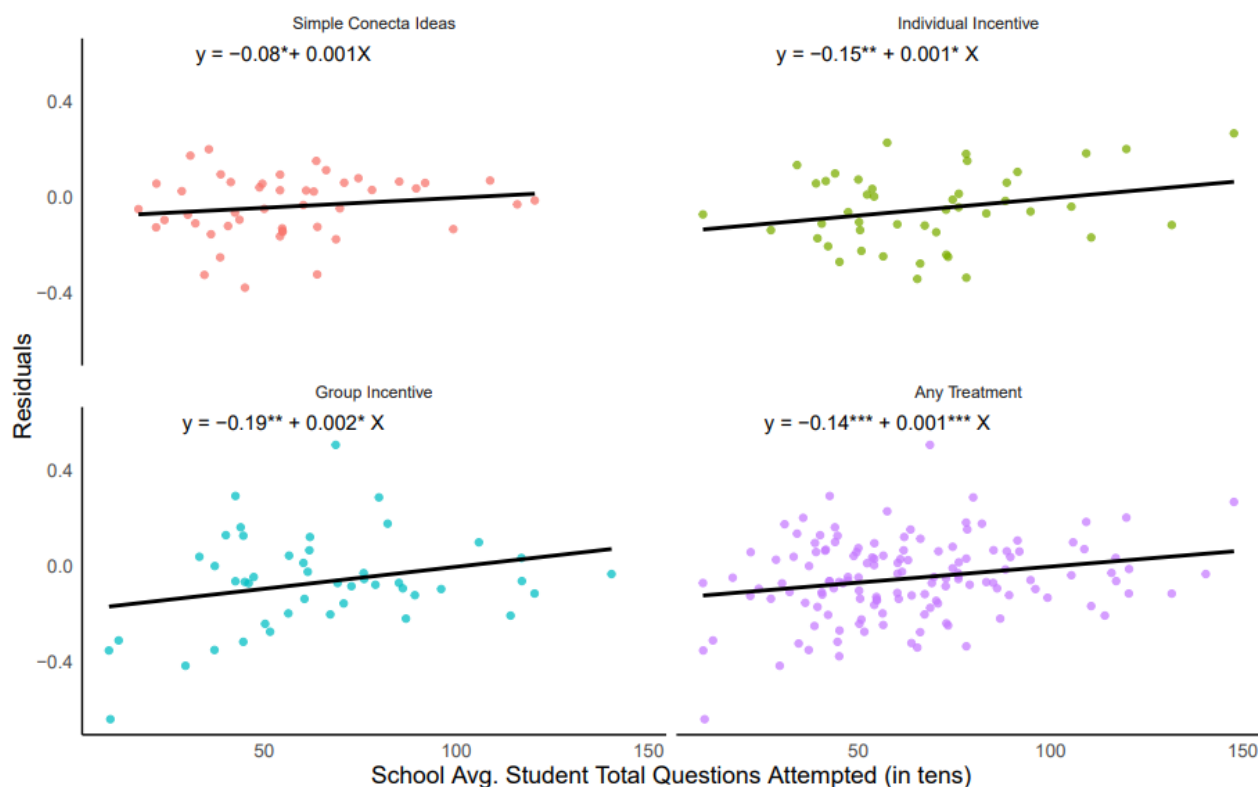
Intentamos desentrañar si existe una relación causal entre la intensidad de Conecta Ideas y el rendimiento de las pruebas en 2022 a través de varios medios, todos imperfectos. En primer lugar, estudiamos si los puntajes del SIMCE 2022 están por encima o por debajo de lo que esperamos que estén en función de las características de la escuela. Específicamente, estimamos una regresión lineal de los puntajes de las pruebas SIMCE 2022 contra los puntajes de SIMCE 2018 y otras características de la escuela utilizando datos del grupo control. A continuación, utilizamos el coeficiente estimado para predecir las puntuaciones de 2022 para todas las escuelas de nuestros experimentos (tratadas y de control) y calcular los residuos (es decir, la desviación de sus puntuaciones previstas). Finalmente, se estudia la correlación entre el uso de Conecta Ideas y estos residuos. En general, los residuos están ligeramente correlacionados con la intensidad del tratamiento. Para la versión de tratamiento sin incentivos individuales o grupales, el tiempo en la plataforma parece disminuir el rendimiento en la prueba SIMCE 2022. Para los grupos con incentivos, hay una relación positiva. En los tres grupos de tratamiento, una hora adicional de uso de Conecta Ideas se traduce en un aumento de $0,01\sigma$ en las puntuaciones de las pruebas. Esto está en línea con hallazgos previos de Conecta Ideas que documentaron un efecto de 0.27σ y el estudiante promedio en el grupo de tratamiento usó la plataforma durante aproximadamente 28 horas (Araya et al., 2019).

Figura 10: Correlación entre los residuales de desempeño del SIMCE 2022 y el tiempo en plataforma



Notas: La estimación del modelo en cada panel de figura es una regresión lineal donde la variable dependiente es el residuo de su desempeño esperado en el SIMCE 2022 y su desempeño real, y la variable independiente es el promedio de horas totales de los estudiantes de la escuela. Son respectivamente los siguientes: 0,01, 0,08 y 0,09. Véase la Figura R²²⁸ del Apéndice para el mismo análisis, pero los residuales provienen de la regresión de los puntajes del SIMCE 2022 sobre los puntajes del APTUS.

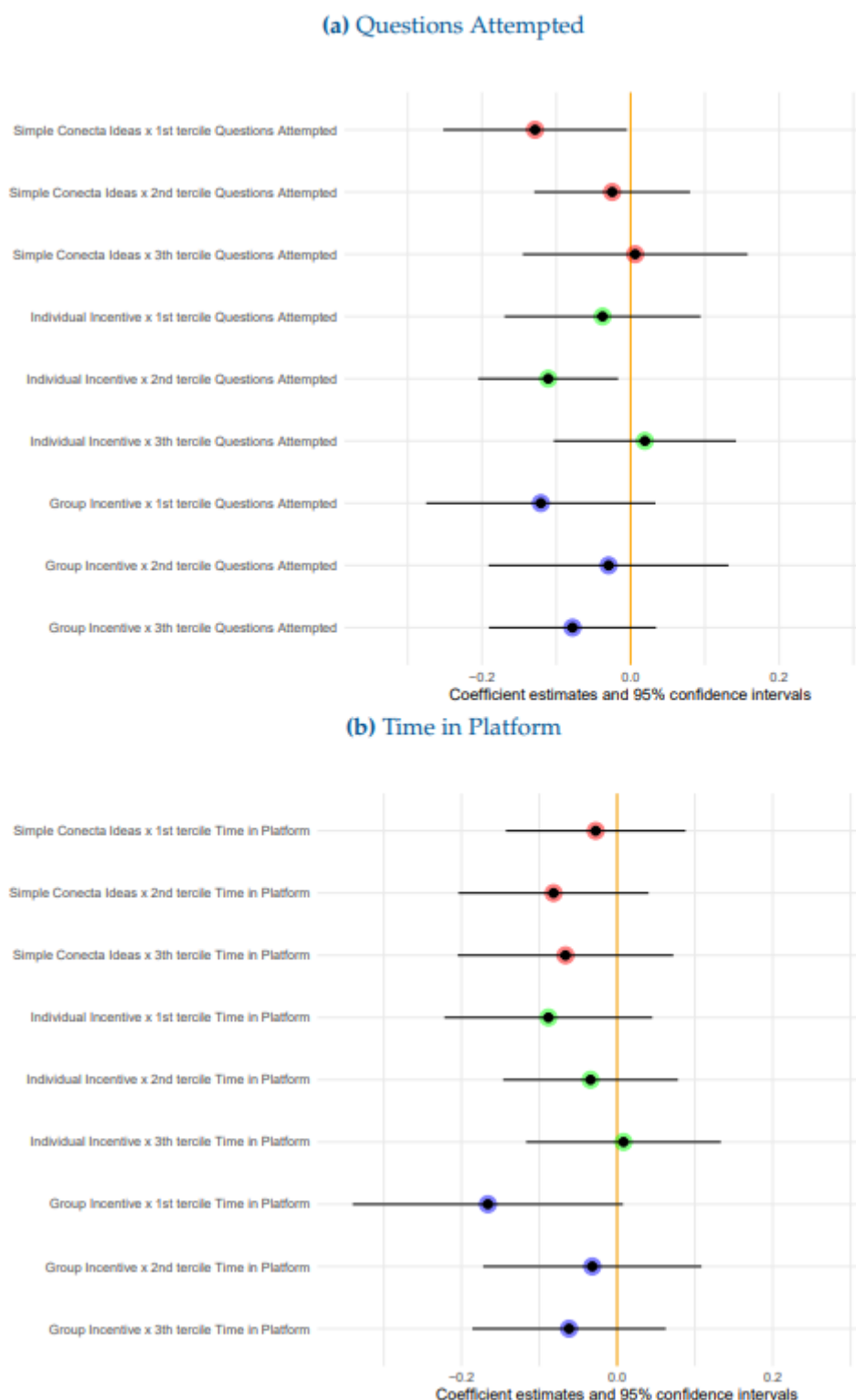
Figura 11: Correlación entre los residuales de desempeño del SIMCE 2022 y las preguntas intentadas.



Notas: La estimación del modelo en cada panel de la figura es una regresión lineal donde la variable dependiente es el residuo de su rendimiento esperado en el SIMCE 2022 y su rendimiento real, y la variable independiente es el promedio de preguntas totales intentadas por los estudiantes de la escuela. Los R^2 son respectivamente los siguientes: 0.03, 0.07 y 0.07. Ver Apéndice Figura B.4 para el mismo análisis, pero los residuos provienen de la regresión de los puntajes del SIMCE 2022 sobre los puntajes de APTUS.

A continuación, se realizó un análisis de heterogeneidad para estimar el efecto (no causal) de los diferentes niveles de tratamiento. Específicamente, clasificamos las escuelas de tratamiento en terciles en función de su uso. A continuación, comparamos cada uno de estos terciles (en cada grupo de tratamiento) con el grupo de control. Esto es equivalente a la interacción de las variables ficticias para cada tercil con las variables ficticias de tratamiento en nuestra regresión principal. En general, los efectos del tratamiento no son mayores (ni positivos) para el tercil más alto de uso. Por otro lado, el tercil más bajo de uso suele tener el efecto de tratamiento más negativo en los PUNTAJES SIMCE 2022 (figura 12). Estos resultados son robustos al utilizar deciles en lugar de terciles (Figuras B.5). En general, estos resultados sugieren que hay un costo fijo para establecer Conecta Ideas, y sin mucho uso, esto conduce a efectos negativos del tratamiento. Sin embargo, incluso los niveles altos de uso (terciles superiores o deciles superiores) no conducen a efectos positivos del tratamiento. Los resultados también son similares si nos centramos en el lenguaje más que en las matemáticas (Figuras B.6 y B.7), es decir, los efectos negativos del tratamiento del lenguaje se concentran en las escuelas de bajo uso.

Figura 12: Efectos del tratamiento según el uso de la plataforma



Notas: Interactuamos la asignación de tratamiento con un maniquí del tercil de distribución de las escuelas de uso de Conecta Ideas. Los coeficientes presentados provienen de la estimación a nivel de estudiante de la regresión lineal donde la variable dependiente es el puntaje de la prueba estandarizada de estudiantes SIMCE 2022. Los puntos son las estimaciones puntuales, mientras que las líneas representan intervalos de confianza del 95%. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela

(municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. Véase la Figura B.5 para la versión del decil del análisis. Las figuras B.6 y B.7 tienen el mismo análisis, pero con el puntaje de la prueba SIMCE 2022 de lengua.

Por último, utilizamos la asignación de tratamiento como instrumento para la intensidad de uso. Dado que cada grupo tiene una sola maniquí de tratamiento, que utilizamos como instrumento para un regresor endógeno continuo (uso), estos efectos deben interpretarse como la respuesta causal promedio. Específicamente, instrumentamos el uso con la asignación aleatoria de tratamientos. En general, no hay un efecto positivo del uso en los puntajes de las pruebas SIMCE (ver Tablas 31 y 32). En cambio, un mayor uso de la plataforma (en términos de preguntas intentadas y tiempo dedicado a la plataforma) afecta negativamente a las puntuaciones de matemáticas para el grupo de tratamiento de incentivos grupales y a las puntuaciones de idiomas para el grupo de incentivos individuales.

Tabla 31: Intensidad del tratamiento como preguntas intentadas en los puntajes de la prueba SIMCE 2022

	Puntaje Promedio Matemáticas	Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje	Puntaje Promedio Lenguaje
	(1)	(2)	(3)	(4)
Conecta Ideas Simple	-0.113	-0.113	-0.031	-0.003
	(0.085)	(0.075)	(0.087)	(0.077)
N	5,322	5,322	5,104	5,104
Incentivo Individual	-0.086	-0.053	-0.156**	-0.111*
	(0.070)	(0.061)	(0.068)	(0.057)
N	5,281	5,281	5,071	5,071
Incentivo Grupal	-0.089	-0.123*	-0.057	-0.069
	(0.090)	(0.070)	(0.069)	(0.056)
N	5,233	5,233	5,019	5,019
Media grupo de referencia	-0.05	-0.05	-0.07	-0.07
Controles	No	Sí	No	Sí
FE: estratos	X	X	X	X
Errores agrupados por: rbd	X	X	X	X
Nivel de Regresión	Estudiante	Estudiante	Estudiante	Estudiante

Notas: Las estimaciones se realizaron para cada grupo de tratamiento individualmente. Por ejemplo, el panel Ideas Simples Conecta hace referencia al uso de instrumentación con la asignación de tratamiento utilizando solo Ideas Simples Conecta y los grupos de control en la estimación, excluyendo los otros grupos de

tratamiento. La variable de preguntas intentadas es en términos de 100 preguntas intentadas por unidad. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. En el Apéndice se muestran los resultados de la estimación de la OLS endógena y de la primera etapa (véanse las tablas B.13, B.14 y B.15), * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tabla 32: Intensidad del tratamiento como tiempo en la aplicación en los puntajes de la prueba SIMCE 2022

	Puntaje Promedio Matemáticas	Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje	Puntaje Promedio Lenguaje
	(1)	(2)	(3)	(4)
Conecta Ideas Simple				
Tiempo en la aplicación	-0.008 (0.006)	-0.008 (0.006)	-0.002 (0.006)	0.000 (0.006)
N	5,322	5,322	5,104	5,104
FE: estratos	X	X	X	X
Errores agrupados por: rbd	X	X	X	X
Incentivo Individual				
Tiempo en la aplicación	-0.007 (0.005)	-0.004 (0.005)	-0.012** (0.005)	-0.009* (0.004)
N	5,281	5,281	5,071	5,071
FE: estratos	X	X	X	X
Errores agrupados por: rbd	X	X	X	X
Incentivo Grupal				
Tiempo en la aplicación	-0.007 (0.007)	-0.010* (0.006)	-0.004 (0.005)	-0.005 (0.004)

N	5,233	5,233	5,019	5,019
FE: estratos	X	X	X	X
Errores agrupados por: rbd	X	X	X	X
Media grupo de referencia	-0.05	-0.05	-0.07	-0.07
Controles	No	Sí	No	Sí
Nivel de Regresión	Estudiante	Estudiante	Estudiante	Estudiante

Notas: Las estimaciones se realizaron para cada grupo de tratamiento individualmente. Por ejemplo, el panel Ideas Simples Conecta hace referencia al uso de instrumentación con la asignación de tratamiento utilizando solo Ideas Simples Conecta y los grupos de control en la estimación, excluyendo los otros grupos de tratamiento. La variable de tiempo en la aplicación se expresa en términos de horas. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. En el Apéndice se muestran los resultados de la estimación de la OLS endógena y de la primera etapa (véanse las tablas B.13, B.14 y B.15). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Análisis de desbordamiento

En esta sección, estudiamos si el tratamiento tuvo algún efecto indirecto en las secciones que no recibieron el tratamiento, es decir, un efecto en las secciones no tratadas de las escuelas tratadas. Se compararon las secciones no tratadas en las escuelas tratadas con todas las secciones en las escuelas del grupo control (Tabla 33). En general, los resultados sugieren que no hay efectos secundarios (positivos o negativos).

Tabla 33: Efecto de desbordamiento en los puntajes de la prueba SIMCE 2022

	Puntaje Promedio Matemáticas	Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje	Puntaje Promedio Lenguaje
	(1)	(2)	(3)	(4)
Clase de Control en Escuela Tratada	0.008	-0.002	0.040	0.056
	(0.072)	(0.049)	(0.055)	(0.039)
N	6,639	6,639	6,361	6,361
Media grupo de referencia	-0.05	-0.05	-0.07	-0.07
Controles FE: estratos	No X	Sí X	No X	Sí X
Errores agrupados por: rbd	X	X	X	X
Nivel de Regresión	Estudiante	Estudiante	Estudiante	Estudiante

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Estas estimaciones provienen de los datos de las secciones no tratadas en las escuelas tratadas y de todas las secciones en las escuelas de control. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Análisis de robustez

En esta sección, estudiamos la solidez de estos resultados para controlar las puntuaciones rezagadas de las pruebas a nivel individual. Específicamente, enviamos a la Agencia de la Calidad de la Educación una base de datos compuesta por el DNI del estudiante (RUT), el carné escolar (RBD) y el decil de la distribución de puntajes APTUS al que pertenecen. El total de estudiantes en esta base de datos tiene 9.508 estudiantes de las 235 escuelas del experimento.

A continuación, estimamos los efectos del tratamiento siguiendo la ecuación 2, pero también controlando el decil al inicio del estudio. Los resultados son similares a los de la muestra principal (Tabla 34).

Tabla 34: Efectos del tratamiento en los puntajes de la prueba SIMCE 2022 utilizando la submuestra emparejada de estudiantes

		Puntaje Promedio Matemáticas	Puntaje Promedio Matemáticas	Puntaje Promedio Lenguaje	Puntaje Promedio Lenguaje
		(1)	(2)	(3)	(4)
Conecta Ideas Simple		0.003 (0.053)	0.000 (0.035)	0.064 (0.058)	0.065 (0.045)
Incentivo Individual		-0.035 (0.050)	-0.008 (0.039)	-0.099* (0.050)	-0.074* (0.038)
Incentivo Grupal		-0.025 (0.061)	-0.017 (0.047)	0.002 (0.051)	0.006 (0.043)
N		5,439	5,439	5,233	5,233
Promedio grupo ref.		-0.05	-0.05	-0.07	-0.07
Dif. Conecta Ideas vs Indiv.		0.038 (0.47)	0.008 (0.84)	0.163 (0.47)	0.139 (0.84)
Dif. Conecta Ideas vs Grupal		0.028 (0.66)	0.017 (0.71)	0.063 (0.35)	0.059 (0.28)
Dif. Indiv. vs Grupal		0.01 (0.86)	0.009 (0.85)	0.1 (0.09)	0.08 (0.11)
Cualquier Tratamiento		-0.019 (0.66)	-0.008 (0.8)	-0.014 (0.73)	-0.003 (0.92)
Controles	No	Sí	No	Sí	
FE: estratos	X	X	X	X	
Errores Agrupados por: rbd	X	X	X	X	
Nivel Reg.:	de Estudiante Emparejado	Estudiante Emparejado	Estudiante Emparejado	Estudiante Emparejado	Estudiante Emparejado

Notas: Los puntajes de las pruebas SIMCE están estandarizados para todos los estudiantes. Las estimaciones se realizan con el universo de estudiantes emparejados que nos dio la Agencia de la Calidad de la Educación. Controlamos los siguientes aspectos: puntajes promedio de la prueba estandarizada de matemáticas y lenguaje de la escuela en la prueba SIMCE 2018, el puntaje de la escuela de la Teoría de Respuesta al Ítem de la prueba APTUS, el sexo y la edad del estudiante, el logaritmo del número de estudiantes que tuvo la escuela en la prueba SIMCE 2018 y las características de la escuela como la ruralidad, tipo de escuela (municipal o subvencionada) y bajo nivel socioeconómico. También se considera el diseño de aleatorización con efectos fijos de estratos. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Resultados de la evaluación de procesos

En esta sección se resumen algunos de los principales hallazgos de la evaluación del proceso, en particular aquellos que pueden explicar el efecto nulo de Conecta Ideas en los puntajes de las pruebas de matemáticas y los efectos negativos en los puntajes de lenguaje. Consulte el informe de evaluación del proceso para obtener una descripción más detallada de los hallazgos.

Contexto

Para entender los resultados de la evaluación de impacto, primero consideramos el contexto: ¿Qué elementos ambientales podrían influir en la implementación? Los colegios participantes son colegios públicos y privados concertados, todos ellos a 40 minutos de distancia de una estación de metro. Entre ellos, 235 establecimientos educacionales optaron por participar y ahora se encuentran en el grupo de tratamiento o control. De estos, el 53% son públicos, mientras que el 47% son colegios privados subvencionados. Esta distribución difiere de la división general de la Región Metropolitana de 28% de escuelas públicas y 72% privadas subvencionadas, lo que indica una mayor proporción de escuelas públicas en nuestro grupo de participantes (ver Figura 2 para un mapa de las escuelas y Tabla 2 que compara las escuelas en los experimentos con otras en el área).

Pandemia de COVID-19

Un factor clave para entender los resultados de la evaluación de impacto es la pandemia de COVID-19. Los estudiantes no pudieron asistir a clases presenciales durante casi dos años (la mayor parte de 2020 y 2021). El cierre de las escuelas afecta al desarrollo de los estudiantes al menos en dos aspectos clave: las interacciones sociales y el desarrollo, y su rendimiento académico y niveles de aprendizaje.

Varios maestros y directores informaron que la ausencia de las aulas físicas resultó en un aumento de la violencia en las interacciones escolares cuando se reanudaron las clases.

"[...] La pandemia tuvo un impacto significativo, ya que muchos niños llegaron con cero autonomías. Teníamos niños en 4° de primaria que no sabían usar cuadernos y otros elementos básicos. También hubo muchos problemas de violencia en las interacciones sociales, con muchos conflictos en los primeros meses. Adaptarse a la rutina y a la relación con la familia también fue un reto. Fue un proceso súper complejo". (Profesora con curso de IC #1, 2022).

En el primer semestre de 2022, muchos estudiantes tuvieron dificultades para asimilar los contenidos de 4° grado. Varios carecían de acceso a una computadora, a una conexión estable a Internet o tenían conocimientos tecnológicos limitados. El profesorado afirma que, a menudo, los padres completan los deberes de los alumnos.

"[...] Varias brechas son visibles; Parece que les cuesta interiorizar el contenido. En general, todas las clases presentan una resistencia al aprendizaje. Vemos que algo pasó en sus cerebros, y interiorizar el conocimiento y ponerlo en práctica es difícil para ellos". (Profesora con curso de CI #2, 2022)

Los maestros notaron un aumento significativo en la proporción de niños que no sabían leer. Esto retrasó la asimilación de nuevos contenidos.

Sin embargo, directivos y docentes afirmaron que el regreso a clases presenciales, a pesar de presentar dificultades, fue beneficioso para los estudiantes. El regreso a clases permitió a los estudiantes rezagados ponerse al día, mejorar las habilidades sociales y formar hábitos de estudio.

Adicionalmente, los docentes relataron que, durante la pandemia y las etapas de retorno a la presencialidad, surgieron retos que eran desconocidos para su quehacer pedagógico, además de haber experimentado un alto nivel de estrés que los afectó emocionalmente:

"[...] ... La parte emocional y psicológica, porque uno estaba traumatizado, porque primero fue una lucha para que los niños se unieran a estas clases en línea, que los niños prestaron atención...". (Profesora con curso de CI #3, 2022)

En este contexto, la implementación del programa Conecta Ideas no fue tan engorrosa para el equipo docente porque tuvieron que aprender a usar la tecnología durante la pandemia. Aunque no todos los docentes continuaron utilizando las herramientas TIC para impartir sus clases una vez finalizado el cierre de las escuelas, sí adquirieron las nociones básicas de cómo hacerlo.

En resumen, las principales consecuencias de la pandemia de COVID-19 fueron el aumento de estudiantes sin saber leer, la pérdida de hábitos de estudio como la puntualidad o la asistencia regular a clases, pero los docentes estaban más familiarizados con el uso de computadoras y otros programas educativos en sus clases.

Fidelidad

A la hora de realizar una intervención experimental, es importante tener en cuenta que la implementación en la mayoría de los casos difiere del diseño estipulado. Los incidentes imprevistos, la asistencia de los estudiantes y otras circunstancias pueden, en última instancia, disminuir el efecto de la intervención. En esta sección, discutiremos las instancias que impidieron que la implementación del tratamiento fuera ideal. También anotaremos si la ejecución superó las expectativas.

Retraso en la contratación de coordinadores de Conecta Idea

La contratación de los coordinadores del laboratorio de Conecta Ideas se llevó a cabo entre marzo y junio de 2022. El proceso de contratación se llevó a cabo muy cerca del inicio de la implementación.¹⁶ Esto comprimió el cronograma que afectó la implementación al principio. Por ejemplo, dos de ellos fueron contratados una semana antes de que comenzara la implementación, ya que son de nacionalidad extranjera, y llegaron a trabajar en el programa esa misma semana. En otro caso, se contrató a dos coordinadores de laboratorio para reemplazar a otros coordinadores cuando el programa ya se había implementado durante un par de semanas en dos escuelas diferentes. Uno renunció porque no le gustaba la dinámica de trabajo: la escuela a la que tenía que ir a trabajar estaba en una zona "peligrosa" y el trabajo le resultaba demasiado exigente. El otro fue despedido por falta de compromiso: llegó tarde a las clases en más de una ocasión, y en otras, no llegó del todo.

Retraso por implementación Aptus Test

La aplicación de la prueba Aptus como medición de línea base no estaba considerada en el cronograma inicial. La decisión de incorporarla, producto de las interrogantes sobre la aplicación de SIMCE que había en ese momento, generó un retraso en la implementación. Esto produjo molestia y tensiones con las escuelas, y disminuyó el entusiasmo original por el programa CI.

¹⁶ Con el fin de tener una fuente adicional de control de los resultados se construyó una línea de base mediante la aplicación de un test de Aptus de desempeño en matemáticas

Limitaciones de la infraestructura

Un elemento esencial para el funcionamiento del programa es la disponibilidad de computadoras operativas (o tabletas) con una conexión estable a Internet para cada estudiante. Las encuestas de evaluación de procesos revelaron que muchas escuelas carecían de la infraestructura adecuada.

De acuerdo con los coordinadores del laboratorio, varios establecimientos no cumplieron con el número requerido de computadoras por estudiante. A veces, las sesiones se llevaban a cabo en un aula normal con tabletas proporcionadas por Conecta Ideas, la escuela o los estudiantes. En otros casos, las computadoras fueron compartidas por dos estudiantes. Según los coordinadores del laboratorio, en el 35% de las escuelas, los estudiantes tuvieron que compartir dispositivos en alguna medida. Además, 16 establecimientos (12% de las escuelas tratadas) tuvieron que compartir equipos la mayor parte del tiempo.

El hecho de que los estudiantes tuvieran que compartir equipos de cómputo generó diferentes problemas. Por ejemplo, las banderas, los incentivos simbólicos que proporciona la plataforma por cada ejercicio realizado (y que se utilizaron como base para los sorteos en los tratamientos de incentivos individuales y grupales), solo fueron acumuladas por el estudiante que inició sesión en las computadoras, incluso si compartía la computadora con un compañero de clase. Esto generó frustración entre los estudiantes cuyos usuarios no habían iniciado sesión, pero habían participado en la sesión de Conecta Ideas. También disminuyó el tiempo que cada estudiante tenía para realizar ejercicios individuales. Además, las preguntas de la plataforma Conecta Ideas no fueron diseñadas como ejercicios grupales.

Otro problema era la mala calidad de la conexión a Internet en muchas escuelas. Según los coordinadores del laboratorio, el 13% de las escuelas tuvieron problemas de internet la mayor parte del tiempo. Maestros y directores de escuelas afirmaron que cuando hubo un corte de energía o se cortó el internet, la sesión de Conecta Ideas no se llevó a cabo.

Formación del profesorado

En el diseño inicial, se ofrecería capacitación a los docentes de las escuelas que recibieron el programa Conecta Ideas sobre el uso de la plataforma. Sin embargo, a los coordinadores no se les asignó ninguna responsabilidad en la capacitación de los maestros. De hecho, el 73% de los docentes indicaron en la encuesta que no recibieron ninguna formación para aprender a utilizar la plataforma; De ellos, el 80% declaró que les hubiera gustado que se les ofreciera uno. Las entrevistas con los docentes y los coordinadores de laboratorio dejaron claro que era necesario tener más claridad sobre este asunto y que debía haber una formación formal para los docentes. La falta de capacitación dificultó la implementación del programa. Por ejemplo, si los profesores entendieran mejor la plataforma, podrían diseñar las clases regulares de matemáticas para que fueran complementarias a la sesión de Ideas Conecta o trabajar con el coordinador del laboratorio sobre qué tipo de ejercicios beneficiarían más a los estudiantes.

Coordinación entre los profesores y el personal de Conecta Ideas

El hecho de que los maestros y el laboratorio coordinados estén alineados en sus objetivos es clave para el éxito del programa. La comunicación entre docentes y coordinadores fue permanente y fluida. En general, se coordinaron para decidir los objetivos de aprendizaje que se abordarían en la próxima sesión de Conecta Ideas y el nivel de dificultad de los ejercicios. También colaboraron en la gestión de la dinámica del aula. Los coordinadores ayudaron al profesor a ver el progreso de los estudiantes utilizando datos de la plataforma. Esta dinámica positiva entre docentes y coordinadores fomentó un ambiente propicio para Conecta Ideas y una implementación armoniosa.

Reajustes en el tiempo designado para las sesiones de laboratorio

A principios de 2022, se autorizó a las escuelas a reducir el tiempo de clase de 90 a 60 minutos para varias asignaturas. Esto se debió a la incertidumbre generada por la pandemia de COVID-19. Por lo tanto, el plan de implementación original se basó en clases de matemáticas que duraban 60 minutos. Sin embargo, en julio de 2022, el gobierno solicitó a las escuelas que volvieran al horario tradicional de 90 minutos. Este reajuste obligó a una reestructuración total del plan de implementación (es decir, asignación de coordinadores de laboratorio

a las escuelas). Sin embargo, AutoMind pudo reestructurar el horario de los 12 coordinadores de laboratorio para cubrir las 137 escuelas tratadas.

Duración de la sesión

Como se mencionó anteriormente, la duración de las sesiones de Conecta Idea finalmente se ajustó a 90 minutos. Sin embargo, fue difícil usar los 90 minutos de manera efectiva. Por ejemplo, muchos estudiantes llegaron tarde a las clases de las 8:00 a.m. Los estudiantes también tenían que ir de su aula habitual al laboratorio de informática, lo que les llevaba entre 10 y 15 minutos y, en algunos casos, hasta 30 minutos. En general, las sesiones de Conecta Ideas duraron entre 45 y 75 minutos en la práctica. Por lo tanto, el tiempo efectivo en la plataforma fue mucho menor de lo planeado. Además, a veces los maestros de la escuela se tomaban de 15 a 20 minutos al comienzo de las clases para dar una explicación tradicional sobre lo que trabajarían los estudiantes.

Plataforma fácil de usar

El uso o no de la plataforma depende de lo complejo que sea su uso. Una herramienta educativa compleja para los estudiantes genera molestia, mientras que una herramienta lo suficientemente amigable genera interés. Cuando se les preguntó sobre la facilidad de uso de Conecta Ideas, el 93% de los estudiantes respondieron que les resultó fácil de navegar y usar. En general, los estudiantes entendieron rápidamente la lógica detrás del software, supieron cómo navegar por él, ingresaron actividades y ejercicios en cada sesión y siguieron el orden lógico para resolverlos. Entendieron las instrucciones que dio la plataforma sin mayores problemas y sabían que si resolvían correctamente el ejercicio, recibirían banderas como recompensa.

Dificultad del ejercicio

Una promesa de Conecta Ideas es la capacidad de "enseñar al nivel adecuado" (es decir, atendiendo a las necesidades de los estudiantes y a los niveles reales de aprendizaje). Al menos, tal y como perciben profesores y alumnos, un acierto de Conecta Ideas está en la dificultad de los ejercicios. Los ejercicios se adaptaron a las necesidades educativas de los estudiantes. Por ejemplo, el 69% de los docentes estuvo de acuerdo en que "los ejercicios de la plataforma Conecta Ideas son adecuados para el nivel de aprendizaje de los estudiantes". Por otro lado, desde la perspectiva de los estudiantes, los ejercicios fueron accesibles. El 80% de los estudiantes encuestados dijo que los ejercicios fueron fáciles, mientras que el 20% restante dijo que eran difíciles de entender.

Sin embargo, los ejercicios no se adaptaron a los contextos de las escuelas (sobre todo teniendo en cuenta las posibles pérdidas de aprendizaje durante la pandemia, lo que hizo que los estudiantes se quedaran rezagados en comparación con sus compañeros en años anteriores), lo que provocó que los estudiantes menos privilegiados se sintieran marginados.

Aplazamiento en las solicitudes de incentivos

El cronograma en el que se anunciaron y otorgaron los incentivos no fue el previsto. La idea original era que los incentivos se anunciaran a finales de julio de 2022. Sin embargo, debido a los diferentes retrasos, se anunciaron en septiembre. Además, la idea era hacer tres rondas de incentivos. Sin embargo, solo se llevaron a cabo dos rondas de incentivos debido al tiempo limitado entre la prueba SIMCE 2022 realizada en la tercera semana de noviembre y la finalización de cursos en la primera quincena de diciembre.

Por otro lado, los incentivos deben motivar a los estudiantes para que trabajen. El premio que se entregó en el sorteo fue un cómic para colorear que explica qué es STEM. Los coordinadores del laboratorio coincidieron en que, al principio, los estudiantes estaban muy entusiasmados con los incentivos, y estos generaron muchas expectativas. Sin embargo, la sensación general era que el cómic para colorear no era lo que los estudiantes esperaban y se sintieron decepcionados. Además, la entrega de incentivos se retrasó más de una semana, lo que aumentó el sentimiento de decepción entre los estudiantes y los desinteresó por los incentivos.

Capacidad de los coordinadores de Conecta Ideas

Si bien los coordinadores de laboratorio estaban adecuadamente preparados en términos de conocimiento de los contenidos, por ejemplo, el 80% de los docentes estuvo totalmente de acuerdo y el 18% estuvo de acuerdo

con la afirmación de que "el coordinador del laboratorio de Conecta Ideas ha podido tener un buen dominio de la materia que necesitaba ser enseñada". Sin embargo, una cuestión importante era simplemente que había demasiados estudiantes para que pudieran abordar plenamente todas las dudas e inquietudes de los estudiantes que surgieron durante las sesiones.

Sobrecarga de trabajo para los coordinadores

Otro problema en el papel de los coordinadores de laboratorio era su gran carga de trabajo. Cuatro coordinadores de laboratorio (50% de los entrevistados) señalaron que tener que planificar, asistir e implementar clases en 11 o 12 escuelas en una semana es una gran carga de trabajo, además de las actividades administrativas que se requieren, que muchas veces surgen de manera espontánea y con urgencia. También señalaron que la cantidad de escuelas que visitaron no permitía la personalización de los estudiantes, ya que es muy complejo aprender los nombres de todos, su nivel de aprendizaje y su contexto social, una situación que los maestros notaron y resintieron.

Práctica común

En esta sección, discutiremos los comportamientos y hábitos comunes observados en las escuelas y entre maestros y estudiantes durante el período de implementación.

En resumen, la intervención alteró el statu quo en las escuelas: los profesores experimentaron un aumento de la carga de trabajo y los estudiantes mostraron una mayor motivación en las clases de matemáticas. Sin embargo, cuestiones más complejas, como la violencia y la indisciplina escolar, pueden afectar la eficacia de la implementación.¹⁷ Estas cuestiones estaban previstas y quedaban fuera del alcance de la intervención. En particular, el aumento de la motivación de los estudiantes se destaca como un éxito significativo, lo que destaca el atractivo de la plataforma. Si bien no hubo problemas críticos que pusieran en peligro el efecto de la intervención, ciertos aspectos, en particular la prueba de práctica SIMCE, requirieron una consideración cuidadosa debido a su potencial para sesgar los resultados.

Vulnerabilidad de las escuelas e inconvenientes extracurriculares

La vulnerabilidad de las escuelas durante la implementación se hizo evidente frente a un contexto violento que repercutió en las escuelas de dos maneras. En primer lugar, el acoso escolar es frecuente en las escuelas del experimento y también fue común durante las sesiones de Conecta Ideas.¹⁸ Esto era especialmente común cuando los estudiantes tenían que trabajar juntos cuando no había suficientes computadoras para cada uno. Un coordinador del laboratorio señaló que la pasividad del profesor, sumado a la falta de reglas, generó un ambiente hostil entre los estudiantes, lo que dificultó el desarrollo de la sesión de Ideas Conecta.

"[...] Además, en esa escuela [...], ha habido muchos problemas de violencia y acoso escolares, y tiene mucho sentido para mí porque los niños no tienen reglas y se tratan muy mal. Es complejo llevar a cabo la sesión en esa escuela, y ella [la profesora] ha tenido esa actitud pasiva desde el principio. Tiene mucho sentido para mí, cuánto te involucras y cuánto impactará esto a los niños". (Coordinadora Femenina de Ideas Conecta #1, 2022)

En segundo lugar, hubo muchos eventos que afectaron el funcionamiento normal de las escuelas. Los días festivos, las huelgas y las prácticas del SIMCE provocaron la pérdida de varias sesiones de taller. El robo de materiales tecnológicos, como computadoras, implicó que los estudiantes no pudieran recibir la implementación del taller Conecta Ideas en más de una ocasión.

Otro factor relevante para el aprendizaje del alumnado destacado por el equipo docente es la implicación de las familias. Sin embargo, dado que las escuelas del experimento están ubicadas en contextos socialmente vulnerables, los estudiantes reciben poco apoyo en casa.

¹⁷ Si bien estos problemas no afectan la validez interna de nuestras estimaciones, el programa puede tener un mayor impacto cuando se implementa en entornos sin estos problemas. Nuestro experimento es incapaz de responder a esa pregunta.

¹⁸ Si bien no tenemos datos al respecto de las escuelas de control, en virtud de la asignación aleatoria, el acoso escolar debería tener la misma prevalencia en las escuelas de tratamiento y control.

Condiciones de trabajo de los docentes

Hay una alta rotación de maestros en algunos establecimientos educativos. Esto afecta el proceso de aprendizaje de los estudiantes e interrumpe la implementación de Conecta Ideas.

La alta carga de trabajo que tenían los profesores y el estrés y agotamiento que experimentaban son factores que son relevantes a la hora de su participación en las sesiones de Conecta Ideas. Algunos docentes señalaron que el programa se presentaba como una tarea extra para ellos.

Motivación de los estudiantes

A pesar de que se tuvieron que recuperar las reglas de convivencia entre los estudiantes debido al regreso a clases presenciales, la implementación despertó dinámicas positivas de comportamiento. Por ejemplo, los equipos directivos y docentes señalaron que muchos de los estudiantes tuvieron una excelente acogida a la plataforma Conecta Ideas y mostraron motivación respecto a su uso. Señalaron que la plataforma era útil para las clases y su aprendizaje. Además, varias historias de profesores y directivos describen el aumento de la motivación para aprender matemáticas debido a los ejercicios de la plataforma Conecta Ideas. Un profesor señaló que "sí, ha habido un aumento de la motivación; Están súper motivados con las matemáticas". En general, el personal docente está de acuerdo en que la plataforma se convirtió en una forma entretenida para que los estudiantes aprendieran matemáticas.

Intervenciones externas

En cuanto a la ocurrencia de intervenciones formales distintas a la implementación de Conecta Ideas, no existieron como tales. Sin embargo, el equipo de AutoMind realizó algunas actividades adicionales que no estaban contempladas en el diseño original de la intervención. Estas intervenciones tenían el potencial de sesgar los resultados observados.

Actividad de títeres

La primera intervención no planificada consistió en una actividad dirigida a desarrollar la voz interior de los niños para mejorar su capacidad de articular su razonamiento, particularmente en preguntas metacognitivas, retroalimentación de compañeros y mientras actuaban como monitores. Esta dinámica incluyó proporcionar a los estudiantes títeres e involucrarlos en actividades de canto. Sin embargo, como esta actividad se desvió de la teoría del cambio establecida y del diseño estandarizado del programa, afectó la interpretación de los resultados de la evaluación de impacto. Por ejemplo, si hubiéramos encontrado un efecto positivo en el tratamiento, no habría estado claro si fue impulsado por la actividad de los títeres o por Conecta Ideas. En consecuencia, esta actividad se detuvo y solo se implementó en escuelas seleccionadas (teníamos conocimiento de al menos 6 escuelas).

Comunicación con los padres y tutores

AutoMind implementó varias estrategias de comunicación para interactuar con los padres y tutores de los estudiantes. Estas estrategias tenían dos propósitos: en primer lugar, permitían a las escuelas informar a las familias sobre la implementación de programas educativos innovadores, fomentando conexiones más sólidas entre las familias y las escuelas y mejorando el compromiso de la escuela con el programa. En segundo lugar, la comunicación tenía como objetivo fomentar el apoyo en el hogar para la práctica de los estudiantes. Sin embargo, debido a la gran cantidad de escuelas participantes, la comunicación constante con los tutores solo se realizó en algunas escuelas.¹⁹

Prueba práctica SIMCE 2022

La última actividad no incluida en la teoría del cambio, que potencialmente sesgó los resultados, fue que AutoMind realizó una prueba de práctica SIMCE en escuelas de tratamiento. Esta prueba, alineada con los objetivos de aprendizaje de la prueba estandarizada 2022, fue administrada por el equipo implementador durante las sesiones de Conecta Ideas entre la última semana de octubre y la primera semana de noviembre. El equipo evaluador identificó que la incorporación de esta prueba práctica podría introducir sesgos en el estudio (lo que llevaría a exagerar el efecto de Conecta Ideas), ya que se relaciona directamente con la medida

¹⁹ No tenemos datos sobre qué escuelas tuvo lugar y en cuáles no.

de resultados en la evaluación de impacto. En consecuencia, su aplicación se detuvo a mitad del proceso. Las pruebas prácticas del SIMCE se realizaron en el 64% de las escuelas participantes en el programa.

Si hubiéramos encontrado un efecto positivo de Conecta Ideas, habríamos intentado analizarlo sin las escuelas donde se realizó esta prueba de práctica. Sin embargo, esto nos habría dejado con un pequeño número de escuelas de tratamiento y estimaciones ruidosas. Dada la falta de efectos positivos del programa (y que la prueba de práctica probablemente lleva a exagerar los efectos de Conecta Ideas), decidimos no profundizar en este tema.

Conecta Ideas en la práctica

Para dar más información sobre la implementación de Conecta Ideas, utilizamos datos recopilados directamente por el software relacionados con el uso. Para cada sesión de Conecta Ideas, conocemos al estudiante que se conectó, las preguntas intentadas, la respuesta a estas preguntas y otras características de cada sesión.

Tenemos un total de 4.434 alumnos de 231 colegios en la base de datos de Conecta Ideas. Al final del año escolar, los estudiantes habían respondido un total de 2.848.689 preguntas.²⁰

La mediana de los estudiantes responde entre 500 y 600 preguntas en total (durante la duración del tratamiento), y cada pregunta dura entre 0,82 y 0,85 minutos en promedio (Tabla 35). Pasando a los agregados a nivel escolar, la escuela media tiene 33 estudiantes, que responden a un total de 17.000 a 19.000 preguntas (Tabla 36).

En general, los estudiantes de las escuelas tratadas pasaron un tiempo promedio de 17,16 minutos por semana en la plataforma y realizaron 22 ejercicios semanales. Esto incluye semanas en las que los estudiantes no se conectaron. Solo si nos centramos en las semanas en las que los alumnos se conectaron, obtenemos los siguientes resultados: 38 minutos a la semana en la plataforma y 50 ejercicios a la semana. Esto equivale a aproximadamente 642 ejercicios en el transcurso del experimento y 8,5 horas pasadas en la plataforma. Estos números sugieren que los estudiantes están perdiendo efectivamente tiempo de instrucción de matemáticas debido al tratamiento, ya que las sesiones de Conecta Ideas que se llevan a cabo durante una clase regular de matemáticas están destinadas a durar entre 60 y 90 minutos.

Tabla 35: Estadísticas descriptivas a nivel de estudiante del uso de Conecta Ideas

Variable	Simple, N = 1,492	Incentivo Individual, N = 1,449	Incentivo Grupal, N = 1,493
Tiempo de Respuesta Promedio			
Mediana (IQR)	0.85 (0.66 - 1.15)	0.82 (0.64 - 1.09)	0.83 (0.65 - 1.09)
Media (SD)	0.96 (0.49)	0.90 (0.39)	0.94 (0.47)
Rango	0.02 - 4.92	0.00 - 4.19	0.18 - 5.38
Nota Promedio			
Mediana (IQR)	6.21 (5.90 - 6.45)	6.21 (5.90 - 6.43)	6.23 (5.96 - 6.47)
Media (SD)	6.11 (0.49)	6.13 (0.45)	6.16 (0.44)
Rango	2.50 - 6.88	2.11 - 7.00	1.50 - 7.00

²⁰ El número de escuelas difiere de las 235 escuelas iniciales a las que se les asignó un tratamiento porque cuatro escuelas se negaron a aceptar el tratamiento después de que fueron asignadas al azar a él. Consulte la sección [3.1.7.3](#) para obtener más información.

En la Escuela			
Mediana (IQR)	1.00 (0.98 - 1.00)	1.00 (0.94 - 1.00)	1.00 (0.97 - 1.00)
Media (SD)	0.96 (0.12)	0.92 (0.18)	0.93 (0.19)
Rango	0.00 - 1.00	0.00 - 1.00	0.00 - 1.00
Número de Respuestas			
Mediana (IQR)	500 (294 - 765)	618 (386 - 934)	608 (368 - 898)
Media (SD)	568 (378)	692 (426)	669 (427)
Rango	1 - 2721	1 - 3064	1 - 2917

Notas: Esta tabla presenta estadísticas resumidas de la base de datos de uso de Conecta Ideas. Cada observación es un estudiante. El primer panel presenta información sobre el tiempo promedio de respuesta a las preguntas, el segundo panel tiene información sobre la calificación promedio (que está entre 1 y 7), el tercer panel es la proporción de preguntas intentadas durante el horario escolar y el último panel tiene el número total de preguntas intentadas por el estudiante.

Tabla 36: Estadísticas descriptivas a nivel de escuela del uso de Conecta Ideas

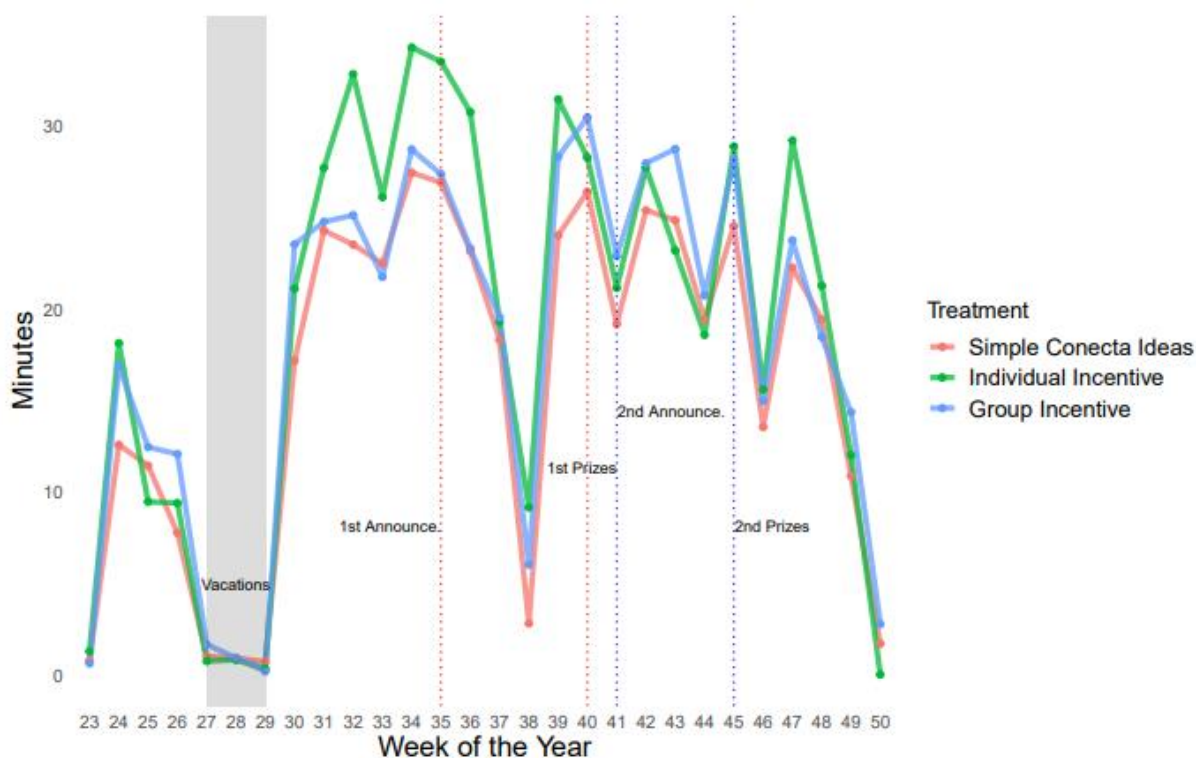
Variable	Simple, N = 46	Incentivo Individual, N = 45	Incentivo Grupal, N = 46
Tiempo Promedio de Respuesta			
Mediana (IQR)	0.81 (0.71 - 0.95)	0.79 (0.67 - 0.92)	0.80 (0.67 - 0.92)
Media (SD)	0.85 (0.27)	0.81 (0.21)	0.86 (0.29)
Rango	0.21 - 1.76	0.34 - 1.30	0.47 - 1.83
Nota Promedio			
Mediana (IQR)	6.24 (6.04 - 6.34)	6.20 (6.08 - 6.32)	6.26 (6.11 - 6.36)
Media (SD)	6.18 (0.23)	6.19 (0.20)	6.23 (0.19)
Rango	5.31 - 6.54	5.33 - 6.52	5.60 - 6.54
En la Escuela			
Mediana (IQR)	0.97 (0.93 - 0.99)	0.97 (0.93 - 0.98)	0.97 (0.95 - 0.99)
Media (SD)	0.93 (0.13)	0.92 (0.16)	0.93 (0.16)
Rango	0.42 - 1.00	0.04 - 1.00	0.00 - 1.00
Número de Respuestas			
Mediana (IQR)	17,209 (11,296 - 24,177)	19,413 (14,118 - 27,872)	19,701 (14,274 - 28,074)
Media (SD)	18,414 (9,847)	22,287 (11,145)	21,712 (11,734)
Rango	2,574 - 48,949	2,109 - 55,237	2,072 - 54,961
Estudiantes			
Mediana (IQR)	33 (27 - 40)	32 (26 - 38)	33 (28 - 36)
Media (SD)	32 (10)	32 (7)	32 (8)
Rango	sept-48	20 - 46	16 - 47

Notas: Esta tabla presenta estadísticas resumidas de la base de datos de uso de Conecta Ideas. Cada observación es una escuela. El primer panel presenta información sobre el tiempo promedio de respuesta a las preguntas, el segundo panel tiene información sobre la calificación promedio (que está entre 1 y 7), el tercer

panel es la proporción de preguntas intentadas durante el horario escolar, y el cuarto panel tiene el número total de preguntas intentadas por el estudiante, y el último panel es el número de estudiantes.

A continuación, se estudia la evolución a lo largo del curso escolar en el tiempo de permanencia en la plataforma (Figura 15) y en el número de preguntas intentadas por semana (Figura 16). Las cifras muestran el período de vacaciones y otros eventos importantes (por ejemplo, cuándo se anuncian los incentivos y cuándo se otorgan a los grupos incentivados).²¹ Después del período de vacaciones, el uso de la plataforma aumentó en todos los grupos de tratamiento. Sin embargo, el grupo de incentivo individual pasó más tiempo en la plataforma e intentó hacer más preguntas. Esto no puede explicarse por los incentivos, ya que los estudiantes y los profesores no tenían ni idea de ellos en ese momento. Los tres grupos comenzaron a moverse juntos después de que se anunciaron los incentivos (en la semana 35), lo que sugiere que los incentivos grupales aumentaron el uso. Por lo tanto, parece que los incentivos individuales, en todo caso, tuvieron un efecto negativo en el uso, mientras que los incentivos grupales tuvieron un efecto positivo. La entrega de los premios y el anuncio de una nueva ronda de incentivos no tuvieron un efecto perceptible en el uso.

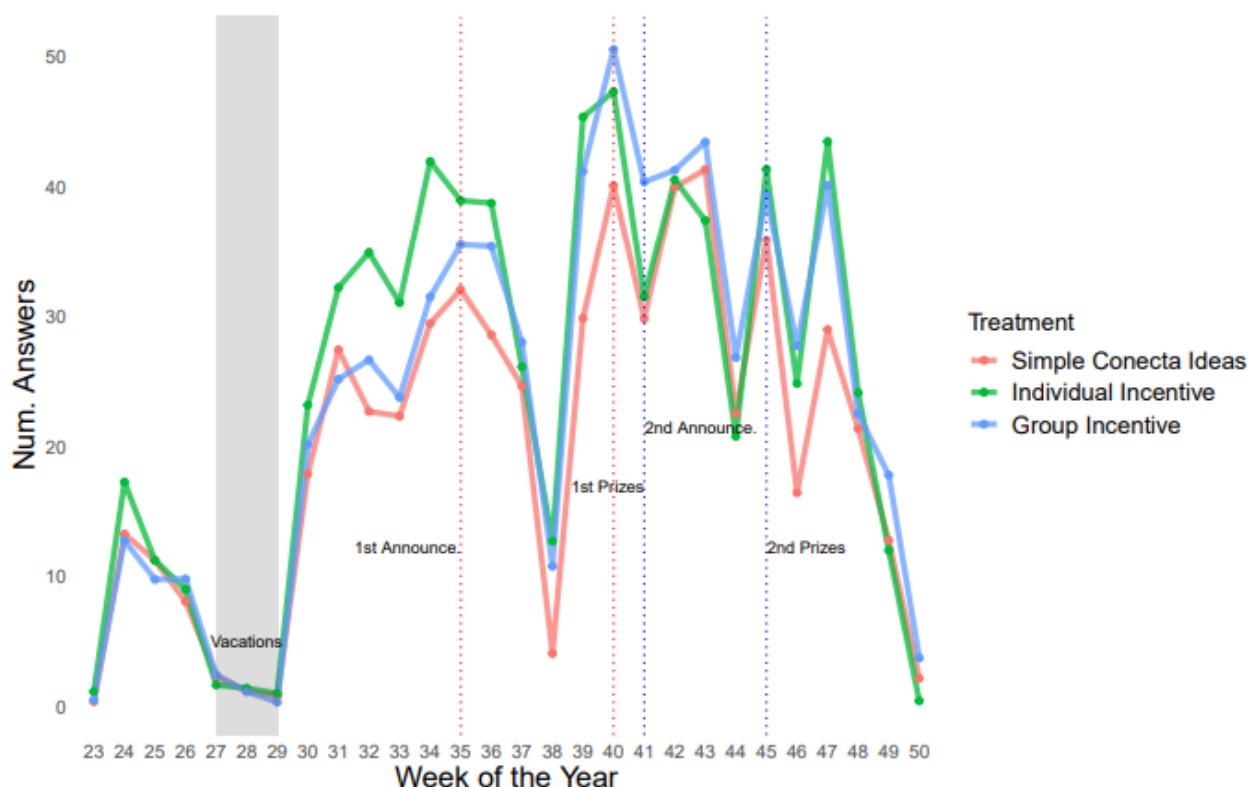
Figura 13: Tiempo de permanencia semanal en la plataforma Conecta Ideas



Notas: Esta figura muestra el tiempo promedio que cada estudiante pasa en la plataforma cada semana. El promedio se calcula para cada semana y entre todos los estudiantes de cada grupo de tratamiento. Las líneas punteadas muestran cuándo se anunciaron los incentivos a profesores y estudiantes y cuándo se distribuyeron los premios.

²¹ El uso es menor en las semanas 36, 37 y 38, que corresponden a las primeras semanas de septiembre, cuando hay varios días festivos nacionales que podrían explicar el menor uso en esas semanas. Por otro lado, las semanas 47 a 50 fueron las dos últimas semanas de noviembre y las dos primeras de diciembre. Eran tres feriados y el SIMCE ya estaba implementado (16 y 17 de noviembre), por lo que en esta época el trabajo en las escuelas se ralentiza y en algunos casos comienzan las vacaciones. Para esa fecha, ya se había realizado el segundo sorteo para sortear los premios.

Figura 14: Número de preguntas intentadas por semana en la plataforma Conecta Ideas



Notas: Esta figura muestra el número promedio de preguntas que cada estudiante intenta en la plataforma cada semana. El promedio se calcula para cada semana y entre todos los estudiantes de cada grupo de tratamiento. Las líneas punteadas muestran cuándo se anunciaron los incentivos a profesores y estudiantes y cuándo se distribuyeron los premios.

Para formalizar esta intuición, estudiamos el efecto de anunciar los incentivos en los grupos de tratamiento incentivados en relación con la intervención simple utilizando un marco de estudio de eventos. Específicamente, estimamos la siguiente ecuación a través de OLS:

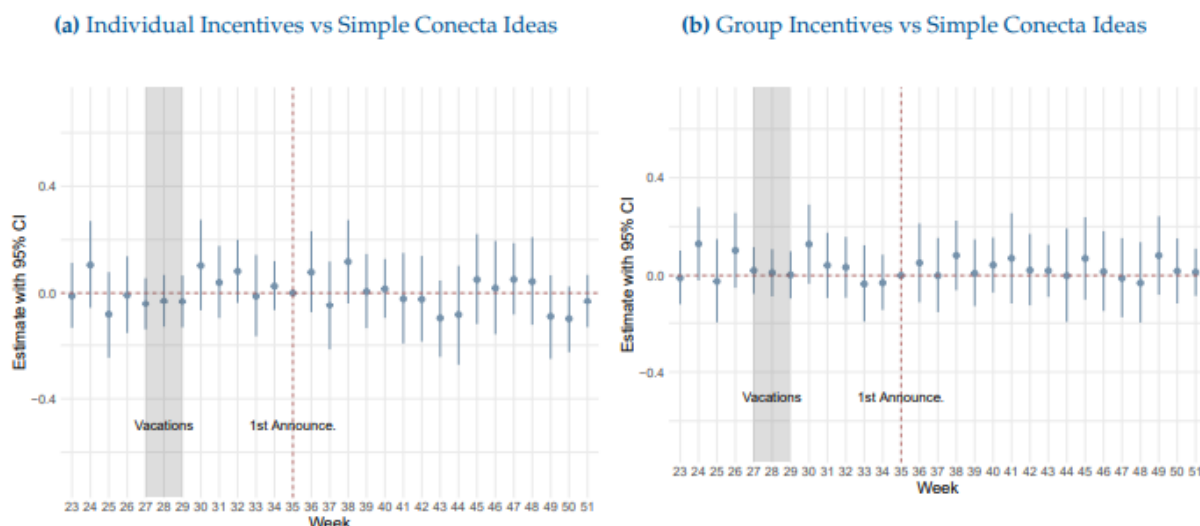
$$Y_{ist} = \alpha_s + \alpha_t + \sum_{\tau \neq 34} \beta_{\tau} I_{\tau} \times II_s + \sum_{\tau \neq 34} \lambda_{\tau} I_{\tau} \times GI_s + \varepsilon_{ist}$$

donde Y_{ist} es el resultado de interés para el estudiante i en las escuelas s en la semana t (por ejemplo, el número de preguntas respondidas o el tiempo que pasa en la plataforma), α_s y α_t son los efectos fijos de la escuela y la semana, I_{τ} es una variable dicotómica igual a uno si la semana t es igual a τ , II_s y GI_s son variables ficticias para los incentivos individuales y grupales, y ε_{ist} es el término de error. Agrupamos los errores estándar a nivel escolar, que es el nivel en el que se aleatorizó el tratamiento (Abadie et al., 2023). Los coeficientes de interés son β_{τ} y λ_{τ} que muestran la evolución del tratamiento en relación con la semana en que se anunció el tratamiento (ver Figuras 15-17)

En general, parece que el grupo de incentivos individuales estaba utilizando la plataforma relativamente más en las semanas previas al anuncio de los incentivos. Sin embargo, después de que se anunciaron los incentivos,

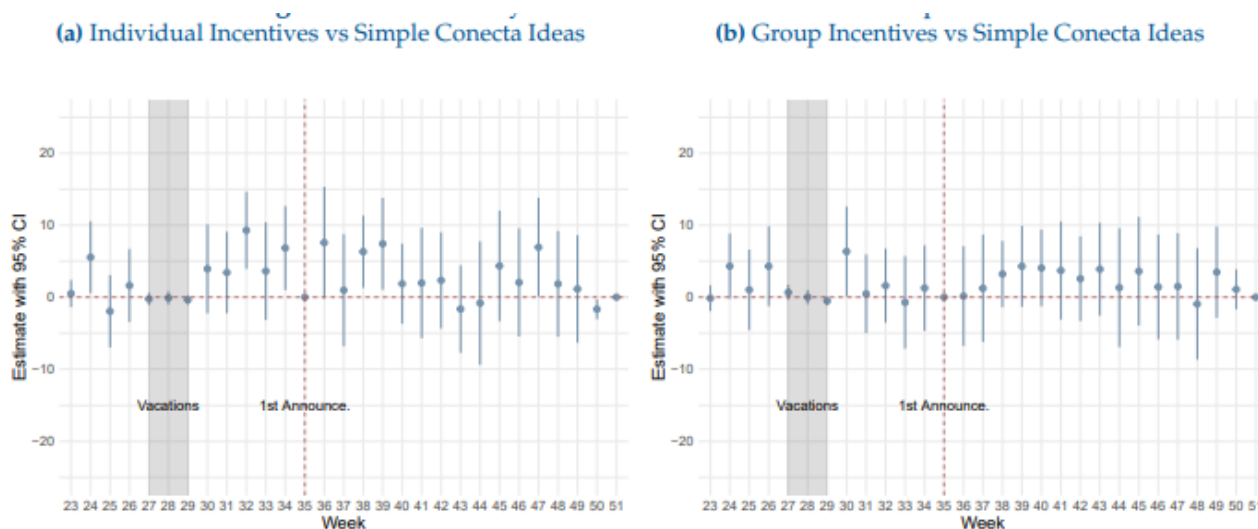
el tiempo dedicado a la plataforma y la cantidad de preguntas no cambiaron. Hubo un ligero aumento (aunque no estadísticamente significativo) en el uso después de que se anunciaron los incentivos para los incentivos grupales.

Figura 15: Estudio de eventos para la probabilidad de uso de la plataforma Conecta Ideas



Notas: Estas figuras presentan los gráficos de estudio de eventos siguiendo la Ecuación 4. El resultado es un maniquí que indica si el estudiante usó la plataforma al menos una vez durante la semana. El período de referencia es la semana 35, cuando se anunciaron los premios a los niños. El panel de la izquierda compara el grupo de tratamiento de incentivo individual con el grupo con la versión simple de Conecta Ideas. El panel de la derecha compara el grupo de tratamiento de incentivo individual con el grupo con la versión simple de Conecta Ideas. En ambos, se utiliza un panel equilibrado.

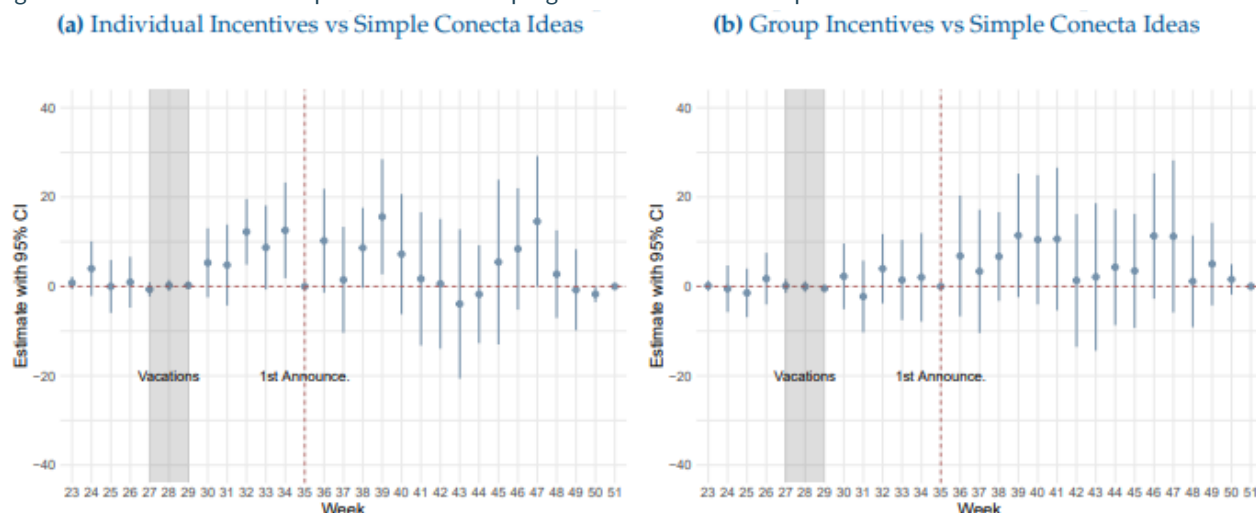
Figura 16: Estudio de eventos para el tiempo en la plataforma Conecta Ideas



Notas: Estas figuras presentan los gráficos de estudio de eventos siguiendo la Ecuación 4. El resultado es el número de preguntas que se intentan cada semana. El período de referencia es la semana 35, cuando se anunciaron los premios a los niños. El panel de la izquierda compara el grupo de tratamiento de incentivo individual con el grupo con la versión simple de Conecta Ideas. El panel de la derecha compara el grupo de

tratamiento de incentivo individual con el grupo con la versión simple de Conecta Ideas. En ambos, se utiliza un panel equilibrado.

Figura 17: Estudio de eventos para el número de preguntas intentadas en la plataforma Conecta Ideas



Notas: Estas figuras presentan los gráficos de estudio de eventos siguiendo la Ecuación 4. El resultado es el número de preguntas que se intentan cada semana. El período de referencia es la semana 35, cuando se anunciaron los premios a los niños. El panel de la izquierda compara el grupo de tratamiento de incentivo individual con el grupo con la versión simple de Conecta Ideas. El panel de la derecha compara el grupo de tratamiento de incentivo individual con el grupo con la versión simple de Conecta Ideas. En ambos, se utiliza un panel equilibrado.

Para resumir este efecto, estimamos el siguiente modelo:

$$Y_{ist} = \alpha_s + \alpha_t + \beta I_{t>35} \times II_s + \lambda I_{t>35} \times GI_s + \varepsilon_{ist}$$

donde todo está como antes, pero es un dummy que indica si los incentivos ya han sido anunciados (en la semana 35). Por lo tanto, y capturar el efecto de los incentivos en el uso de la plataforma. Los incentivos individuales no tuvieron ningún efecto (o un ligero efecto negativo) en el uso de la plataforma. Los incentivos grupales tuvieron un ligero efecto positivo en el uso. Dado el efecto mayoritariamente nulo sobre el uso de los incentivos, no es sorprendente que los efectos del tratamiento sobre el aprendizaje sean similares en todos los grupos de tratamiento.

Tabla 37: Cambio en los efectos del tratamiento después del primer anuncio de incentivos

	Al Menos Un Intento	Tiempo en la Aplicación	Respuestas por Semana
	(1)	(2)	(3)
Incentivo Individual post	x -0.018 (0.022)	-0.436 -1.073	-0.031 -2.543
Incentivo Grupal post	x 0.000 (0.025)	0.694 -1.254	4.872** -2.237

N	128,586	128,586	128,586
Media del grupo de referencia	0.43	15.65	19.58
Diferencia Indiv. Incentivo x post vs	0.018 (0.43)	1.131 (0.3)	4.903 (0.05)
Grupal Incentivo x post			
Nivel de Regresión	Semana- Estudiante	Semana- Estudiante	Semana- Estudiante
FE: semana del año	X	X	X
FE: rbd	X	X	X
Errores agrupados por:	rbd	rbd	rbd

Notas: El tratamiento Simple se utiliza como grupo de referencia en esta tabla. El tratamiento es interactuado por un maniquí que indica si la observación es pasada la semana 35, cuando se anunciaron los incentivos. El modelo se estima mediante un panel equilibrado. También se considera el diseño de aleatorización con efectos fijos de estratos y la temporalidad con efectos fijos de la semana del año. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tabla 38: Cambio en los efectos del tratamiento después del primer anuncio de incentivos (solo observaciones post-vacaciones)

	Al Menos Un Intento	Tiempo en la Aplicación	Respuestas por Semana
	(1)	(2)	(3)
Incentivo Individual x post	-0.047* (0.028)	-3.076** -1.480	-4.141 -3.131
Incentivo Grupal x post	0.005 (0.031)	0.589 -1.579	3.856 -2.690
N	97,548	97,548	97,548
Media del grupo de referencia	0.51	19.01	24.11

Dif.				
Incentivo				
Indiv. x				
post vs	0.052	3.666	7.997	
Incentivo	(0.08)	(0.02)	(0.01)	
Grupal x				
post				
Nivel de	Estudiante-	Estudiante-	Estudiante-	
Regresión	semana	semana	semana	
FE:				
semana	X	X	X	
del año				
FE: rbd	X	X	X	
Errores				
agrupados	X	X	X	
por: rbd				

Notas: El tratamiento Simple se utiliza como grupo de referencia en esta tabla. El tratamiento es interactuado por un maniquí que indica si la observación es pasada la semana 35, cuando se anunciaron los incentivos. El modelo se estima mediante un panel equilibrado. Las vacaciones ocurrieron durante la semana 27 y la semana 29. También se considera el diseño de aleatorización con efectos fijos de estratos y la temporalidad con efectos fijos de la semana del año. Los errores se agrupan por escuela. La diferencia de coeficientes se comprobó con una prueba de Chi-cuadrado. En el cuerpo principal de la tabla, los números entre paréntesis son errores estándar. El número entre paréntesis junto a las diferencias entre grupos y para el efecto de "cualquier tratamiento" son los valores * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Costos

Los costos del programa se pueden dividir en cinco grandes categorías: coordinadores de laboratorio, computación en la nube, soporte técnico y mantenimiento, y otros costos. Cada uno de ellos tiene un coste fijo y otro variable. A continuación, ampliamos cada una de estas categorías.

Coordinadores de laboratorio y otros costos de mano de obra

Un coordinador de laboratorio gana ~1.500 dólares al mes. Cada coordinador puede dirigir hasta 16 sesiones de Ideas Conecta mensualmente, cada una con 30 estudiantes y con una duración de Escriba aquí la ecuación. 90 minutos.²² Los coordinadores toman licencia médica por enfermedades relacionadas con el trabajo el 10% del tiempo (~20% en invierno). Estas ausencias a menudo se anuncian horas antes de una sesión. Por lo tanto, por cada cinco coordinadores, es necesario un sustituto a tiempo completo. Por lo tanto, el costo anual por estudiante para una sesión semanal de 90 minutos es de USD 43.5 (=USD 1,500 / mes * 12 meses * 1.16 (el costo del sustituto) / (16 sesiones * 30 estudiantes)).

Además, Conecta Ideas contrata a un profesor experto para apoyar a los coordinadores de laboratorio y proporcionar retroalimentación. Estos profesores expertos ganan 3.500 dólares mensuales y pueden impartir hasta 100 cursos. Totalmente utilizado (con 30 estudiantes por sección), el costo anual por estudiante es de USD 14. En nuestro caso, se ofrecieron 141 cursos diferentes como parte del experimento y, por lo tanto, el costo (para dos profesores expertos) fue de USD 19,85 por estudiante por año.

²² Tenga en cuenta que en nuestra implementación, algunos coordinadores informaron que iban a más de 10 escuelas diferentes por semana.

Computación en la nube

El coste mensual de los servidores de Amazon Web Services (AWS) para 50.000 estudiantes es de ~ 2.500 dólares, lo que equivale a aproximadamente 0,05 dólares por estudiante al mes, o alrededor de 0,6 dólares al año. Este costo incluye componentes fijos y variables, con un mínimo de USD 800 mensuales.

Soporte técnico y mantenimiento

El costo anual de dos ingenieros, cada uno de los cuales gana USD 4.500 mensuales, es fijo. Pueden admitir hasta 1.000 cursos. El costo anual total es el mayor de 12 por USD 9.000 o 12 por USD 9 multiplicado por el número de cursos (n). Para el experimento, este costo se traduce en 24,48 USD por año. Sin embargo, a medida que Conecta Ideas se expande, este costo debería aproximarse a cero por estudiante.

Otros costos

Hay un director de proyecto, que gana USD 8.000 por mes, que puede administrar hasta 5.000 cursos, lo que genera un costo anual de USD 96.000. Esto se traduce en un costo de USD 0.64 por estudiante si se utiliza en su totalidad. Para el experimento, este costo equivale a USD 21,76 por estudiante por año.

Además, Conecta Ideas agrega los siguientes costos a cualquier proyecto: 25% de gastos generales, 10% para contingencias y 20% de ganancias.

Resumen

En resumen, el costo por estudiante (por año) para 147 secciones tratadas es de USD 170,8 (. A escala (con más de 5.000 secciones tratadas), muchos de estos costos disminuyen, ya que son principalmente fijos, y el costo por estudiante (por año) pasa a ser de USD 91,05 $(.) = (43.5 + 19.85 + 0.6 + 24.48 + 21.76) * 1.55 = (43.5 + 14 + 0.6 + 0.64) * 1.55$

Conclusiones

El estudio evalúa el impacto de un software de aprendizaje ("Conecta Ideas") en los niveles de logro de los estudiantes en Chile. El programa fue asignado aleatoriamente, lo que permite recuperar el efecto causal de reemplazar las clases regulares de matemáticas por sesiones de Conecta Ideas. Estimamos el impacto utilizando datos del Sistema de Medición de Calidad de la Educación (SIMCE). Hay tres grupos de tratamiento diferentes (con 47 escuelas cada uno): uno con una versión "simple" o "simple vainilla" de Conecta Ideas, en el que los estudiantes reciben puntos simbólicos por completar ejercicios en la plataforma. Un segundo grupo recibe incentivos individuales, en los que los puntos simbólicos sirven como boletos de rifa para un premio. Un tercer grupo recibe incentivos grupales, en los que se suman los puntos simbólicos de cada aula y estos sirven como boletos de rifa de un premio para toda la clase. Además, hay un grupo de control (con 94 escuelas) donde las clases de matemáticas se dan como de costumbre. Las principales conclusiones del estudio son:

- La sustitución de las clases regulares de matemáticas por sesiones de Conecta Ideas no mejoró el rendimiento de los estudiantes de 4° grado en matemáticas medido por la prueba estandarizada SIMCE
- Conecta Ideas tuvo un efecto negativo, aunque estadísticamente insignificante, en las puntuaciones de matemáticas.
- Proporcionar incentivos individuales o grupales no cambió significativamente el uso de la plataforma. Consistentemente, el impacto en los puntajes de matemáticas es similar si los estudiantes reciben incentivos individuales o grupales.
- Hay algunas pruebas sugestivas de efectos secundarios negativos de las ideas de Conecta en las puntuaciones lingüísticas. Hay un efecto negativo y estadísticamente significativo en las puntuaciones lingüísticas para el grupo con incentivos individuales (las estimaciones puntuales son negativas, pero no estadísticamente significativas para los otros dos grupos de tratamiento).

- La evaluación del proceso sugiere que hubo varios problemas con la implementación del programa, incluyendo retrasos en el inicio del programa, infraestructura inadecuada en las escuelas, pérdida de tiempo en la transición entre el aula regular y el laboratorio de computación, y premios inadecuados para los incentivos para motivar a los estudiantes. Estos problemas pueden ayudar a explicar la falta de efecto en las puntuaciones de matemáticas. No está claro por qué las puntuaciones lingüísticas pueden haber bajado.

Además, encontramos que

- Conecta Ideas no tiene ningún efecto sobre la autopercepción o motivación de los estudiantes.
- Un análisis dosis-respuesta proporciona resultados contradictorios. El uso de la plataforma se correlaciona con los puntajes de las pruebas. Sin embargo, esto también es cierto para las puntuaciones previas al tratamiento. El uso de un enfoque de variables de instrumento sugiere que el uso de la plataforma más no mejora significativamente las puntuaciones en matemáticas y lenguaje (incluso puede disminuir las puntuaciones).
- No hay evidencia de efectos heterogéneos del tratamiento en cuanto al sexo, el estado de NSE o el rendimiento inicial.

Integración del proceso y de las evaluaciones de impacto

Apoyo a la teoría del cambio

Los resultados de la evaluación de impacto sugieren que la teoría del cambio no funcionó como se esperaba en la práctica.²³ El programa no mejoró el rendimiento de los estudiantes en matemáticas.

La evaluación del proceso proporciona algunas ideas sobre por qué este puede ser el caso. Hubo varios problemas con la implementación, incluyendo retrasos en el inicio del programa, infraestructura inadecuada en las escuelas, que obligó a los estudiantes a compartir computadoras durante las sesiones de Conecta Ideas, pérdida de tiempo en la transición entre el aula regular y el laboratorio de computación, y problemas de internet, y premios inadecuados para los incentivos para motivar a los estudiantes. Sin embargo, el programa también contaba con algunos componentes que no estaban incluidos en el diseño original, como las pruebas prácticas del SIMCE 2022, que podrían sesgar los resultados a favor de Conecta Ideas.

En general, los estudiantes de las escuelas tratadas pasaron un tiempo promedio de ~ 17,16 minutos por semana en la plataforma y realizaron ~22 ejercicios semanales. Esto equivale a aproximadamente 642 ejercicios en el transcurso del experimento y 8,5 horas pasadas en la plataforma. Estos números sugieren que los estudiantes están perdiendo efectivamente tiempo de instrucción de matemáticas debido al tratamiento, ya que las sesiones de Conecta Ideas que se llevan a cabo durante una clase regular de matemáticas están destinadas a durar entre 60 y 90 minutos. Esta pérdida de tiempo de instrucción podría explicar la falta de efectos (positivos) del tratamiento en los puntajes de las pruebas.

Interpretación

Estos resultados sugieren que las sesiones de Conecta Ideas no son un sustituto adecuado de las clases regulares de matemáticas. Si bien una evaluación previa había mostrado efectos positivos de complementar las lecciones regulares de matemáticas con sesiones de Conecta Idea (Araya et al., 2019), esta sugiere que reemplazar las lecciones regulares puede afectar negativamente el rendimiento de los estudiantes, incluso en otras asignaturas diferentes a las matemáticas.

Como concluye una revisión reciente del aprendizaje asistido por computadora (CAL, por sus siglas en inglés), "Descubrimos que el simple hecho de proporcionar a los estudiantes acceso a la tecnología produce resultados en gran medida mixtos" (Escueta, Nickow, Oreopoulos, & Quan, 2020). Por lo tanto, nuestra evidencia se alinea con hallazgos previos que arrojan efectos nulos o negativos de "EduTech".

²³ Por ejemplo, las sesiones condujeron a la pérdida de tiempo real de instrucción. Además, los docentes nunca fueron capacitados en el uso de Conecta Ideas. Los incentivos no fueron suficientes para motivar a los estudiantes.

Sin embargo, la revisión también destacó que "la calidad de la implementación es una característica importante de CAL, y el grado en que un software ayuda con éxito a los profesores a involucrar a los estudiantes en el tiempo dedicado al aprendizaje académico que coincide adecuadamente con los niveles de aprendizaje reales de los estudiantes puede ser un determinante importante de la capacidad del software CAL para mejorar drásticamente el aprendizaje de los estudiantes. Se necesita mucha más investigación para ayudarnos a aislar los mecanismos de cuándo y por qué ciertos programas de CAL mejoran la formación de habilidades cognitivas y dónde estos hallazgos pueden ser generalizables" (Escueta et al., 2020). Dado que la implementación distó mucho de ser ideal en esta evaluación, no está claro si una implementación de mayor calidad (por ejemplo, más meses de exposición, con menos pérdida de tiempo en cada sesión) aún podría conducir a resultados positivos.

Nuestros resultados no son concluyentes sobre los efectos de combinar incentivos (individuales o grupales) a los estudiantes con software educativo, ya que los premios otorgados a los estudiantes no los motivaron.

Limitaciones y lecciones aprendidas

Este estudio refleja los efectos de Conecta Ideas tal como se implementó. Es posible que tener más coordinadores de laboratorio, sesiones más cortas o largas con más o menos frecuencia y alterar otros detalles de la intervención pueda tener un efecto diferente (y positivo) en el aprendizaje de los estudiantes.

Nuestros resultados no deben tomarse como una prueba definitiva de que el software Conecta Ideas no tiene ningún impacto en el aprendizaje de los estudiantes per se. Más bien, reemplazar las lecciones regulares de matemáticas con sesiones de Conecta Ideas, dadas las limitaciones y el contexto actuales (por ejemplo, en términos de infraestructura) no es beneficioso para los estudiantes.

Para futuras iteraciones, si los estudiantes reciben incentivos, es importante realizar un piloto y grupos focales para identificar premios que motiven a los estudiantes.

Investigaciones futuras

Una pregunta que aún queda es si el efecto positivo que se mostró al complementar las lecciones regulares de matemáticas con las sesiones de Conecta Ideas (Araya et al., 2019) está impulsado por el tiempo adicional que los estudiantes dedican a aprender matemáticas o por el software en sí. Por lo tanto, un esfuerzo de investigación futuro podría comparar lecciones de matemáticas adicionales (más allá de las clases regulares) con sesiones adicionales de Conecta Idea (que también ocurren además de las clases regulares de matemáticas).

La cuestión de si los incentivos individuales o grupales funcionan mejor para motivar a los estudiantes también sigue abierta. Un esfuerzo de investigación futuro podría replicar nuestro diseño experimental, pero usar premios que motiven a los estudiantes.

VIII. Referencias

- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1), 1–35.
- Abdie, H., & Valentin, D. (2007). Multiple correspondence analysis. *Encyclopedia of Measurement and Statistics*.
- Agencia de Calidad de la Educación. (2022). Programa de evaluaciones educativas. directores(as) y docentes. educación básica y media. 2022. Retrieved from <https://www.agenciaeducacion.cl/noticias/conoce-el-programa-de-evaluaciones-educativas/>
- Alasino, E., Ramírez, M. J., Romero, M., Schady, N., & Uribe, D. (2024). Learning losses during the covid-19 pandemic: Evidence from Mexico. *Economics of Education Review*, 98, 102492.
- Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430), 431–442.
- Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, 592(7854), 403–408.
- Araya, R., Ortiz, E., Bottan, N., & Cristia, J. (2019). Does gamification in education work? experimental evidence from Chile. IDB Working Paper Series-Inter-American Development Bank (IDB-WP-982).
- Baker, F. B., & Kim, S.-H. (2017). The basics of item response theory using R. Springer.
- Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007, 08). Remedying Education: Evidence from Two Randomized Experiments in India*. *The Quarterly Journal of Economics*, 122(3), 1235–1264. Retrieved from <https://doi.org/10.1162/qjec.122.3.1235> doi: 10.1162/qjec.122.3.1235
- Barrow, L., Markman, L., & Rouse, C. E. (2009, February). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy*, 1(1), 52–74. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/pol.1.1.52> doi: 10.1257/pol.1.1.52
- Bau, N., & Das, J. (2020, February). Teacher value added in a low-income country. *American Economic Journal: Economic Policy*, 12(1), 62–96.
- Bethausen, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of "the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, 7(3), 375–385.
- Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O., & Cruz-Aguayo, Y. (2015). One laptop per child at home: Short-term impacts from a randomized experiment in Peru. *American Economic Journal: Applied Economics*, 7(2), 53–80.
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, 314(7080), 572.
- Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., . . . Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa. *Journal of Economic Perspectives*, 31(4), 185–204.
- Brunetti, A., Buchel, K., Jakob, M., Jann, B., & Steffen, D. (2023). Inadequate teacher content knowledge "and what could be done about it: evidence from El Salvador. *Journal of Development Effectiveness*, 1–24.
- Canales, A., & Webb, A. (2018). Educational achievement of indigenous students in Chile: School composition and peer effects. *Comparative Education Review*, 1–43.
- Carril, A. (2015, November). RANDTREAT: Stata module to randomly assign treatments uneven treatments and deal with misfits. *Statistical Software Components*, Boston College Department of Economics. Retrieved from <https://ideas.repec.org/c/boc/bocode/s458106.html>

- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. doi: 10.18637/jss.v048.i06
- Columbia. (2024). Item response theory. Retrieved from <https://www.publichealth.columbia.edu/research/population-health-methods/item-response-theory>
- Cristia, J., Ibarra, P., Cueto, S., Santiago, A., & Severín, E. (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics*, 9(3), 295–320.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Das, J., & Zajonc, T. (2010). India shining and bharat drowning: Comparing two indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics*, 92(2), 175–187.
- De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
- DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications. 83
- Conecta Ideas Impact Evaluation
- Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4, 3895–3962.
- Escueta, M., Nickow, A. J., Oreopoulos, P., & Quan, V. (2020, December). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, 58(4), 897–996. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/jel.20191507> doi: 10.1257/jel.20191507
- Evans, D. K., & Yuan, F. (2022). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis*, 44(3), 532–540.
- Fairlie, R. W., & Robinson, J. (2013a). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. *American Economic Journal: Applied Economics*, 5(3), 211–240.
- Fairlie, R. W., & Robinson, J. (2013b, July). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. *American Economic Journal: Applied Economics*, 5(3), 211–40. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/app.5.3.211> doi: 10.1257/app.5.3.211
- Jackson, K., & Makarin, A. (2018, August). Can online off-the-shelf lessons improve student outcomes? evidence from a field experiment. *American Economic Journal: Economic Policy*, 10(3), 226–54. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/pol.20170211> doi: 10.1257/pol.20170211
- Kawulich, B. B. (2005, May). Participant observation as a data collection method. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2). Retrieved from <https://www.qualitative-research.net/index.php/fqs/article/view/466> doi: 10.17169/fqs-6.2.466
- Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130), 297–300.
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. Retrieved from <https://doi.org/10.3102/0034654315581420> doi: 10.3102/0034654315581420
- Lai, J., & Widmar, N. O. (2021). Revisiting the digital divide in the covid-19 era. *Applied Economic Perspectives and Policy*, 43(1), 458–464. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/aepp.13104> doi: <https://doi.org/10.1002/aepp.13104>

- Le, S., Josse, J., & Husson, F. (2008). Factominer: an r package for multivariate analysis. *Journal of statistical software*, 25, 1–18.
- Malamud, O., & Pop-Eleches, C. (2011). Home computer use and the development of human capital. *The Quarterly journal of economics*, 126(2), 987–1027.
- Markopoulos, A. P., Fragkou, A., Kasidiaris, P. D., & Davim, J. P. (2015). Gamification in engineering education and professional training. *International Journal of Mechanical Engineering Education*, 43(2), 118–131.
- Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019, 04). Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania. *The Quarterly Journal of Economics*, 134(3), 1627–1673. doi: 10.1093/qje/qjz010
- Mbiti, I., Romero, M., & Schipper, Y. (2019, May). Designing effective teacher performance pay programs: Experimental evidence from tanzania (Working Paper No. 25903). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w25903> doi: 10.3386/w25903
- McKenzie, D., & Ozier, O. (2019). Why ex-post power using estimated effect sizes is bad, but an ex-post mde is not. Retrieved from <https://blogs.worldbank.org/impactevaluations/why-ex-post-power-using-estimated-effect-sizes-bad-ex-post-mde-not>
- Moscoviz, L., & Evans, D. K. (2022). Learning loss and student dropouts during the covid-19 pandemic: A review of the evidence two years after schools shut down. Center for Global Development, Working Paper, 609.
- Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review*, 109(4), 1426–60.
- Naik, G., Chitre, C., Bhalla, M., & Rajan, J. (2020). Impact of use of technology on student learning outcomes: Evidence from a large-scale experiment in india. *World Development*, 127, 104736. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0305750X19303857> doi: <https://doi.org/10.1016/j.worlddev.2019.104736>
- Nunnally, J. C. (1994). *Psychometric theory 3e*. Tata McGraw-hill education. 84 Conecta Ideas Impact Evaluation
- Patrinos, H. A., Vegas, E., & Carter-Rau, R. (2022). An analysis of covid-19 student learning loss (Policy Research Working Paper Series No. 10033). The World Bank: The World Bank.
- Romero, M., & Sandefur, J. (2021, 11). Beyond Short-Term Learning Gains: the Impact of Outsourcing Schools in Liberia After Three Years. *The Economic Journal*, 132(644), 1600–1619. Retrieved from <https://doi.org/10.1093/ej/ueab087> doi: 10.1093/ej/ueab087
- Romero, M., Sandefur, J., & Sandholtz, W. A. (2020). Outsourcing education: Experimental evidence from Liberia. *American Economic Review*, 110(2), 364–400. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/aer.20181478> doi: 10.1257/aer.20181478
- Saunders, R. P., Evans, M. H., & Joshi, P. (2005). Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. *Health promotion practice*, 6(2), 134–147.
- Schady, N., Holla, A., Sabarwal, S., Silva, J., & Yi Chang, A. (2023). Collapse and recovery: How the COVID-19 pandemic eroded human capital and what to do about it (Tech. Rep.). The World Bank: The World Bank.
- Schleicher, A. (2019). *Pisa 2018: Insights and interpretations*. oecd Publishing.
- Singh, A. (2015). How standard is a standard deviation? A cautionary note on using SDs to compare across impact evaluations in education. Retrieved 31/07/2017, from <http://blogs.worldbank.org/impactevaluations/how-standard-standard-deviation-cautionary-note-using-sds-compare-across-impact-evaluations>

Singh, A. (2020). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, 18(4), 1770–1813.

Singh, A., Romero, M., & Muralidharan, K. (2022, October). Covid-19 learning loss and recovery: Panel data evidence from india (Working Paper No. 30552). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w30552> doi: 10.3386/w30552

Unicef. (2021). Covid-19 and school closures: One year of education disruption. United Nations Children's Fund (UNICEF), New York. van der Linden, W. J. (2017). *Handbook of item response theory*. CRC Press.