

Araya, Roberto; Arias Ortiz, Elena; Bottan, Nicolas L.; Cristia, Julian

Working Paper

¿Funciona la gamificación en la educación? Evidencia experimental de Chile

IDB Working Paper Series, No. IDB-WP-982

Provided in Cooperation with:

Inter-American Development Bank (IDB), Washington, DC

Suggested Citation: Araya, Roberto; Arias Ortiz, Elena; Bottan, Nicolas L.; Cristia, Julian (2019) : ¿Funciona la gamificación en la educación? Evidencia experimental de Chile, IDB Working Paper Series, No. IDB-WP-982, Inter-American Development Bank (IDB), Washington, DC, <https://doi.org/10.18235/0001777>

This Version is available at:

<https://hdl.handle.net/10419/208170>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>

DOCUMENTO DE TRABAJO DEL BID N° IDB-WP-982

¿Funciona la gamificación en la educación?

Evidencia experimental de Chile

Roberto Araya
Elena Arias Ortiz
Nicolas Bottan
Julian Cristia

Banco Interamericano de Desarrollo
Departamento de Investigación y Economista Jefe

Julio de 2019

¿Funciona la gamificación en la educación?

Evidencia experimental de Chile

Roberto Araya*

Elena Arias Ortiz**

Nicolas Bottan***

Julian Cristia****

*Universidad de Chile

**Banco Interamericano de Desarrollo

***Cornell University

****Banco Interamericano de Desarrollo

Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo

¿Funciona la gamificación en la educación?: evidencia experimental de Chile / Roberto Araya, Elena Arias Ortiz, Nicolas Bottan, Julián Cristia.

p. cm. — (Documento de trabajo del BID ; 982)

Incluye referencias bibliográficas.

1. Gamification-Chile. 2. Educational games-Chile. 3. Educational technology-Chile. 4. Computer-assisted instruction-Chile. 5. Mathematics-Study and teaching-Chile. 6. Educational innovations-Chile. I. Araya, Roberto. II. Arias Ortiz, Elena. III. Bottan, Nicolas. IV. Cristia, Julián. V. Banco Interamericano de Desarrollo. Departamento de Investigación y Economista Jefe. VI. Serie.

IDB-WP-982

<http://www.iadb.org>

Copyright © 2019 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Después de un proceso de revisión por pares, y con el consentimiento previo y por escrito del BID, una versión revisada de esta obra podrá reproducirse en cualquier revista académica, incluyendo aquellas referenciadas por la Asociación Americana de Economía a través de EconLit, siempre y cuando se otorgue el reconocimiento respectivo al BID, y el autor o autores no obtengan ingresos de la publicación. Por lo tanto, la restricción a obtener ingresos de dicha publicación sólo se extenderá al autor o autores de la publicación. Con respecto a dicha restricción, en caso de cualquier inconsistencia entre la licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas y estas declaraciones, prevalecerán estas últimas.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



Resumen¹

La gamificación, o introducción de elementos de juegos en contextos ajenos a juegos, tiene el potencial de mejorar el aprendizaje aumentando la motivación de los alumnos. Sin embargo, hay escasa evidencia rigurosa sobre su efectividad. En este documento, evaluamos experimentalmente un programa de tecnología innovadora que utiliza la gamificación para aumentar el aprendizaje de matemática en escuelas primarias con bajos resultados en Chile. El programa ConectaIdeas comprende dos sesiones semanales en un laboratorio de computación durante las cuales los alumnos utilizan una plataforma online para trabajar en ejercicios de matemática. La plataforma registra cuántos ejercicios los alumnos realizan e incluye diferentes tipos de competencias individuales y grupales para promover la motivación de los alumnos. Nuestra evaluación indica que el programa generó grandes efectos positivos en el aprendizaje de matemática, de 0,27 desviaciones estándar, en el examen nacional estandarizado chileno (no se encontraron efectos en lenguaje). El programa también influyó en diversos resultados no académicos de maneras positivas y negativas. Por un lado, aumentó la preferencia de los alumnos para utilizar la tecnología en el aprendizaje de matemática y promovió la idea entre los alumnos de que la inteligencia se puede mejorar cuando uno estudia intensamente. Por otro lado, el programa aumentó la ansiedad asociada a estudiar matemática y redujo las preferencias de los alumnos para trabajar en equipo. Estos resultados sugieren que la gamificación puede ser una herramienta potente para mejorar el aprendizaje de los alumnos, pero que su uso puede acarrear consecuencias no deseadas.

Códigos JEL: I21, I28

Palabras clave: educación, gamificación, tecnología, experimentos

¹ Este proyecto no se habría materializado sin el apoyo y el compromiso de Ben Petrazzini. Queremos agradecer especialmente a Analía Jaimovich, Nora Cuevas (alcaldesa de San Bernardo), Ruth Carrillo, Dina Herrera, José Sotelo y Julio Torres por sus importantes contribuciones. También agradecemos a Xiomara Pulido, Maritza Moraga, Jimena Romero, Obed Ulloa y Anastasiya Yarygina por su valiosa ayuda como asistentes de investigación. El equipo del Centro de Investigación Avanzada en Educación de la Universidad de Chile, que incluía a Carlos Aguirre, Raúl Gormaz, Paulina Jaure y Arturo Prat, jugó un rol crucial en la implementación de este proyecto. Agradecemos los excelentes comentarios y sugerencias de Paulo Bastos, Samuel Berlinski, Marcelo Cabrol, Alejandrina Cristia, Colleen Eddy, Sebastián Espinoza, Alfonso Flores-Lagunes, Sebastián Gallegos, Ana María Ibañez, Mike Lovenheim, Julián Messina, Doug Miller, Francisco Pardo, Pamela Reyes-Santander, Amy Schwartz, Emiliana Vegas, Jouni Viiri, Ana Yarygina y los participantes del seminario en el taller de la División de Educación del Banco Interamericano de Desarrollo, la Conferencia del Pacífico para Economía del Desarrollo de 2019, la Conferencia Anual de la Sociedad de Investigación y Efectividad en la Educación de 2019, la Conferencia sobre Investigación para Mejorar los Sistemas Educativos de 2019 y el taller de verano de Daniel Patrick Moynihan Syracuse/ Seminario Cornell de Verano de Educación y Políticas Sociales. Agradecemos a María Paz Espinosa y Camila González Nakada de MIDE UC y a Cecilia Lecaros, de APTUS por su apoyo en las actividades de recopilación de datos, y también agradecemos el financiamiento del Centro Internacional de Investigaciones para el Desarrollo y el BID. Por último, Roberto Araya agradece al Financiamiento Basal para Centros Científicos y Tecnológicos de Excelencia FB0003 PIA-CONYCIT y agradece igualmente el financiamiento inicial de la Corporación de Fomento de la Producción (CORFO) y del Fondo Nacional de Desarrollo Tecnológico y Productivo (FONTEC) que apoyó el desarrollo del programa. Las opiniones expresadas en este documento pertenecen a los autores y no deberían atribuirse al Banco Interamericano de Desarrollo.

1. Introducción

La introducción de elementos de juego en contextos ajenos a juegos -conocido como gamificación- se ha convertido en una estrategia cada vez más habitual utilizada en educación, en salud y en las empresas para motivar a las personas a adoptar ciertas conductas deseadas. Por ejemplo, el dispositivo “Fitbit” mide el número de pasos que una persona da al día, envía un mensaje de felicitación cuando se alcanza una meta de número de pasos y permite que la competencia entre usuarios genere aún más motivación. Como este caso demuestra, la idea básica de la gamificación es que la introducción de elementos sencillos del juego como puntos, insignias y rankings pueden transformar una tarea aburrida en una actividad atractiva.

La gamificación puede jugar un importante rol en la educación, teniendo en cuenta que la motivación de los alumnos ha sido reconocida hace tiempo como un elemento central del aprendizaje (Weiner 1990). Sin embargo, puede que existan desventajas en su uso relacionados con potenciales reducciones de la motivación intrínseca, aumentos de la ansiedad y efectos motivacionales de corta duración (Barata *et al.* 2013; McDaniel, Lindgren y Friskics 2012, Hanus y Fox 2015). A pesar de estas posibles desventajas, la gamificación en educación es una industria floreciente. Impulsada por el aumento del acceso a dispositivos conectados a internet en todo el mundo, organizaciones como Duolingo y Khan Academy prestan apoyo a más de 10 millones de alumnos todos los meses.² La investigación sobre este tema también ha aumentado notablemente. Según Google Scholar, el número de artículos publicados anualmente que contienen las palabras “gamification” y “education” pasó de 140 en 2010 a 3.570 en 2014 y llegó a 9.570 en 2018 (se buscan estas palabras en inglés dado que la mayoría de los artículos académicos están en este idioma).

Sin embargo, ¿funciona la gamificación? Es decir, ¿es verdad que los programas educativos que introducen elementos de juegos para aumentar la motivación generan grandes aumentos en el aprendizaje? Desafortunadamente, a pesar del gran número de estudios sobre la gamificación en la educación, se observa una escasez de evidencia empírica rigurosa que mida sus efectos en el aprendizaje.

Este artículo pretende contribuir a cerrar esta brecha del conocimiento evaluando experimentalmente un programa que utiliza la gamificación intensamente para mejorar los resultados académicos. El programa, denominado ConectaIdeas, se propone mejorar el aprendizaje de matemática entre los alumnos de cuarto básico en Chile. Los alumnos que participan practican ejercicios de matemática en una plataforma online durante sesiones de aprendizaje semanales de 90 minutos durante el horario escolar regular. El programa emplea un conjunto de estrategias de gamificación para promover el uso intensivo de la plataforma

² Duolingo es una aplicación para el aprendizaje de lenguas. La aplicación tiene 25 millones de usuarios cada mes y su valor en el mercado en 2017 era de USD 700 millones (Buchanan 2018). Khan Academy es una organización sin fines de lucro que proporciona contenidos y ejercicios de matemática, ciencias, historia y otras disciplinas. Su sitio web tenía 12 millones de usuarios al mes en 2016 (<https://khanacademyannualreport.org/>).

de aprendizaje. En primer lugar, la plataforma muestra a cada alumno cuántos ejercicios ha completado y compara esta cifra con el promedio de la sección. Además, los alumnos pueden observar el número de ejercicios completados por cada alumno en la sección. En segundo lugar, se muestran “publicidades” a los alumnos frecuentemente las cuales buscan motivarlos promoviendo la idea de que la inteligencia se puede mejorar cuando uno estudia intensamente. En tercer lugar, secciones enteras de alumnos de una escuela participan en competencias semanales con secciones en otras escuelas para ver que grupo realiza un mayor número promedio de ejercicios en la plataforma de aprendizaje. En cuarto lugar, secciones de alumnos también participan en competencias “en vivo” entre escuelas cada dos meses en las cuales los alumnos compiten en la solución de problemas de matemática incorporadas en un juego online.

¿Mejora ConectaIdeas el aprendizaje de los alumnos? Para responder a esta pregunta llevamos a cabo una evaluación experimental en 24 escuelas primarias públicas de alumnos de bajos ingresos en Santiago, Chile. Los alumnos no sólo provenían de un contexto socioeconómicamente vulnerable sino también tenían retrasos en el aprendizaje: en promedio, tenían un puntaje de 0,7 desviaciones estándar por debajo del promedio nacional en matemática. Para medir impactos, asignamos aleatoriamente una sección de cuarto básico en cada una de estas escuelas al grupo de tratamiento, y asignamos la otra sección en ese grado al grupo de control. Por ejemplo, la sección de 4^{to} “A” en una escuela participante en el estudio pudo haber sido asignada al grupo tratamiento mientras que la sección de 4^{to} “B” de esa misma escuela pudo haber sido asignada al grupo de control. Los datos de línea de base se recopilaron en marzo de 2017 y el programa se implementó inmediatamente después y luego siguió hasta noviembre de 2017 (el año escolar en Chile comienza en marzo y acaba en diciembre).

Nuestro principal indicador de resultado lo obtuvimos del examen nacional estandarizado de Chile que se llevó a cabo en noviembre de 2017 (después de 7 meses de exposición al programa). Se trata de una evaluación en papel implementada anualmente en todas las escuelas del país para monitorear el aprendizaje en matemática y lenguaje. La medición de los efectos en pruebas de este tipo es importante porque la evidencia muestra que los tamaños del efecto de intervenciones educativas varían considerablemente entre diferentes tipos de pruebas. De hecho, Hill *et al.* (2018) analizaron un grupo de evaluaciones experimentales en educación en Estados Unidos y documentaron que el efecto promedio en las pruebas estandarizadas amplias era de 0,07 desviaciones estándar, en comparación con un efecto promedio de 0,23 en las pruebas estandarizadas enfocadas en ciertos temas (por ejemplo, fracciones) y de 0,44 en las pruebas especializadas elaboradas para intervenciones específicas. Por lo tanto, utilizar un examen estandarizado nacional amplio como indicador de resultado principal de nuestra evaluación permite estimar cómo el escalamiento del programa podría afectar los puntajes en la principal evaluación utilizada por un Ministerio de Educación para monitorear la calidad y la equidad del aprendizaje.

Los resultados señalan que ConectaIdeas generó una importante mejora en el aprendizaje de matemática, la cual es estadísticamente significativa. Nuestra especificación preferida muestra un efecto de 0,27 desviaciones estándar. Los efectos en los resultados de matemática son similares en las diferentes muestras de alumnos definidas por género, educación de la madre y rendimiento académico en la línea de base. Aunque el programa se proponía mejorar el aprendizaje en matemática, podría haber generado efectos en otras disciplinas. Sin embargo, los efectos estimados en lenguaje son cercanos a cero y estadísticamente no significativos.

De forma de poner en contexto los efectos encontrados, los comparamos con los de otras evaluaciones educativas que también han analizado los efectos en el examen nacional estandarizado en Chile. Una importante evaluación midió los efectos de ampliar la jornada escolar de cuatro a siete horas al día en las escuelas del país. Este programa, el cual es considerado una de las principales reformas educativas recientes de Chile y que generó un aumento masivo del gasto en educación, mejoró los resultados en matemática y lenguaje en 0,06 desviaciones estándar (Bellei 2009). A su vez, un programa que proporcionaba planes de clases y materiales a los docentes mejoró los puntajes de las pruebas de matemática y lenguaje en 0,07 y 0,09 desviaciones estándar, respectivamente (Bassi, Meghir y Reynoso 2016). Por lo tanto, los efectos de ConectaIdeas en matemática son aproximadamente cuatro veces mayores que los efectos de estas dos intervenciones.

Para evaluar de manera integral al programa ConectaIdeas, también examinamos los efectos en indicadores de resultado no académicos. Por el lado positivo, observamos que el programa aumentó la preferencia de los alumnos por utilizar computadoras en la enseñanza de matemática, lo cual puede ser importante en un contexto de creciente acceso y uso de la tecnología en diferentes ámbitos de la vida. También encontramos que el programa aumentó la probabilidad de que los alumnos creen que pueden mejorar su inteligencia si se esfuerzan al estudiar. No encontramos evidencia de efectos en la motivación intrínseca para matemática o en el autoconcepto en matemática. Por el lado negativo, encontramos que el programa aumentó la ansiedad asociada con el estudio de matemática y también redujo las preferencias de trabajar en equipo.

Además, pudimos documentar cuánto, cuándo y dónde los alumnos utilizaron la plataforma online utilizando los datos que se van registrando de forma automática en dicha plataforma. Observamos que prácticamente todos los alumnos en el grupo de tratamiento utilizaron la plataforma y que, en promedio, los alumnos utilizaron la plataforma 27 horas durante el período de 7 meses de exposición al programa. Una pregunta clave es si los efectos académicos positivos se pueden explicar parcialmente por el hecho de que los alumnos practicaban matemática en sus casa. Sin embargo, la evidencia es inequívoca en este sentido, ya que el uso en el hogar representa sólo un 2% del tiempo de conexión y, por lo tanto, no puede explicar los resultados observados. También documentamos que el tiempo dedicado a la plataforma se mantiene constante durante la implementación del programa. Esta patrón

contrasta con la marcada disminución en el uso a lo largo del tiempo documentada en programas que proporcionan computadoras o internet para el uso en el hogar (Malamud *et al.*, 2019). Por lo tanto el programa ConectaIdeas pudo evitar el fuerte efecto de novedad observado en estos otros programas.

Las estimaciones de la evaluación experimental ya descrita corresponden a la implementación del programa durante el año escolar de 2017. Sin embargo, este gran efecto registrado, ¿es sólo un resultado particular o representa el efecto típico del programa? Para explorar este aspecto, generamos estimaciones no experimentales que aprovechan la implementación de ConectaIdeas en 11 escuelas en Santiago entre 2011 y 2016, junto con datos longitudinales a nivel de las escuelas de los exámenes nacionales estandarizados. Utilizando un enfoque de diferencias en diferencias, observamos que ConectaIdeas generó efectos positivos y estadísticamente significativos de entre 0,19 a 0,22 desviaciones estándar en matemática y efectos no estadísticamente significativos en lenguaje. Estos resultados sugieren que los grandes impactos documentados en la evaluación experimental son representativos del efecto típico del programa.

Nuestro estudio está relacionado con una amplia literatura en educación y ciencias de la computación que ha analizado diferentes aspectos relacionados con la gamificación en educación. Los estudios han analizado teóricamente las ventajas y desventajas potenciales de diferentes modelos de gamificación en educación, han documentado ejemplos de su introducción en contextos concretos y han proporcionado alguna evidencia cualitativa y cuantitativa sobre sus efectos en los resultados de los alumnos (Denny 2013; Domínguez *et al.* 2013; Mekler *et al.* 2017). En general, los estudios de esta literatura han llegado a la conclusión de que incorporar la gamificación puede aumentar la motivación y la participación de los alumnos (Lister 2015; Alsawaier 2018).³ Sin embargo, hay escasa evidencia empírica rigurosa sobre los efectos en los resultados académicos de las intervenciones educativas que utilizan gamificación (Markopoulos *et al.* 2015).

De nuestra parte, no hemos encontrado estudios en la literatura económica que hayan evaluado rigurosamente los efectos de un programa que utilizaba intensamente la gamificación para mejorar los resultados de aprendizaje. Sin embargo, hay dos vertientes de esa literatura que están relacionadas con nuestro estudio. La primera vertiente incluye las evaluaciones de las intervenciones que utilizaron incentivos monetarios para aumentar la motivación de los alumnos. Los estudios que han evaluado los efectos de ofrecer incentivos monetarios a los alumnos generalmente han encontrado efectos positivos, aunque modestos, en los resultados académicos (Bettinger 2012; Fryer 2011). Una excepción es el estudio de Li *et al.* (2014), que señala que cuando se proporcionó incentivos individuales a los alumnos, los efectos en el aprendizaje fueron pequeños, pero que cuando los incentivos se

³ También hay investigaciones de larga data y recientes sobre cómo incorporar juegos y torneos en el aula, sin utilizar tecnología, para aumentar la motivación (Edwards *et al.*, 1972; Slavin, 2010).

proporcionaron para promover competencias entre grupos (y colaboración dentro de los grupos) los efectos en el aprendizaje fueron grandes.

La segunda vertiente incluye estudios experimentales que evaluaron los efectos en el aprendizaje de los programas de enseñanza asistidos por computadora. Las evaluaciones experimentales implementadas en India (Banerjee *et al.* 2007 y Muralidharan, Singh y Ganimian 2019), China (Lai *et al.* 2013, 2015; Mo *et al.* 2014) y Estados Unidos (Dynarski *et al.* 2007; Wijekumar *et al.* 2009; Rutherford *et al.* 2014) han mostrado efectos positivos de estas intervenciones en el aprendizaje, aunque los efectos de los programas implementados en Estados Unidos han sido considerablemente más pequeños.

La principal contribución de este estudio es que presenta una evaluación integral experimental de los efectos de un programa educativo que utiliza la gamificación intensivamente. En particular, el estudio presenta varias ventajas. En primer lugar, presenta estimaciones no sesgadas y precisas debido al diseño de aleatorización dentro de la escuela y al gran número de alumnos que participaron en el estudio (unos 1.100). En segundo lugar, evalúa un programa que se implementa en las escuelas públicas durante la jornada escolar, lo cual es relevante para considerar una futura implementación a gran escala en escuelas públicas. En tercer lugar, mide los efectos en los resultados académicos utilizando un examen estandarizado nacional amplio. En cuarto lugar, informa sobre los efectos del programa en la motivación intrínseca, el autoconcepto, la ansiedad, la creencia de los estudiantes respecto a si estudiar intensamente puede afectar la inteligencia y las preferencias por el trabajo en equipo y relacionado con el uso de la tecnología en la escuela. Por último, el estudio complementa las estimaciones experimentales de un año con estimaciones no experimentales de varios años para proporcionar resultados más concluyentes respecto a los efectos del programa.

El artículo se estructura de la siguiente forma. La Sección 2 describe el programa ConectaIdeas. La Sección 3 detalla el diseño experimental, los datos y la estrategia de identificación, y documenta el balance en la línea de base. La Sección 4 presenta efectos sobre los resultados académicos y la Sección 5 recoge evidencia sobre mecanismos potenciales. La Sección 6 presenta pruebas de robustez que incluyen estimaciones no experimentales de los efectos del programa. Por último, en la Sección 7 se presentan las conclusiones.

2. El programa ConectaIdeas

El programa ConectaIdeas fue desarrollado por un equipo dirigido por el investigador Roberto Araya, actualmente en el Centro de Investigación Avanzada en Educación (CIAE) de la Universidad de Chile. El objetivo del equipo consistía en diseñar un programa que pudiera generar grandes aumentos en el aprendizaje de matemática de alumnos de condición

socioeconómica baja. El principio guía del proyecto era que la introducción de elementos de juegos en la enseñanza de matemática, facilitada por la tecnología, podía aumentar la motivación de los alumnos y generar mayor aprendizaje. Después de años de desarrollo a pequeña escala, el programa ConectaIdeas se implementó entre 2011 y 2016 en 11 escuelas en la comuna de Lo Prado, en Santiago de Chile. Durante este período, el equipo mejoró el diseño y desarrolló protocolos de implementación detallados.

El programa implementado durante la evaluación experimental de 2017 consistía en proporcionar a los alumnos sesiones semanales de aprendizaje de matemática de 90 minutos en el laboratorio de computación. Una de estas sesiones reemplazaba las clases tradicionales de matemática en el aula, mientras que la otra representaba tiempo adicional de clases de matemática.⁴ En una clase típica, los alumnos trabajaban para solucionar el mismo conjunto de 20 a 30 ejercicios que se les asignaba y que estaban relacionados con los temas cubiertos en la enseñanza regular de matemática e incluidos en el programa nacional de estudios. Cuando solucionaban estos problemas, los alumnos recibían retroalimentación automática respecto a si sus respuestas eran correctas o no. Los coordinadores del laboratorio, contratados y supervisados por el equipo del Centro de Investigación Avanzada en Educación, eran responsables de llevar a cabo las sesiones de aprendizaje en el laboratorio de computación en colaboración con los docentes regulares de aula. Los coordinadores del laboratorio eran antiguos docentes que tenían un día de formación, y una supervisión continua del equipo de implementación (los docentes no recibían capacitación formal, pero el programa promovía el aprendizaje con la práctica).

El programa incluía varias estrategias de gamificación. El Gráfico 1 contiene una captura de pantalla de la plataforma que describe varias de estas estrategias. La primera estrategia se centra en motivar a los alumnos realizando un seguimiento de sus avances y estableciendo comparaciones con sus compañeros. Como muestra el Gráfico 1, al alumno se le presenta un gráfico que recoge el número de ejercicios que ha completado cada semana (la línea azul oscura en el gráfico). Esta información se muestra con la intención de motivar al alumno para que su esfuerzo sea visible y concreto. El gráfico también incluye una línea para el promedio de la sección (la línea azul claro en el Gráfico 1). Esta información busca activar los efectos motivacionales presentes en las comparaciones sociales que han demostrado ser importantes en diferentes ámbitos, como la conservación de la energía y el esfuerzo de los trabajadores (Cialdini *et al.* 2007; DellaVigna y Pope 2018).⁵

⁴ En Chile, la gran mayoría de los alumnos, incluidos los que participan en el estudio, asisten a la escuela durante aproximadamente siete horas al día. El horario regular incluye un tiempo de enseñanza obligatoria que las escuelas tienen que asignar a disciplinas específicas, pero también una parte del tiempo que las escuelas pueden asignar a cualquier disciplina. Las escuelas normalmente asignan estas horas a matemática y lenguaje porque son las disciplinas evaluadas en el examen nacional estandarizado. La sesión adicional de matemática del programa ConectaIdeas empleaba parte del tiempo que las escuelas pueden asignar a cualquier disciplina.

⁵ Otra pantalla muestra el ranking de los alumnos individuales en la sección que están ordenados por el número de ejercicios acumulados completados durante el año escolar, la semana pasada o la sesión actual.

La segunda estrategia busca motivar a los alumnos transmitiendo la idea de que la inteligencia es maleable y que se puede mejorar realizando un esfuerzo mientras se estudia - es decir, teniendo una “mentalidad de crecimiento” (Dweck 2006). Con este fin, la plataforma muestra a los alumnos “publicidades” personalizadas que destacan este mensaje. El Gráfico 1 muestra un ejemplo de una de estas publicidades. En este caso, al alumno se le presenta una imagen de un niño tocando el piano y un mensaje que dice “el esfuerzo, y sólo el esfuerzo, *nombre del alumno*, es el camino a la perfección” (donde *nombre del alumno* se reemplaza con el nombre del alumno, por ejemplo, Juan Pérez). Estas imágenes se presentan durante 20 segundos y van acompañadas de un audio del mensaje generado por la computadora. Las imágenes y los mensajes presentados provienen de un conjunto de 10 ejemplos que ponen de relieve la importancia de esforzarse cuando uno estudia.

La tercera estrategia se centra en la motivación grupal en vez de individual. En particular, se organizan competencias entre secciones de alumnos en términos del número promedio de ejercicios completados por semana en la plataforma educativa. Volviendo al Gráfico1, vemos que en el lado derecho de la pantalla se muestran fotos de diferentes secciones de alumnos. Se trata de un subconjunto del ranking de las secciones que participan en el programa las cuales están ordenadas de arriba a abajo según el número promedio de ejercicios completados en la semana. La foto en la cuarta posición (empezando desde arriba) corresponde a la sección del alumno que está utilizando la computadora. Las tres secciones superiores que se muestran corresponden a las que están justo por encima de la sección de ese alumno en términos del número promedio de ejercicios completados, y las tres secciones de abajo son aquellos que se encuentran inmediatamente por debajo en el ranking. Los alumnos pueden hacer clic en cualquiera de estas fotos para conocer el nombre del colegio y de la sección con la cual están compitiendo.

La cuarta estrategia también pretende motivar a los alumnos activando dinámicas sociales positivas y la colaboración dentro del aula. Con ese fin, se organizan torneos “en vivo” cada dos meses. En ellos, los alumnos compiten solucionando ejercicios de matemática incorporados en un juego online. Para este torneo, se programa una hora en la que todas las secciones de alumnos participantes de ConectaIdeas estén conectados a la plataforma. Luego, se arman parejas de alumnos en escuelas diferentes. Estos dos alumnos juegan el “juego de la espiral” que se muestra en el Gráfico 2, en el que se turnan solucionando ejercicios de matemática y mueven “fichas” con el objetivo de colocarlas todas en el centro de la espiral (la celda con el número 143). Cada cinco minutos, los puntajes individuales acumulados por cada alumno se promedian a nivel de la sección y un miembro del equipo del programa ConectaIdeas, quien actúa como “presentador”, informa a los alumnos cuáles son las escuelas con mejores resultados y procura entusiasmar a los participantes.

Además de estas estrategias centradas en la gamificación, la plataforma ConectaIdeas también incluye herramientas para facilitar el trabajo de los docentes. Concretamente, la plataforma proporciona a los docentes y a los coordinadores de laboratorio un panel con

información en tiempo real sobre el número de preguntas contestadas y el número de preguntas correctas de cada alumno. En este panel, los alumnos están ordenados desde aquellos que necesitan más apoyo (aquellos que han contestado pocas preguntas o tienen una baja tasa de respuestas correctas) a aquellos que necesitan menos apoyo. La plataforma también presenta un panel que muestra las tasas de respuestas correctas por pregunta de modo de poder identificar las preguntas en las que todos los alumnos necesitan apoyo. Además, el sistema genera informes de retroalimentación para los coordinadores del laboratorio, los docentes y los directores de escuela.

Por último, presentamos información sobre los costos por alumno asociados con la implementación de ConectaIdeas. Estos costos incluyen los salarios para un coordinador de proyecto y para los coordinadores de laboratorio, así como los costos de diferentes insumos como computadoras, conexiones a internet, software, cloud computing y la coordinación general del proyecto. El costo por alumno de la implementación de ConectaIdeas para la intervención implementada en 2017 fue de alrededor de USD 150, lo que representa un aumento del 5% en el gasto público por alumno de escuela primaria en Chile.

3. Metodología

3.1 Diseño y selección de la muestra

Implementamos una evaluación experimental para evaluar los efectos del programa ConectaIdeas. El equipo de ConectaIdeas fue encargado de seleccionar 24 escuelas públicas ubicadas en Santiago (para simplificar la logística) que tenían al menos 2 secciones de cuarto grado de primaria. Además, las escuelas tenían que haber sido clasificadas por el Ministerio de Educación en las dos categorías más bajas de estatus socioeconómico (de cinco categorías) para probar si ConectaIdeas también podía cerrar las brechas socioeconómicas en rendimiento académico.

El proceso de selección comenzó a finales de enero de 2017 con la identificación de 22 comunas que tenían escuelas que satisfacían los criterios descritos más arriba. Los directores de las comunas fueron contactados primero por correo electrónico y luego por teléfono. El equipo de ConectaIdeas luego visitó a los 11 directores que contestaron y manifestaron su interés. Después, se llevaron a cabo sesiones informativas con los directores de las comunas y los directores de escuela en nueve comunas. En el paso final, el equipo llevó a cabo visitas técnicas en las escuelas para verificar los requisitos técnicos. Las visitas técnicas estaban programadas en 31 escuelas en 6 comunas. Sin embargo, después de visitar cuatro comunas (la Pintana, Maipú, Quinta Normal y San Bernardo) el equipo confirmó 24 escuelas que cumplían con los requisitos arriba mencionados y, por lo tanto, se puso fin al proceso de selección. Es importante mencionar que el procedimiento de selección no involucró que las escuelas tomaran decisiones para autoseleccionarse en el programa.

El Cuadro 1 presenta estadísticas obtenidas del examen nacional estandarizado de 2016 (conocido en Chile como Sistema de Medición de la Calidad de la Educación, SIMCE) que muestra cómo se desarrolló el proceso de construcción de la muestra. La columna (1) presenta las medias del universo de escuelas en todo el país y las columnas (2) a (5) presentan medias de muestras de escuelas que surgen de restringir progresivamente la muestra para incluir los requisitos de elegibilidad. En particular, la columna (2) se limita a escuelas en Santiago y la columna (3) limita aún más la muestra a escuelas en las dos últimas categorías en términos de estatus socioeconómico. Luego, la columna (4) limita aún más la muestra a escuelas con al menos dos secciones en cuarto grado de primaria y la columna (5) presenta la muestra de escuelas que participaron en el estudio. Los resultados en este cuadro evidencian que la composición de la muestra del estudio es bastante similar a la muestra de las escuelas de estatus socioeconómico bajo en Santiago, con dos excepciones: la matriculación en las escuelas del estudio es mayor (debido al requisito de las dos secciones) y sus alumnos tienen resultados aún más bajos en matemática y lenguaje. De hecho, los alumnos participantes del estudio presentan resultados más bajos que el alumno promedio en el país (0,60 desviaciones estándar en lenguaje y 0,68 en matemática).⁶

Adoptamos un diseño con asignación aleatoria a nivel de sección dentro de cada escuela. Es decir, en cada una de las 24 escuelas participantes en el estudio, asignamos aleatoriamente una de las dos secciones de cuarto grado al grupo de tratamiento. Estas secciones participaron en el programa ConectaIdeas. Las otras secciones fueron asignadas al grupo de control y recibieron enseñanza de matemática de forma tradicional. Para las tres escuelas en la muestra que tenían más de dos secciones, sólo incluimos las dos primeras (es decir, las secciones “A” y “B”) en la evaluación. La aleatorización se llevó a cabo antes del levantamiento de los datos de la línea de base y se informó a las escuelas del estatus de tratamiento de cada sección *después* de finalizar dicho levantamiento. Hubo una fidelidad perfecta en términos de la implementación del programa en las secciones de tratamiento y control. Es decir, todas las secciones asignadas al tratamiento participaron en ConectaIdeas y ninguna de las secciones de control participó en el programa. Por último, el porcentaje de estudiantes que participaron de la línea de base y que no participaron en el examen estandarizado nacional de 2017 fue bajo y balanceado entre los grupos de tratamiento y control.⁷

⁶ Los puntajes de las pruebas en el examen nacional estandarizado se normalizaron utilizando la media y la desviación estándar a nivel de todo el país.

⁷ Las tasas de “attrition” para estimar los efectos en los resultados de matemática fueron del 10% para el grupo de tratamiento y del 8% para el grupo de control. Una regresión del estatus de “attrition” en un indicador de tratamiento (controlando por los efectos fijos de la escuela) produce un coeficiente de 0,02 y un error estándar de 0,01. Para lenguaje, las tasas de attrition fueron del 10% para el grupo de tratamiento y del 9% para el grupo de control. La regresión analoga para lenguaje produce un coeficiente de 0,00 y un error estándar de 0,01.

3.2. Estrategia de identificación

La estimación del efecto del programa es sencilla debido a la asignación aleatoria de las secciones a tratamiento dentro de las escuelas. La ventaja de este diseño es que permite tener en cuenta las características de las escuelas, las cuales pueden influir en ambas secciones, al incluirse efectos fijos por escuela. Además, dado que el “inter-cluster correlation” a nivel de secciones es cercano a cero, una vez que se agregan los efectos fijos de las escuelas, nuestro diseño es casi tan preciso como un diseño que involucra la aleatorización a nivel individual.

Estimamos los efectos del programa bajo dos especificaciones principales, la primera de las cuales implica estimar la siguiente ecuación:

$$y_{ics}^{post} = \alpha_1 + \beta_1 * Tratamiento_{cs} + \phi_s + \varepsilon_{ics} \quad (1)$$

donde y_{ics}^{post} es la variable de resultado en la línea final (por ejemplo, el puntaje de la prueba de matemática en el examen nacional estandarizado) para el alumno i , en la sección c , en la escuela s . $Tratamiento_{cs}$ es un indicador igual a uno si la sección fue asignada al grupo de tratamiento y cero en caso contrario. ε_{ics} es el término de error, que no debería estar correlacionado con la variable de tratamiento debido a la asignación aleatoria, y ϕ_s son efectos fijos por escuela. El coeficiente de interés, β , estima el efecto promedio del programa en la variable de resultado.

En una segunda especificación, también controlamos por el valor de la variable de resultado en la línea de base:

$$y_{ics}^{post} = \alpha_2 + \beta_2 * Tratamiento_{cs} + \gamma_2 * y_{ics}^{pre} + \phi_s + \varepsilon_{ics} \quad (2)$$

donde y_{ics}^{pre} es el puntaje del estudiante en la línea de base de la asignatura respectiva. Es decir, cuando se estiman los efectos en matemática, controlamos por el puntaje en matemática en la línea de base y cuando estimamos los efectos en lenguaje controlamos por el puntaje en esa asignatura en la línea de base. Dado que el aprendizaje está estrechamente correlacionado a lo largo del tiempo, controlar por los puntajes de la línea de base normalmente aumenta la precisión estadística. Además, hacer este puede acomodar potenciales diferencias en puntaje en la línea de base. Por consiguiente, esta es nuestra especificación preferida.

Por último, todas las estimaciones presentadas en este documento incluirán errores estándar estimados con “cluster” a nivel de sección (la unidad de aleatorización). Una preocupación potencial es que, dado que estamos usando “cluster” de los errores estándar con 48 secciones, podríamos estar exagerando la precisión de nuestras estimaciones debido a contar con un número relativamente modesto de grupos (Cameron y Miller 2015). Por

consiguiente, en la sección de robustez, mostramos resultados adicionales utilizando estrategias alternativas para estimar los errores estándar.

3.3. Los datos

Nuestro análisis se basa en una combinación de datos provenientes del examen nacional estandarizado en Chile, de encuestas y de la plataforma ConectaIdeas. El principal indicador de resultado del estudio corresponde al puntaje de la prueba de matemática del examen nacional estandarizado administrado el 7 y 8 de noviembre de 2017. También, analizamos los efectos en lenguaje para explorar potenciales efectos indirectos en esta disciplina. El examen nacional estandarizado se ha llevado a cabo anualmente desde 1998 con todos los alumnos de cuarto grado de primaria al final del año académico, y es ampliamente utilizado para monitorear los resultados educativos. Estas pruebas son importantes para los docentes y los directores porque están vinculadas a incentivos monetarios, y bajos puntajes pueden activar medidas administrativas, incluyendo visitas a las escuelas y la introducción de cambios en cómo se gestionan estas escuelas. También administramos pruebas de matemática y lenguaje como parte del estudio en la línea de base, en la línea media (después de 4 meses de exposición) y en la línea final (después de 7 meses de exposición).

También recogimos datos de los alumnos respecto del autoconcepto en matemática, la motivación intrínseca en matemática, la preferencia por tener clases de matemática en el laboratorio de computación (en vez de en el aula regular), tener una mentalidad de crecimiento y la preferencia por el trabajo en equipo. Estos datos primarios sobre las percepciones de los alumnos fueron complementados con datos secundarios de un cuestionario incluido en el examen nacional estandarizado que explora el autoconcepto de los alumnos en matemática y si los alumnos sienten ansiedad en relación con las pruebas de matemática, las notas y los deberes. Por último, analizamos los datos registrados por la plataforma ConectaIdeas para documentar cómo fueron utilizadas las computadoras en la enseñanza de matemática.

3.4 Aleatorización y equilibrio

En esta sección, analizamos si la aleatorización generó grupos similares de tratamiento y control. Con ese fin, el Cuadro 2 presenta medias de los puntajes en la línea de base para los grupos de tratamiento y control (en las columnas (1) y (2), respectivamente). A su vez, la columna (3) presenta las diferencias estimadas entre los grupos de tratamiento y control al incluirse efectos fijos por escuela, y la columna (4) presenta el tamaño de la muestra en cada variable analizada. No hay diferencias significativas en los puntajes promedios para lenguaje entre el grupo de tratamiento y control. Sin embargo, los alumnos del grupo tratamiento tuvieron resultados en promedio más bajos que los alumnos del grupo control, en 0,08

desviaciones estándar en la prueba de matemática (esta diferencia es significativa al 10%). A pesar de que se trata de una diferencia modesta, estos resultados proporcionan una motivación adicional para controlar por el rendimiento académico en la línea de base cuando se estiman los efectos en este área.

El Cuadro 2 también presenta estadísticas para las características de los alumnos construido utilizando datos del cuestionario administrado junto con el examen nacional estandarizado en noviembre de 2017. Los resultados indican que la composición de los grupos de tratamiento y control son similares. Las diferencias en las características analizadas son pequeñas y sólo estadísticamente significativas al 10% para la educación de la madre.⁸

4. Resultados principales

¿Mejoró ConectaIdeas el aprendizaje de los alumnos? Para responder a esta pregunta, el Cuadro 3 presenta los efectos del programa en los resultados académicos de matemática medidos en el examen nacional estandarizado de 2017. Los resultados indican que el programa ConectaIdeas generó grandes mejoras en el aprendizaje de matemática. En nuestra primera especificación, la cual no controla por los resultados de línea de base en matemática, el efecto estimado es de 0,22 desviaciones estándar. En nuestra especificación preferida, la cual controla por los resultados de línea de base en matemática, el efecto es ligeramente mayor (0,27 desviaciones estándar). En cualquiera de los dos casos, los efectos estimados son estadísticamente significativos a nivel del 1%.

Aunque el programa se centró exclusivamente en matemática, podría haber generado efectos en lenguaje. Por ejemplo, el programa podría haber motivado a los alumnos a estudiar más en general, o podría haber llevado a emplear el tiempo de estudio de lenguaje en el estudio de matemática. Sin embargo, los resultados indican que el programa no afectó los resultados en lenguaje.

Los efectos documentados en matemática parecen grandes no sólo cuando se comparan con los de otras evaluaciones educativas llevadas a cabo en Chile (como se señaló en la introducción) sino también cuando se compara con referencias habituales de análisis de políticas educativas. Una de estas referencias se relaciona con las brechas de aprendizaje entre alumnos de diferentes niveles socioeconómicos. En particular, los alumnos cuyas madres terminaron la escuela secundaria tienen mejores resultados que aquellos cuyas madres no acabaron este nivel en 0,51 desviaciones estándar en el examen estandarizado nacional de matemática. Por lo tanto, ConectaIdeas podría cerrar aproximadamente el 50% de esta brecha de aprendizaje (0,27/0,51). Una segunda referencia utilizada habitualmente

⁸ La principal muestra en el estudio incluye a los alumnos que participaron en el examen de línea de base de matemática y en el examen de matemática estandarizado nacional. Por consiguiente, ésta es la muestra que se analiza cuando se exploran las diferencias en las características de los alumnos en el Cuadro 2.

involucra comparar los efectos con la progresión habitual en el aprendizaje que los alumnos experimentan en un año. Desafortunadamente, no tenemos datos de Chile acerca de cuánto mejoran los alumnos sus resultados académicos en matemática en un año. Sin embargo, Hill *et al.* (2008) documentan que los alumnos de cuarto de primaria en Estados Unidos mejoran su aprendizaje en 0,52 desviaciones estándar en un año. Suponiendo que la progresión académica de los alumnos en Chile es similar a la de Estados Unidos, podemos afirmar que los alumnos que participaron en ConectaIdeas progresaron aproximadamente un 50% más que sus contrapartes en el grupo de control (0,27/0,52).

El Cuadro 4 analiza si ConectaIdeas generó efectos heterogéneos en submuestras de estudiantes definidos por género, educación de la madre y rendimiento académico en la línea de base. El análisis sobre estos efectos heterogéneos se centra en matemática dado que es la materia objetivo del programa. Para empezar, los efectos son ligeramente mayores para los niños que para las niñas (0,29 versus 0,24 desviaciones estándar), pero esta diferencia en efectos no es estadísticamente significativa. Cuando se exploran los efectos por educación de la madre, observamos que éstos son casi iguales (0,28 versus 0,29 desviaciones estándar). Este patrón de efectos similares en diferentes submuestras también está presente cuando dividimos la muestra por resultados académicos en la línea de base. Es decir, los efectos en los alumnos con puntajes por debajo de la mediana en la prueba de matemática de la línea de base son idénticos a los de aquellos que tuvieron un puntaje inicial por encima de la mediana. En conclusión, estos resultados indican que las diferentes subpoblaciones de alumnos definidas por género, educación de la madre y resultados académicos en la línea de base experimentaron similares efectos positivos de ConectaIdeas.

5. Los mecanismos

5.1 Evidencia sobre el uso de la plataforma

ConectaIdeas generó efectos grandes en el aprendizaje de matemática pero, ¿por qué es efectivo el programa? ¿Se debe a que los alumnos están utilizando la plataforma intensivamente? Y, en ese caso, ¿este uso intensivo se produce principalmente en la escuela o también en la casa? Para responder a estas preguntas y comprender mejor los mecanismos que explican los efectos documentados, utilizamos datos longitudinales a nivel individual del uso de la plataforma. Nos centramos en el uso entre finales de marzo de 2017, cuando la intervención comenzó, y el 6 de noviembre de 2017, justo antes del examen nacional estandarizado.

Los datos de registro muestran que la plataforma se utilizó intensivamente, y que los alumnos la utilizaron en promedio por 27 horas. Además, el alumno promedio estuvo conectado a la plataforma 43 días (de aproximadamente 210 días durante el período de implementación) y la utilizó cada vez durante 39 minutos. La utilización se concentró

principalmente en la escuela, ya que representó el 98% del total de uso de la plataforma.⁹ Estos resultados también son corroborados por los Gráficos 3 y 4, que muestran que el uso de la plataforma se concentró fundamentalmente de lunes a viernes y durante las horas del día en que las escuelas estaban abiertas. Es decir, que el uso de la plataforma en la escuela jugó un rol central en esta intervención.

Si las estrategias de gamificación de ConectaIdeas generaron un uso intensivo de la plataforma en la escuela, ¿por qué no produjeron un mayor uso en el hogar? El bajo uso en el hogar se puede considerar un problema de diseño. Dado que los alumnos que utilizan la plataforma ConectaIdeas sólo pueden trabajar en ejercicios que les han sido asignados por sus docentes, los alumnos no pueden usar la plataforma durante el fin de semana para practicar a menos que les hayan asignado ejercicios. Y, de hecho, los coordinadores de laboratorio y los docentes no recibieron instrucciones para asignar ejercicios a los alumnos para que practicasen fuera de la escuela. Por lo tanto, en implementaciones futuras de ConectaIdeas sería interesante asignar ejercicios a los alumnos como deberes para la casa con el fin de explorar si el uso de la plataforma en el hogar también puede contribuir a mejorar el aprendizaje.

Utilizando datos de la plataforma también documentamos que había escasa heterogeneidad entre diferentes escuelas en términos del número de sesiones de tecnología llevadas a cabo. La escuela promedio implementó 49 sesiones de tecnología de matemática y las escuelas en los percentiles 10 y 90 tuvieron 42 y 55 sesiones, respectivamente. Por último, el Gráfico 5 presenta la distribución del uso de la plataforma por mes. Los resultados señalan que el uso de la plataforma fue similar a lo largo del año escolar (una vez que se excluyen los meses con uso incompleto).¹⁰

5.2 Efectos en los resultados no académicos

La introducción de elementos de juego en la enseñanza puede generar efectos en una gama de resultados más allá de los resultados académicos en matemática y lenguaje. Por consiguiente, analizamos los efectos en los resultados no académicos utilizando datos de nuestra encuesta final a los alumnos y a partir del cuestionario que se administró junto con el examen nacional estandarizado.

En particular, construimos índices para medir las dimensiones relevantes. Por ejemplo, construimos un índice de motivación intrínseca en matemática utilizando nueve

⁹ Clasificamos el uso como “en la escuela” si tuvo lugar en los días en que las escuelas estaban abiertas (es decir, días de la semana que no eran días festivos ni vacaciones) y entre los horarios en que las escuelas estaban abiertas.

¹⁰ El uso en marzo y noviembre es mínimo porque estos meses se incluyen sólo parcialmente en el período de análisis. El uso fue bajo en abril porque las escuelas estaban comenzando el programa, y en julio debido a las vacaciones de invierno de dos semanas.

preguntas incluidas en la encuesta final a los alumnos. Estas preguntas se obtuvieron de traducir al castellano las preguntas de esta dimensión incluidas en los exámenes de matemática de cuarto grado de primaria (IEA 2014) de los Estudios Internacionales de Tendencias en Matemática y Ciencias (TIMSS). Todas las preguntas se transforman en indicadores que son iguales a uno si el alumno está de acuerdo con una afirmación, estandarizados utilizando la desviación media y estándar, promediado en diferentes ítems para la misma dimensión y más tarde estandarizado nuevamente para una interpretación más fácil.¹¹ El Cuadro 5 presenta los efectos estimados obtenidos aplicando regresiones de estos índices en un indicador de tratamiento y efectos fijos por escuela (es decir, estimando la ecuación 1).

Los resultados señalan efectos positivos estadísticamente significativos en dos dimensiones que están bien alineados con las expectativas que se podrían tener dada la naturaleza del programa.¹² Para comenzar, la base de la gamificación involucra producir una experiencia de aprendizaje más atractiva y, encontramos que el 79% de los alumnos en el grupo de tratamiento declara preferir las clases de matemática en el laboratorio de computación en lugar de hacerlo en el aula tradicional. En cambio, sólo el 59% de los alumnos en el grupo de control declara preferir las clases de matemática en el laboratorio de computación. Esta diferencia se traduce en un efecto positivo de 0,40 desviaciones estándar en las preferencias de los alumnos de tener las clases de matemática en el laboratorio de computación. Además, una de las características de ConectaIdeas implicaba presentar publicidades personalizadas a los alumnos para motivar la adopción de una mentalidad de crecimiento, y documentamos un efecto positivo de 0,14 desviaciones estándar en este área.

En cambio, hay dos ámbitos en los que no encontramos efectos estadísticamente significativos, aunque algunos efectos se podrían haber previsto. El primero versa sobre la motivación intrínseca, es decir, el goce inherente en el aprendizaje de matemática. Dado que ConectaIdeas enfatiza la realización de ejercicios de matemática para aumentar los puntajes y tener mejores resultados en las competencias individuales y grupales, podría reducir la motivación intrínseca en esta materia. Sin embargo, no encontramos evidencia que apoye esta expectativa. De hecho, el efecto de la motivación intrínseca es positivo aunque no es estadísticamente significativo. El segundo efecto versa sobre el autoconcepto en matemática o la percepción que los alumnos tienen de sus propias habilidades para solucionar ejercicios de matemática. Dado que ConectaIdeas produjo grandes aumentos en el rendimiento académico en matemática, podríamos esperar efectos positivos en este ámbito, aunque, en realidad, no encontramos efectos estadísticamente significativos.

¹¹ El cuadro A.1 en el Apéndice presenta información detallada sobre las preguntas utilizadas para construir los indicadores no académicos.

¹² En esta sección todos los resultados que son mencionados como estadísticamente significativos se refieren al nivel del 5%.

A su vez, hay dos ámbitos en los que encontramos efectos estadísticamente significativos que se pueden considerar como no deseables. En particular, encontramos efectos positivos estadísticamente significativos de 0,13 desviaciones estándar en la ansiedad asociada a estudiar matemática, que se podría vincular con las comparaciones sociales y las competencias individuales y grupales incluidas en ConectaIdeas. También documentamos efectos negativos estadísticamente significativos de 0,21 desviaciones estándar en las preferencias por el trabajo en equipo. Este resultado puede ser sorprendente si se considera que ConectaIdeas promovió la colaboración dentro del aula por medio de la organización de competencias entre secciones de estudiantes. Una explicación potencial de este resultado inesperado es que algunos alumnos podrían haber notado las desventajas de trabajar en equipo (por ejemplo, el vínculo más débil entre el propio esfuerzo y los resultados finales) luego de haber participado varias veces en las competencias de equipo de ConectaIdeas.

En el análisis descrito, medimos los efectos en seis indicadores diferentes. Dado que estamos realizando múltiples tests estadísticos, la probabilidad de encontrar efectos estadísticamente significativos en algún indicador se incrementa más allá del nivel de confianza elegido. Para abordar este problema, seguimos el procedimiento sugerido por Benjamini *et al.* (2006) e implementado en Stata por Anderson (2008) para producir valores- q , que se pueden interpretar como análogos a los valores- p una vez que tenemos en cuenta que estamos realizando múltiples tests estadísticos.¹³ Los efectos en las preferencias por tener las clases de matemática en el laboratorio de computación, en la mentalidad de crecimiento, en la ansiedad asociada a estudiar matemática y en las preferencias por el trabajo en equipo siguen siendo estadísticamente significativos después de este ajuste.

Por último, exploramos la heterogeneidad de los efectos de ConectaIdeas en los resultados no académicos estimando los efectos de tratamiento para las submuestras definidas por género, educación de la madre, y resultados académicos en la línea de base (Cuadro 6). Los efectos documentados más consistentes son el aumento de la preferencia por tener clases de matemática en el laboratorio de computación, que son estadísticamente significativos para todas las submuestras. Los efectos positivos en la mentalidad de crecimiento y ansiedad asociada a estudiar matemática, así como los efectos negativos en las preferencias por el trabajo en equipo, que se documentaron para la muestra en su conjunto, también se observan para cada submuestra, aunque los efectos son estadísticamente significativos sólo en algunos casos. Por otro lado, algunos efectos positivos estadísticamente significativos en el autoconcepto en matemática y la motivación intrínseca ahora están presentes en algunas submuestras. Sin embargo, se debería ser cauto cuando se interpretan estos coeficientes debido a la reducción de potencia estadística asociada con el hecho de centrarse en

¹³ Los valores- q son el valor crítico más bajo al que se rechaza una hipótesis nula cuando se controla por la tasa de descubrimientos falsos. La tasa de descubrimientos falsos es la proporción prevista de hipótesis nulas rechazadas que, de hecho, son verdaderas. Para estimar los valores- q , tenemos que especificar una familia de valores- p relacionados. En este ejercicio, consideramos que los 6 valores- p presentados en el Cuadro 5 pertenecen a la misma familia.

submuestras y también debido a la mayor probabilidad de encontrar algún efecto estadísticamente significativo cuando se realizan múltiples tests estadísticos.

Para presentar una evaluación más concluyente sobre este tema, analizamos si los efectos estimados son estadísticamente significativos para cada resultado y en cada dimensión. Por ejemplo, producimos el valor-p para la diferencia en los efectos de ConectaIdeas en el autoconcepto en matemática entre niños y niñas. En este análisis, encontramos sólo tres casos donde los efectos eran estadísticamente significativos y diferentes en distintas submuestras. En particular, observamos que los efectos en las preferencias por el trabajo en equipo son más negativos para las niñas que para los niños, que los efectos en las preferencias por las clases de matemática en el laboratorio de computación son más positivos en los alumnos cuyas madres tienen niveles más bajos de educación, y los efectos en las preferencias por las clases de matemática en el laboratorio de computación son mayores en los alumnos con resultados académicos en la línea de base más bajos. Sin embargo, una vez que tenemos en cuenta que estamos aplicando pruebas de múltiples hipótesis para este análisis (siguiendo el procedimiento descrito en Benjamini *et al.*, 2006), no encontramos diferencias estadísticamente significativas en los efectos de ConectaIdeas en ninguno de los resultados-dimensiones analizados.¹⁴

6. Pruebas de robustez

6.1 Estimación de errores estándar

Para estimar los efectos de ConectaIdeas en el rendimiento académico de los estudiantes, estimamos los errores estándar con clusters a nivel de sección. Dado que tenemos 48 clusters, podría ocurrir que las fórmulas utilizadas para calcular los errores estándar pueden generar estimaciones conservadoras. Para lidiar con este problema, calculamos los errores estándar siguiendo varias especificaciones diferentes. Para comenzar, calculamos los errores estándar siguiendo el procedimiento sugerido por Cameron y Miller (2015) con bootstrapping a nivel de sección. Además, calculamos los errores estándar con clusters a nivel de escuela (para nuestra especificación base y también siguiendo el procedimiento sugerido por Cameron y Miller 2015). Además, calculamos los errores estándar promediando los resultados a nivel de sección, ajustados por el rendimiento en la línea de base y realizando una regresión a este nivel que incluye los efectos fijos por escuela (como sugieren Bertrand, Duflo y Mullainathan 2004). Por último, seguimos la metodología sugerida en Ibragimov y Muller (2010), donde el principal modelo se estima por separado para cada escuela, y luego se realiza una prueba-t sobre la distribución de los coeficientes de tratamiento estimados. En todos los casos, las conclusiones presentadas en nuestro análisis principal siguen inalteradas: encontramos

¹⁴ En este ejercicio, consideramos que todos los valores-p de los efectos diferenciales en diferentes dimensiones para los resultados no académicos pertenecen a la misma familia. Nótese que hay 6 resultados no académicos para cada una de las 3 dimensiones para un total de 18 valores-p.

efectos estadísticamente significativos a nivel del 1% para los resultados de matemática y sin efectos significativos para lenguaje (Cuadro A.2 en el Apéndice).

6.2. Efectos de derrame en el grupo de control

Es posible que la introducción del programa pueda haber influido en la conducta de los docentes y los alumnos en las secciones de control. Por ejemplo, los docentes en las secciones de control pueden haber realizado un mayor esfuerzo para compensar por los efectos potenciales del programa, o puede que se hayan sentido desalentados porque no recibieron más apoyo. En cualquiera de estos dos casos, la diferencia en los resultados académicos entre alumnos en las secciones de tratamiento y las de control no reflejarían los efectos causales de ConectaIdeas. A pesar de que estos efectos de derrame dentro de las escuelas pueden jugar un rol importante en ciertas intervenciones (por ejemplo, en aquellas que involucran el suministro de información) puede que en este contexto esta posibilidad esté atenuada. Esto se debe a que el equipo de implementación controlaba la plataforma ConectaIdeas y no permitía a los alumnos en las secciones de control tener acceso a ella.

Aún así, exploramos empíricamente esta posibilidad generando estimaciones de diferencias en diferencias de los efectos del programa en las secciones de control. Para ese fin, utilizamos datos del examen nacional estandarizado y mantuvimos las secciones de control en las 24 escuelas que participaron en la evaluación experimental, así como en las secciones “A” y “B” en un grupo de escuelas de comparación. Estas escuelas de comparación incluyen aquellas que estaban situadas en Santiago, que fueron clasificadas en las dos categorías socioeconómicas inferiores en 2017, que tenían dos o tres secciones en 2017 y que participaron en el examen nacional estandarizado en 2016 y 2017. Para crear un mejor grupo de comparación, estimamos el puntaje de propensión de ser una escuela que participó en el experimento en 2017 utilizando la edad del alumno, el género, la asistencia a una escuela de preprimaria y la educación de la madre. El Cuadro A.3 en el Apéndice presenta los efectos derrame estimados en las secciones de control. Los resultados sugieren que el programa no afectó el aprendizaje de los alumnos en matemática o lenguaje en las secciones de control. Estos resultados están presentes tanto en la especificación base como cuando se utiliza propensity reweighting. Por consiguiente, estos hallazgos sugieren que la diferencia en resultados académicos entre los alumnos en las secciones de tratamiento y los de las secciones de control reflejan los efectos causales de ConectaIdeas.

6.3. Efectos medidos utilizando pruebas académicas del estudio

Además de los efectos estimados utilizando nuestra principal medida de resultados (el examen nacional estandarizado) también medimos los efectos en matemática y lenguaje usando pruebas desarrolladas y administradas por empresas especializadas contratadas como

parte del estudio. Estas pruebas de línea media y línea final se aplicaron como un respaldo en caso de que no llegáramos a tener acceso a los datos de los exámenes nacionales estandarizados.

El Cuadro 4.A en el Apéndice, presenta estos resultados. El panel A muestra que el programa generó efectos de 0,18 desviaciones estándar en el aprendizaje de matemática en la línea media (en nuestra especificación preferida). Estos resultados parecerían ser consistentes con el efecto de 0,27 desviaciones estándar en el examen nacional estandarizado, teniendo en cuenta que la prueba de la línea media se administró cuatro meses después de que comenzara el programa, y el examen nacional estandarizado se administró siete meses después de comenzar el programa. En cambio, los resultados de la línea final aplicado como parte del estudio muestran efectos más pequeños de 0,13 desviaciones estándar en los resultados de matemática.

Hay diversos motivos potenciales por los que documentamos efectos más bajos en la prueba de la línea final en comparación con la prueba de la línea media y el examen nacional estandarizado. Los motivos se refieren a cómo se desarrollaron estas pruebas. Para la prueba de la línea media, la empresa especializada en pruebas encuestó a los docentes en las escuelas del estudio y evaluó a los alumnos en los temas del programa de estudio que fueron cubiertos durante el primer semestre en estas escuelas. En cambio, para la prueba de la línea final, la empresa especializada en pruebas utilizó un examen estándar normalmente administrado en las escuelas interesadas en documentar cuánto están aprendiendo sus alumnos. El grupo de escuelas que contrata estos servicios tiende a incluir escuelas privadas de alto rendimiento en comparación con la población nacional de alumnos en Chile. Los temas cubiertos en cuarto grado de primaria en estas escuelas pueden ser bastante diferentes de los cubiertos en las escuelas que participan en este estudio (fundamentalmente escuelas públicas de bajo rendimiento). Por lo tanto, se podría esperar que los temas que se enfatizaron en la intervención (y que estaban cubiertos en el examen nacional estandarizado) no estaban cubiertos de forma adecuada en la prueba de la línea final aplicado por la empresa especializada en pruebas.¹⁵

¹⁵ En línea con esta explicación, hay un fuerte solapamiento entre los objetivos de aprendizaje que los alumnos practicaron en la plataforma y aquellos que fueron evaluados en la línea media, pero esto no fue lo que ocurrió en la línea final. De los seis objetivos de aprendizaje principales en términos de práctica de los alumnos en la plataforma, la prueba final no evaluó dos de ellos e incluyó sólo una pregunta para otros dos objetivos de aprendizaje. Además, la prueba final incluyó varios ítems relacionados con los ocho objetivos de aprendizaje que representaban la menor práctica en la plataforma. En cambio, estos problemas de falta de cobertura se minimizan con el examen nacional estandarizado, el cual emplea un sistema bajo el cual diferentes alumnos solucionan diferentes subconjuntos de preguntas (en total se incluyen 175 ítems). Una pieza final de evidencia que sugiere que los resultados del examen de la línea final pueden ser menos fiables en comparación con el examen nacional estandarizado es que la correlación entre el examen de la línea final y el examen de la línea de base fue menor que la correlación entre el examen nacional estandarizado y el examen de la línea de base (0,59 versus 0,68). Se encuentra un patrón similar cuando se analizan las correlaciones con el examen de la línea intermedia (0,66 y 0,76, respectivamente).

6.4. Estimaciones no experimentales

Una preocupación potencial es que nuestra evaluación experimental podría haber incidido en la calidad de la implementación del programa. Por lo tanto, es importante medir la efectividad del programa bajo circunstancias más normales. Por ello, en esta sección proporcionamos evidencia no experimental respecto a los efectos de ConectaIdeas utilizando datos longitudinales a nivel de las escuelas de los exámenes nacionales estandarizados en el período antes de la evaluación experimental de 2017.¹⁶ En particular, explotamos la implementación de ConectaIdeas en 11 escuelas en la comuna de Lo Prado en Santiago, desde 2011 hasta 2016, utilizando un enfoque de diferencias en diferencias.¹⁷

Aunque los elementos centrales de ConectaIdeas han permanecido inalterados a lo largo de los años, hay algunas diferencias entre la versión implementada en el período 2011-2016 y la versión de 2017 evaluada experimentalmente. En primer lugar, se esperaba que los alumnos de cuarto grado que participaron en ConectaIdeas a lo largo de 2011-2016 utilizaran la plataforma semanalmente durante 135 minutos, comparados con los 180 minutos esperados de uso para la versión evaluada experimentalmente.¹⁸ En segundo lugar, en el período 2011 a 2014, los alumnos de tercero básico también fueron expuestos al programa, con una sesión de 45 minutos a la semana. Por último, la plataforma sufrió algunas modificaciones menores y ajustes a lo largo de los años.

Construimos un grupo de comparación centrándonos en las escuelas urbanas situadas en Santiago que están clasificadas en las tres últimas categorías socioeconómicas, que participaron consistentemente en el examen nacional estandarizado durante el período y que tenían una matriculación en cuarto grado superior a los ocho alumnos en todos los años. Luego, utilizando características a nivel de las escuelas del período anterior al programa estimamos el puntaje de propensión de recibir el programa como función de los puntajes promedio de matemática y lenguaje, la proporción de alumnos que asistieron a preprimaria y la proporción de alumnos cuyas madres eran indígenas. De esta forma, utilizamos propensity score reweighting para generar un grupo de comparación más adecuado.¹⁹

El Cuadro 7 presenta estadísticas que describen las características promedio de las escuelas de tratamiento y comparación en el período anterior a la implementación del programa. A primera vista, estos dos grupos de escuelas parecen bastante diferentes (columna (3)). Por ejemplo, las escuelas tratamiento tienen resultados inferiores a las escuelas de

¹⁶ Araya (2018) evaluó el impacto de ConectaIdeas utilizando una comparación de niveles de aprendizaje entre antes y después del comienzo de la implementación del programa.

¹⁷ Tres de esas escuelas no participaron en el programa en 2013 y 2014.

¹⁸ Los alumnos de cuarto básico que participaron en ConectaIdeas durante 2011-2016 tuvieron aproximadamente 45 minutos más de clases de matemática a la semana, mientras que aquellos que participaron en la evaluación experimental tuvieron aproximadamente 90 minutos más de clases de matemática.

¹⁹ La ponderación para el grupo de control está dada por: $\frac{pscore}{1-pscore}$, mientras que la ponderación para las escuelas del programa es igual a 1.

comparación tanto en matemática como en lenguaje. En la columna (4) limitamos la muestra a aquellas escuelas para las que hay un solapamiento en los puntajes de propensión (es decir, el puntaje de propensión se sitúa entre el puntaje mínimo y el máximo del grupo tratamiento). Al aplicar esta limitación, las diferencias entre las muestras de tratamiento y comparación disminuyen considerablemente. Por último, en la columna (5) mostramos diferencias entre las escuelas del tratamiento y de comparación después de aplicar propensity score reweighting.

Estimamos el siguiente modelo para evaluar el efecto de ConectaIdeas en el rendimiento académico:

$$y_{ist} = \alpha + \beta \text{Tratamiento}_{st} + \tau_t + \phi_s + X_{ist} + \varepsilon_{ist},$$

Donde Tratamiento_{st} es igual a 1 para la escuela s que participó en el programa en el año t y 0 de lo contrario, τ_t son efectos fijos por año, ϕ_s son efectos fijos por escuela y X_{ist} son características de los alumnos, como el género, un indicador para la asistencia a preprimaria, educación de la madre y tamaño de la cohorte de la sección. Por último, β , es el parámetro de interés y estima el efecto promedio de participar en ConectaIdeas en los puntajes de matemática o lenguaje. Los errores estándar se calculan con clusters a nivel de escuela.

Las columnas (1) y (2) en el Cuadro 8 presentan estimaciones de diferencias en diferencias utilizando toda la muestra, mientras que las columnas (3) y (4) limitan la muestra al common support y emplean propensity score reweighting. Los resultados señalan que ConectaIdeas mejoró el aprendizaje en matemática entre 0,19 y 0,22 desviaciones estándar. En general, no se encuentran efectos en lenguaje (solo en una de las cuatro especificaciones se encuentra un efecto positivo el cual es pequeño y significativo al 10 por ciento). Estos resultados son similares a los obtenidos en nuestro diseño experimental, aunque ligeramente inferiores. En general, estos resultados refuerzan la conclusión principal respecto a la efectividad de ConectaIdeas de mejorar el aprendizaje de matemática.

7. Conclusión

Llevamos a cabo una evaluación experimental en 24 escuelas primarias de bajo rendimiento académico en Santiago de Chile para evaluar la efectividad del programa ConectaIdeas. Este programa incorpora varias estrategias de gamificación para aumentar la motivación de los alumnos y de esta forma incrementar el aprendizaje en matemática. Los resultados indican que el programa aumentó el aprendizaje en matemática en 0,27 desviaciones estándar según lo medido en el examen nacional estandarizado chileno. No se observaron importantes efectos en lenguaje. El programa también influyó en los resultados no académicos. Por el

lado positivo, el programa aumentó la preferencia de los alumnos por utilizar computadoras en las clases de matemática y promovió la idea de que la inteligencia se puede mejorar cuando uno estudia intensamente. Por el lado negativo, el programa generó aumentos en la ansiedad asociada con estudiar matemática y redujo las preferencias de los alumnos por trabajar en equipo.

Es importante tener en cuenta algunas características del programa cuando se piensa en extrapolar los resultados a otros contextos o a la hora de incrementar su escala. El programa se implementó en escuelas con alumnos de bajo nivel socioeconómico y rendimiento académico en Santiago, Chile. Por consiguiente, los efectos pueden ser diferentes en otros contextos donde los alumnos tienen niveles más altos de aprendizaje. Por otro lado, para implementar todas las características del programa descritas en este estudio, las escuelas necesitan un laboratorio de computación con una conexión a internet fiable. Esta condición no se cumplirá en numerosas escuelas en países en desarrollo, sobre todo en zonas rurales. Sin embargo, es posible diseñar e implementar una versión del programa que no requerirá acceso a internet (esta versión no incluiría las competencias grupales dado que estas funcionalidades requieren acceso a internet). Por último, el programa evaluado aquí incluía una implementación por parte de coordinadores externos. Para facilitar la escalabilidad y reducir los costos, es relevante probar una versión del programa que puedan implementar los docentes de aula. Estos docentes deberían recibir capacitación y apoyo pedagógico para facilitar el desarrollo de las habilidades necesarias para usar la plataforma de aprendizaje. Además, podría ser fructífero explorar algunas medidas complementarias, como proporcionar una pequeña paga a los docentes para compensarlos por el esfuerzo requerido para adoptar estas nuevas prácticas en su trabajo.

Más allá de estos aspectos relacionados con la extrapolación y la escalabilidad, los importantes impactos académicos documentados en este estudio sugieren que utilizar la gamificación en educación puede ser una estrategia prometedora para aumentar el aprendizaje de los alumnos. Esto es especialmente relevante si se piensa que Chile, al igual que muchos otros países en el mundo en vías de desarrollo y en el mundo desarrollado, están buscando estrategias efectivas para mejorar los niveles de aprendizaje y reducir las brechas de aprendizaje. Además, la utilización de los enfoques de gamificación podría aprovechar las inversiones considerables que numerosos países han hecho para aumentar el acceso a las computadoras y a internet en las escuelas, así como el creciente acceso a dispositivos conectados a internet en los hogares (Arias Ortiz y Cristia 2014).

Sin embargo, se necesita más investigación para comprender mejor cómo las diferentes estrategias de gamificación incluidas en ConectaIdeas (y otras estrategias innovadoras) pueden influir en la participación y el aprendizaje de los alumnos. Además, teniendo en cuenta los efectos de una mayor ansiedad relacionado con el aprendizaje de matemática y menores preferencias a trabajar en equipo, es importante seguir explorando la robustez de estos resultados y maneras potenciales de eliminar estos efectos no deseados.

Referencias

- Alsawaier, R. 2018. "The Effect of Gamification on Motivation and Engagement." *The International Journal of Information and Learning Technology* 35: 56–79.
- Anderson, M. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103: 1481–1495.
- Araya, R. 2018. "Teacher Training, Mentoring or Performance Support Systems?" En *International Conference on Applied Human Factors and Ergonomics*. Springer.
- Arias Ortiz, E., y J. Cristia. 2014. "El BID y la Tecnología para Mejorar el Aprendizaje: Cómo Promover Programas Efectivos?" Nota Técnica del BID No. 670. Banco Interamericano de Desarrollo, Washington, DC.
- Banerjee, A., S. Cole, E. Duflo, y L. Linden. 2007. "Remedying Education: Evidence From Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122: 1235–1264.
- Barata, G., S. Gama, J. Jorge, y D. Gonçalves. 2013. "Engaging Engineering Students with Gamification." En *International Conference on Games and Virtual Worlds for Serious Applications*. IEEE Education Society (Septiembre).
- Bassi, M., C. Meghir, y A. Reynoso. 2016. "Education Quality and Teaching Practices." Documento de Trabajo NBER No. 22719. National Bureau of Economic Research, Cambridge, MA.
- Bellei, C. 2009. "Does Lengthening the School Day Increase Students' Academic Achievement? Results from a Natural Experiment in Chile." *Economics of Education Review* 28: 629–640.
- Benjamini, Y., A. Krieger, y D. Yekutieli. 2006 "Adaptive Linear Step-Up Procedures that Control the False Discovery Rate." *Biometrika* 93: 491–507.
- Bertrand, M., E. Duflo, y S. Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119: 249–275.

- Bettinger, E. 2012. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." *Review of Economics and Statistics* 94: 686–698.
- Buchanan, L. 2018. "The Hottest Education Startup in the U.S. Is a \$700 Million Company Built by a Guatemalan Engineer in Pittsburgh." *Inc.* (Diciembre 2018). Disponible en: <https://www.inc.com/leigh-buchanan/duolingo-700-million-language-learning-startup-pittsburgh-2018-surge-cities.html>
- Cameron, A., y D. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50: 317–372.
- Cialdini, R., L. Demaine, B. Sagarin, D. Barrett, K. Rhoads, y P. Winter. 2006. "Managing Social Norms for Persuasive Impact." *Social Influence* 1: 3–15.
- Claro, S., D. Paunesku, y C. Dweck. 2016. "Growth Mindset Tempers the Effects of Poverty on Academic Achievement." *Proceedings of the National Academy of Sciences* 113: 8664–8668.
- DellaVigna, S., y D. Pope. 2018. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy* 126: 2410–2456.
- Denny, P. 2013. "The Effect of Virtual Achievements on Student Engagement." En *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.
- Domínguez, A., J. Saenz-de-Navarette, L. de-Marcos, L. Fernández-Sanz, C. Pagés, y J.J. Martínez-Herraíz. 2013. "Gamifying Learning Experiences: Practical Implications and Outcomes." *Computers & Education* 63: 380–392.
- Dweck, C. 2006. *Mindset: the New Psychology of Success*. Random House.
- Dynarski, M., R. Agodini, S. Heaviside, T. Novak, N. Carey, y L. Campuzano. 2007. "Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort." Report to Congress. Publication NCEE 2007-4005. U.S. Department of Education, Washington, DC (Marzo).

- Edwards, K., D. DeVries, y J. Snyder. 1972. "Games and Teams: A Winning Combination." *Simulation & Games* 3: 247–269.
- Fryer, R. 2011. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." Documento de Trabajo NBER No. 16850. National Bureau of Economic Research, Cambridge, MA.
- Hanus, M., y J. Fox. 2015. "Assessing the Effects of Gamification in the Classroom: A Longitudinal Study on Intrinsic Motivation, Social Comparison, Satisfaction, Effort, and Academic Performance." *Computers & Education* 80: 152–161.
- Hill, C., H. Bloom, A. Black, y M. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2: 172–177.
- Ibragimov, R., y U. Müller. 2010. "t-Statistic Based Correlation and Heterogeneity Robust Inference." *Journal of Business & Economic Statistics* 28: 453–468.
- International Education Association (IEA). 2014. "Trends in International Mathematics and Science Study 2015." Cuestionario de Estudiante, 4to Grado.
- Lai, F., L. Zhang, X. Hu, Q. Qu, Y. Shi, Y. Qiao, M. Boswell, y S. Rozelle. 2013. "Computer Assisted Learning as Extracurricular Tutor? Evidence from a Randomised Experiment in Rural Boarding Schools in Shaanxi." *Journal of Development Effectiveness* 5: 208–231.
- Lai, F., X. Hu, L. Zhang, y Q. Qu. 2015. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing." *Economics of Education Review* 47: 34–48.
- Li, T., T. Han, L. Zhang, y S. Rozelle. 2014. "Encouraging Classroom Peer Interactions: Evidence from Chinese Migrant Schools." *Journal of Public Economics* 111: 29–45.
- Lister, M. 2015. "Gamification: The Effect on Student Motivation and Performance at the Post-Secondary Level." *Issues and Trends in Educational Technology* 3: 1–22.

- Malamud, O., S. Cueto, J. Cristia, y D. Beuermann. 2019. "Do Children Benefit from Internet Access? Experimental Evidence from Peru." *Journal of Development Economics* 138: 41–56.
- Markopoulos, A., A. Fragkou, P. Kasidiaris, y J. Davim. 2015. "Gamification in Engineering Education and Professional Training." *International Journal of Mechanical Engineering Education* 43: 118–131.
- McDaniel, R., R. Lindgren, y J. Friskics. 2012. "Using Badges for Shaping Interactions in Online Learning Environments." En *IEEE International Professional Communication Conference*.
- Mekler, E., F. Bruhlmann, A. Tuch, y K. Opwis. 2017. "Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance." *Computers in Human Behavior* 71: 525–534.
- Mo, D., L. Zhang, R. Luo, Q. Qu, W. Huang, J. Wang, Y. Qiao, M. Boswell, y S. Rozelle. 2014. "Integrating Computer-Assisted Learning into a Regular Curriculum: Evidence from a Randomised Experiment in Rural Schools in Shaanxi." *Journal of Development Effectiveness* 6: 300–323.
- Muralidharan, K., A. Singh, y A. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109: 1426–1460.
- Organisation for Economic Co-operation and Development (OECD). 2017. "PISA 2015 Results (Volume V): Collaborative Problem Solving." OECD Publishing, París.
- Rutherford, T., G. Farkas, G. Duncan, y M. Burchinal. 2014. "A Randomized Trial of an Elementary School Mathematics Software Intervention: Spatial-Temporal Math." *Journal of Research on Educational Effectiveness* 7: 358–383.
- Slavin, R. 2010. "Co-operative Learning: What Makes Groupwork Work?" En *The Nature of Learning: Using Research to Inspire Practice*, editado por H. Dumont, D. Istance, y F. Benavides. París: OCDE.

- Weiner, B. 1990. "History of Motivational Research in Education." *Journal of Educational Psychology* 82: 616–622.
- Wijekumar, K., J. Hitchcock, H. Turner, P. Lei, y K. Peck. 2009. "A Multisite Cluster Randomized Trial of the Effects of CompassLearning Odyssey® Math on the Math Achievement of Selected Grade 4 Students in the Mid-Atlantic Region." Reporte Final, NCEE 2009-4068. National Center for Education Evaluation and Regional Assistance.

Cuadro 1: Construcción de la muestra - Año previo al tratamiento (2016)

	Todas las escuelas (1)	Restricciones adicionales de la muestra			
		En Santiago (2)	Estatus socioeconómico	Dos o más secciones	Participaron en la evaluación
			bajo (3)	(4)	(5)
<i>Puntajes (normalizados con estudiantes de todo el país)</i>					
Matemática	0,00	0,07	-0,37	-0,25	-0,68
Lenguaje	0,00	0,02	-0,38	-0,32	-0,60
<i>Características de los alumnos</i>					
Mujer	0,50	0,50	0,48	0,50	0,48
Edad	9,61	9,63	9,70	9,68	9,83
Asistencia a preprimaria	0,98	0,99	0,99	0,99	0,98
Madre con educación secundaria	0,72	0,76	0,52	0,55	0,48
Padre en casa	0,60	0,61	0,54	0,55	0,50
Madre indígena	0,11	0,07	0,11	0,12	0,11
<i>Número de alumnos</i>	<i>217.034</i>	<i>84.972</i>	<i>27.048</i>	<i>14.675</i>	<i>1.366</i>
<i>Características de la escuela</i>					
Matriculación en cuarto grado	29,35	47,90	37,41	67,94	56,92
Rural	0,39	0,07	0,14	0,06	0,04
Estatus socioeconómico bajo	0,64	0,41	1,00	1,00	0,96
<i>Número de escuelas</i>	<i>7.395</i>	<i>1.774</i>	<i>723</i>	<i>216</i>	<i>24</i>

Nota: Este cuadro presenta medias de diferentes grupos de escuelas. Se utilizan datos de los exámenes nacionales estandarizados de cuarto grado de 2016. Las puntajes de las pruebas han sido normalizadas restando la media y dividiendo por la desviación estándar de la muestra completa, la cual incluye a todos los alumnos en el país. La columna (1), presenta la media de los alumnos de todas las escuelas del país, la columna (2) restringe la muestra a quienes se encuentran en la zona metropolitana de Santiago, la columna (3) restringe aún más la muestra a escuelas en las dos categorías inferiores (de las cinco) en términos de estatus socioeconómico, la columna (4) restringe aún más la muestra a las escuelas con dos o más aulas, y la columna (5) restringe aún más la muestra a las escuelas que participan en el estudio.

Cuadro 2: Balance en la línea de base - Año de tratamiento (2017)

	Tratamiento (1)	Control (2)	Diferencia (3)	N (4)
<i>Puntajes (normalizados con el grupo de control)</i>				
Matemática	-0.09	0,00	-0,08 (0.05)*	1.089
Lenguaje	-0.05	0,00	-0,04 (0.06)	1.057
<i>Características de los alumnos</i>				
Mujer	0.48	0.47	0,02 (0.02)	1.089
Edad	9.74	9.76	-0.02 (0.03)	1055
Asistencia a preprimaria	0.98	0.99	-0.01 (0.01)	788
Madre con educación secundaria	0,48	0,52	-0,04 (0.02)*	873
Padre en casa	0.53	0.54	-0,01 (0.02)	873
Madre indígena	0,14	0,11	0,02 (0.02)	872
Tiene Internet	0.81	0.82	-0,02 (0.02)	840

Nota: Este cuadro presenta las medias y las diferencias estimadas entre los grupos de tratamiento y de control. Los resultados de los puntajes de la línea de base son construidos utilizando datos del examen implementado como parte del estudio. Los resultados de las características de los alumnos son construidos utilizando datos del examen nacional estandarizado de 2017. La muestra utilizada para analizar los puntajes de la prueba de matemática y las características de los alumnos incluye a los alumnos que participaron en el examen de la línea de base de matemática y en el examen nacional estandarizado de matemática de 2017. La muestra utilizada para analizar los puntajes de la prueba de lenguaje incluye a los alumnos que participaron en el examen de la línea de base de lenguaje y en el examen nacional estandarizado de lenguaje de 2017. Todos los puntajes de las pruebas han sido normalizados restando la media y dividiendo por la desviación estándar del grupo de control. Las columnas (1) y (2) presentan las medias para los grupos de tratamiento y control, respectivamente. La columna (3) presenta las diferencias entre los grupos de tratamiento y control, controlando por efectos fijos por escuela. La columna (4) presenta el número de alumnos en cada muestra. Los errores estándar, reportados en paréntesis, están estimados con clusters a nivel de sección. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Cuadro 3: Efectos en el rendimiento académico

	Tratamiento	Control	Diferencia	Diferencia ajustada	N
	(1)	(2)	(3)	(4)	(5)
Matemática	-0,40	-0,60	0.22 (0.05)***	0.27 (0.04)***	1.089
Lenguaje	-0.61	-0,58	-0.04 (0.05)	-0.01 (0.04)	1.057

Nota: Este cuadro presenta los efectos estimados de ConectaIdeas en el rendimiento académico en matemática y lenguaje. Se utilizan datos del examen nacional estandarizado de 2017. Las etiquetas en las filas corresponden a las variables dependientes. Las columnas (1) y (2) presentan las medias para los grupos de tratamiento y control, respectivamente. La columna (3) presenta las diferencias entre los grupos de tratamiento y control, controlando por efectos fijos por escuela. La columna (4) presenta diferencias ajustadas controlando por efectos fijos por escuela y por el valor de la variable de resultado en la línea de base. La columna (5) presenta el número de alumnos en cada muestra. La muestra utilizada para analizar los puntajes de la prueba de matemática incluye a los alumnos que participaron en el examen de la línea de base de matemática y en el examen nacional estandarizado de matemática de 2017. La muestra utilizada para analizar los puntajes de la prueba de lenguaje incluye a los alumnos que participaron en el examen de la línea de base de lenguaje y en el examen nacional estandarizado de lenguaje de 2017. Los puntajes de las pruebas han sido normalizados restando la media y dividiendo por la desviación estándar de la muestra, la cual incluye a todos los alumnos del país. Los errores estándar, reportados en paréntesis, están estimados con clusters a nivel de sección. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Cuadro 4: Efectos heterogéneos en el rendimiento académico

	Género		Madre con educación secundaria		Puntaje en línea de base	
	Niños	Niñas	Sí	No	Bajo	Alto
	(1)	(2)	(3)	(4)	(5)	(6)
Matemática	0.29 (0.05)***	0.24 (0.06)***	0,28 (0.07)***	0,29 (0.05)***	0,26 (0.06)***	0,26 (0.05)***
<i>N</i>	571	518	434	439	510	579
Lenguaje	0.03 (0.05)	-0.05 (0.05)	0,05 (0.06)	-0,03 (0.05)	-0,03 (0.06)	0,05 (0.05)
<i>N</i>	565	492	420	427	530	527

Nota: Este cuadro presenta los efectos estimados de ConectaIdeas en el rendimiento académico en matemática y lenguaje para distintas submuestras de alumnos. Se utilizan datos del examen nacional estandarizado de 2017. Cada celda corresponde a una regresión. Las etiquetas en las filas corresponden a las variables dependientes. Los títulos de las columnas indican la muestra incluida en la estimación. La muestra utilizada para analizar los puntajes de la prueba de matemática incluye a los alumnos que participaron en el examen de la línea de base de matemática y en el examen nacional estandarizado de matemática de 2017. La muestra utilizada para analizar los puntajes de la prueba de lenguaje incluye a los alumnos que participaron en el examen de la línea de base de lenguaje y en el examen nacional estandarizado de lenguaje de 2017. Los puntajes de las pruebas han sido normalizados restando la media y dividiendo por la desviación estándar de la muestra, la cual incluye a todos los alumnos del país. Los errores estándar, reportados en paréntesis, están estimados con clusters a nivel de sección. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Cuadro 5: Efectos en indicadores no académicos

	Tratamiento (1)	Control (2)	Diferencia (3)	N (4)
Prefiere laboratorio para matemática	0.42	0,00	0.40 (0.06)***	787
Mentalidad de crecimiento	0,10	0,00	0,14 (0.05)***	790
Motivación intrínseca en matemática	0.09	0,00	0,10 (0.08)	797
Autoconcepto en matemática	0,08	0,00	0,12 (0.07)	706
Ansiedad con matemática	0,15	0,00	0,13 (0.05)**	883
Prefiere trabajar en equipo	-0.20	0.00	-0.21 (0.06)***	827

Nota: Este cuadro presenta los efectos estimados de ConectaIdeas en indicadores no académicos. Las etiquetas en las filas corresponden a las variables dependientes. Las columnas (1) y (2) presentan las medias para los grupos de tratamiento y control, respectivamente. La columna (3) presenta las diferencias entre los grupos de tratamiento y control controlando por efectos fijos por escuela. Los errores estándar, reportados en paréntesis, están estimados con clusters a nivel de sección. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Cuadro 6: Efectos heterogéneos en indicadores no académicos

	Género		Madre con educación secundaria		Puntaje en línea de base	
	Niños	Niñas	Sí	No	Bajo	Alto
	(1)	(2)	(3)	(4)	(5)	(6)
Prefiere laboratorio para matemática	0.49 (0.07)***	0.35 (0.07)***	0.28 (0.09)***	0.62 (0.09)***	0.53 (0.09)***	0.29 (0.08)***
Mentalidad de crecimiento	0.19 (0.06)***	0.11 (0.07)	0.12 (0.06)*	0.26 (0.08)***	0.13 (0.09)	0.20 (0.06)***
Motivación intrínseca en matemática	0.06 (0.09)	0.17 (0.11)	0.16 (0.12)	0.07 (0.11)	-0.01 (0.13)	0.20 (0.10)**
Autoconcepto en matemática	0.01 (0.08)	0.24 (0.13)*	0.20 (0.11)*	0.12 (0.09)	0.11 (0.12)	0.23 (0.07)***
Ansiedad con matemática	0.08 (0.06)	0.20 (0.07)***	0.07 (0.08)	0.13 (0.07)*	0.09 (0.07)	0.13 (0.08)*
Prefiere trabajar en equipo	-0.10 (0.08)	-0.34 (0.07)***	-0.26 (0.11)**	-0.09 (0.07)	-0.10 (0.07)	-0.30 (0.08)***

Nota: Este cuadro presenta los efectos estimados de ConectaIdeas en indicadores no académicos para distintas submuestras de alumnos. Se utilizan datos de los exámenes nacionales estandarizados de cuarto básico 2017. Cada celda corresponde a una regresión. Las etiquetas en las filas corresponden a las variables dependientes. Los títulos de las columnas indican la muestra incluida en la estimación. Los errores estándar, reportados en paréntesis, están estimados con clusters a nivel de sección. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Cuadro 7: Estimación no experimental - Balance durante el período previo al tratamiento (2005-2010)

	Tratamiento	Comparación	Diferencia		
			Sin ajustes	Common support	Common support y propensity score reweighting
	(1)	(2)	(3)	(4)	(5)
Puntuaciones de la prueba (Normalizadas con las de todo el país)					
Matemática	-0.47	-0.20	-0.27 (0.04)***	-0.04 (0.04)	0.04 (0.05)
Lenguaje	-0.47	-0.21	-0.27 (0.04)***	-0.04 (0.04)	0.04 (0.04)
Características de los alumnos					
Mujer	0.49	0.48	0.01 (0.01)	0.02 (0.01)	0.03 (0.02)*
Edad	10.60	10.81	-0.21 (0.20)	-0.43 (0.20)**	-0.48 (0.21)**
Asistencia a preprimaria	0.81	0.79	0.02 (0.02)	0.01 (0.02)	0.01 (0.02)
Madre con educación secundaria	0.36	0.52	-0.16 (0.02)***	-0.01 (0.02)	0.03 (0.03)
Padre en casa	0.22	0.22	-0.00 (0.01)	0.00 (0.01)	0.01 (0.01)
Madre indígena	0.15	0.08	0.07 (0.02)***	0.03 (0.02)**	0.00 (0.02)
Tiene Internet	0.69	0.66	0.03 (0.06)	0.04 (0.06)	0.01 (0.06)
Número de escuelas	11	999	1,010	429	429

Nota: Este cuadro presenta las medias y las diferencias estimadas entre los grupos de tratamiento y de comparación utilizados en el análisis de diferencias en diferencias. Se utilizan datos de los exámenes nacionales estandarizados de 2005 a 2010. Los puntajes de las pruebas han sido normalizados restando la media y dividiendo por la desviación estándar de la muestra, la cual incluye a todos los alumnos del país, para cada año. Las columnas (1) y (2) presentan las medias para los grupos de tratamiento y de comparación, respectivamente. Las columnas (3) a (5) presentan las diferencias entre los grupos de tratamiento y de comparación. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Cuadro 8: Estimación no experimental - Efectos en el rendimiento académico

	Diferencias en diferencias (DID)		DID + Propensity score reweighting	
	(1)	(2)	(3)	(4)
Matemática	0.22*** (0.06)	0.22*** (0.06)	0.19*** (0.07)	0.19*** (0.07)
Lenguaje	0.06 (0.04)	0,06 (0.04)*	0.02 (0.04)	0.03 (0.05)
<i>Número de alumnos</i>	655,072	655,072	239,312	239,312
<i>Número de escuelas</i>	1,010	1,010	429	429

Nota: Este cuadro presenta los efectos estimados de diferencia en diferencia de ConectaIdeas en el rendimiento académico en matemática y lenguaje. Se utilizan datos de los exámenes nacionales estandarizados de 2005 a 2010. La unidad de observación es la combinación de una escuela y un año. Cada celda corresponde a una regresión. Cada regresión incluye un indicador de tratamiento, efectos fijos por escuela y efectos fijos por año. Las etiquetas en las filas corresponden a las variables dependientes. Las columnas (1) y (2) incluyen escuelas urbanas en la zona metropolitana de Santiago que se encuentran en las tres categorías inferiores (de las cinco) en términos de estatus socioeconómico y que tuvieron una matriculación mínima de ocho estudiantes en cuarto grado en el período 2005-2010. Las columnas (3) y (4) limitan aún más la muestra a escuelas para las que hay un solapamiento en los puntajes de propensión. Los resultados de las regresiones presentados en las columnas (2) y (4) también incluyen controles que varían en el tiempo. Los puntajes de las pruebas han sido normalizados restando la media y dividiendo por la desviación estándar de la muestra, la cual incluye a todos los alumnos del país, para cada año. El número de escuelas y alumnos presentados en el cuadro corresponde a aquellos incluidos para estimar los efectos sobre los puntajes de las pruebas de matemática. A su vez, 654.365 alumnos de 1.010 escuelas están incluidos para estimar los efectos en los puntajes de las pruebas de lenguaje presentadas en las columnas (1) y (2), y 239.182 estudiantes en 429 escuelas están incluidos para estimar los efectos sobre los puntajes de las pruebas de lenguaje presentadas en las columnas (3) y (4). La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Gráfico 1: Captura de pantalla del panel de control de los alumnos



Gráfico 2: Captura de pantalla del juego del torneo

Conectado a sala 6

The game board is a 10x6 grid of numbers. The numbers are arranged in a specific pattern, with some cells containing icons like bees or coins. The board is divided into several colored regions: red (outer border), yellow (inner border), blue (central area), green (inner area), and pink (central area). The numbers are as follows:

1	2	3	4	5	6	7	8	9	10	11
43	44	45	46	47	48	49	50	51	52	53
42	79	80	81	82	83	84	85	86	87	54
41	78	107	108	109	110	111	112	113	88	55
40	77	106	127	128	129	130	131	114	89	56
39	76	105	126	139	140	141	132	115	90	57
38	75	104	125	138	143	142	133	125	91	58
37	74	103	124	137	136	135	134	117	92	59
36	73	102	123	122	121	120	119	118	93	60
35	72	101	100	99	98	97	96	95	94	61
34	71	70	69	68	67	66	65	64	63	62
33	32	31	30	29	28	27	26	25	24	23
22										

Jugada N°1

Student A
Finalizar turno
Selecciona una ficha...

Turno N°1
66 s

Oro:0 **Err:0**

Student B
Esperando turno...

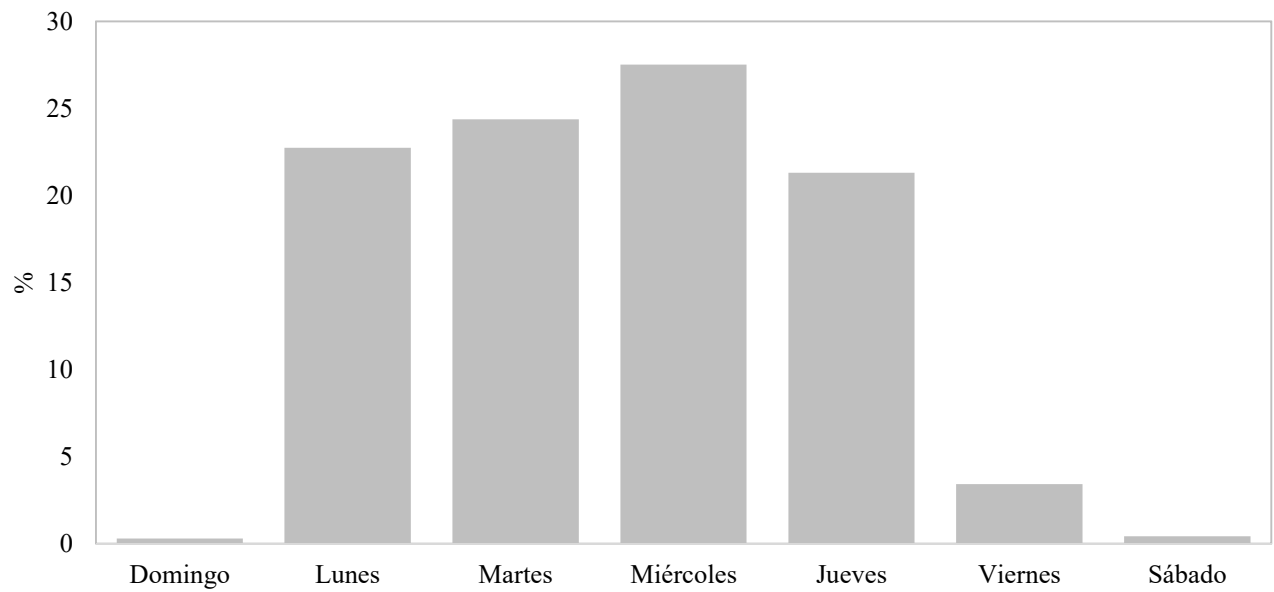
Turno N°0
-

Oro:0 **Err:0**

The number that when subtracted from 30 the result is 18

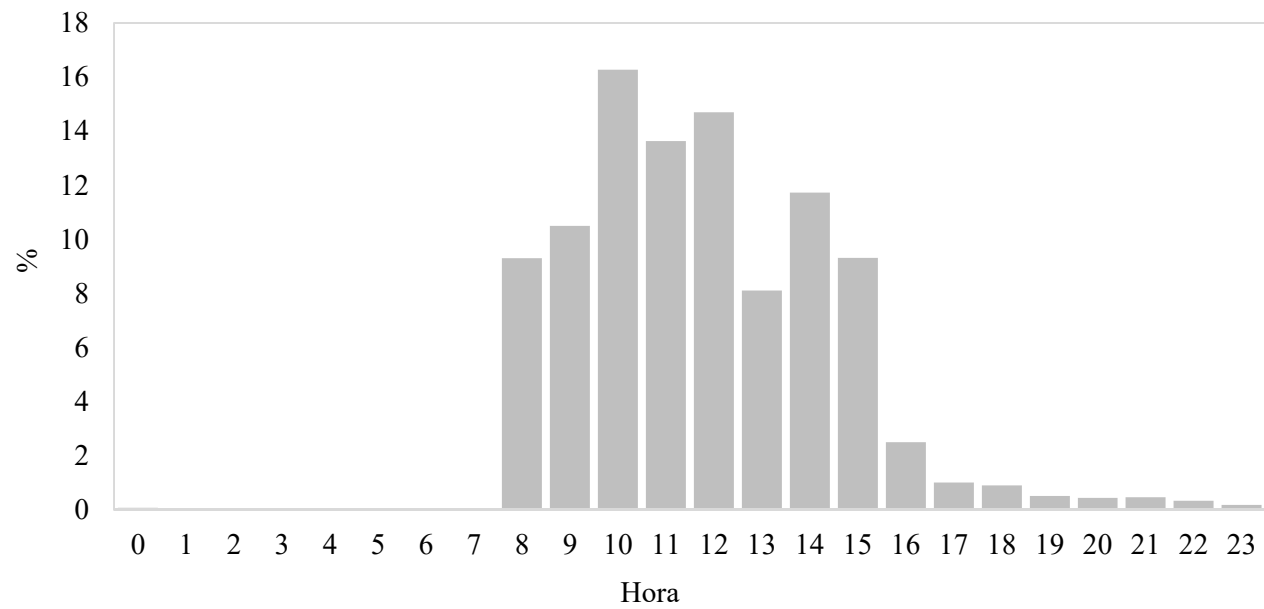
Contar de nuevo...

Gráfico 3: Uso de la plataforma por día de la semana



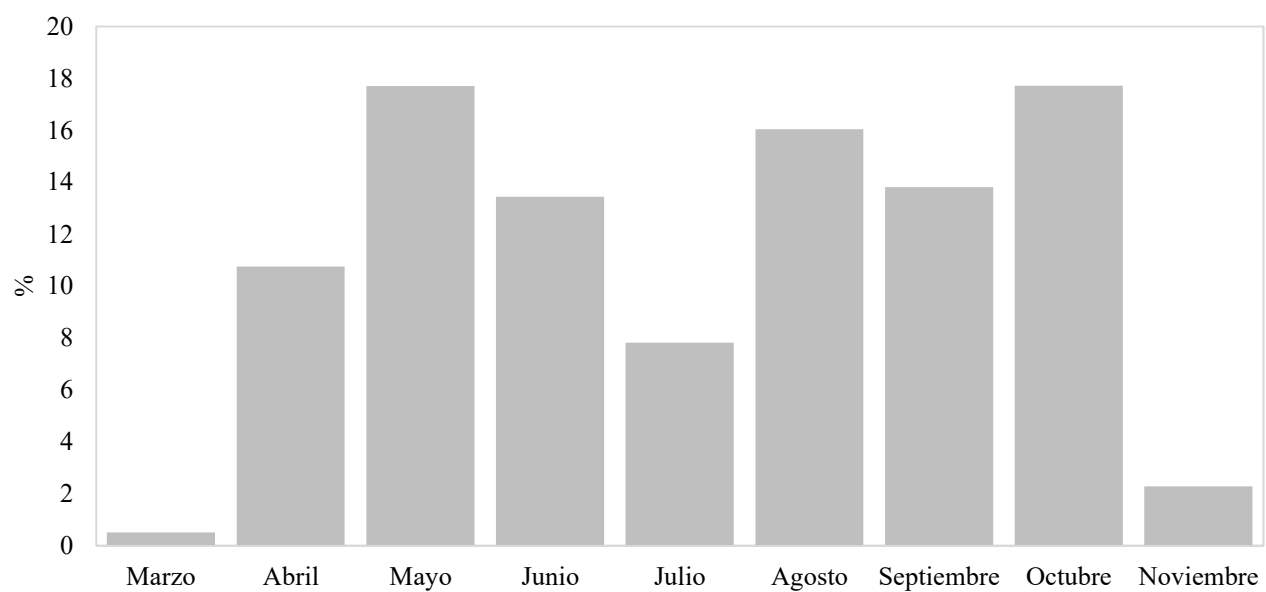
Nota: Este gráfico presenta la distribución del uso de la plataforma por día de la semana para el período entre el 28 de marzo, 2017 y el 6 de noviembre, 2017.

Gráfico 4: Uso de la plataforma por hora del día



Nota: Este gráfico presenta la distribución del uso de la plataforma por hora del día para el período entre el 28 de marzo, 2017 y el 6 de noviembre, 2017.

Gráfico 5: Uso de la plataforma por mes



Nota: Este gráfico presenta la distribución del uso de la plataforma por mes para el período entre el 28 de marzo, 2017 y el 6 de noviembre, 2017.

Cuadro A.1: Construcción de indicadores no académicos

Indicador	Pregunta	Fuente
Prefiere laboratorio para matemática	Prefiero tener clases de matemática en el laboratorio que en el aula	Creado para este estudio
Mentalidad de crecimiento	La inteligencia es algo que no se puede cambiar*	Claro (2016)
	Puedes aprender cosas nuevas, pero no puedes cambiar tu inteligencia*	Claro (2016)
	Mis padres dicen que soy capaz de aprender	SIMCE
Motivación intrínseca en matemática	Me gusta aprender matemática	TIMSS
	Desearía no tener que estudiar matemática*	TIMSS
	Matemática es aburrida*	TIMSS
	Aprendo muchas cosas interesantes en matemática	TIMSS
	Me gusta matemática	TIMSS
	Me gusta cualquier trabajo escolar que involucre números	TIMSS
	Me gusta resolver problemas matemáticos	TIMSS
	Espero con ansias las lecciones de matemática	TIMSS
	Matemática es mi asignatura favorita	TIMSS
Autoconcepto en matemática	Normalmente me va bien en matemática	TIMSS
	Matemática es más difícil para mí que para muchos de mis compañeros de clase*	TIMSS
	Simplemente no soy bueno en matemática*	TIMSS
	Normalmente me va bien en los exámenes de matemática	SIMCE
	Matemática es fácil para mí	SIMCE
Ansiedad con matemática	Mi profesor me dice que soy bueno en matemática	SIMCE
	Me temo que las preguntas de matemática son demasiado difíciles para mí	SIMCE
	Me preocupa tener malas notas en matemática	SIMCE
	Me pongo nervioso antes de los exámenes de matemática	SIMCE
Prefiere trabajar en equipo	Me pongo nervioso cuando no entiendo la tarea de matemática	SIMCE
	Prefiero trabajar en equipo que trabajar solo	PISA
	Me parece que los equipos toman mejores decisiones que los individuos	PISA
	Me parece que el trabajo en equipo aumenta mi propia eficiencia	PISA
	Me gusta cooperar con mis compañeros	PISA

Nota: Este cuadro presenta las preguntas utilizadas para la construcción de los indicadores no académicos y la fuente de cada pregunta. En general, los indicadores no académicos son construidos como índices que combinan diferentes preguntas. Las preguntas cuyo origen es el Estudio Internacional de Tendencias en Matemática y Ciencias (TIMSS), el Informe del Programa Internacional para la Evaluación de Estudiantes (PISA), y Claro (2016), o que están marcados como "Creado para este estudio", se incluyeron en el cuestionario final aplicado como parte de este estudio a los alumnos que participan en la evaluación experimental. En cambio, las preguntas cuyo origen es el Sistema de Medición de la Calidad de la Educación (SIMCE) se incluyeron en el cuestionario que fue administrado como parte del examen estandarizado nacional chileno de 2017. La fuente TIMSS incluye las preguntas 16 y 18 del cuestionario para alumnos de cuarto básico del TIMSS 2015 (IEA 2014). La fuente PISA incluye las preguntas del índice "valoración del trabajo en equipo" desarrollado para el informe PISA 2015 sobre "Resolución colaborativa de problemas" (OCDE 2017). La fuente Claro (2016) incluye las preguntas del estudio de Claro, Paunesku, y Dweck (2016) el cual buscó medir la mentalidad de crecimiento en estudiantes chilenos. La fuente "Creado para este estudio" se refiere a preguntas desarrolladas sin utilizar una referencia externa. Todas las preguntas son escalas tipo Likert de 4 puntos, con la única excepción de la pregunta "Preferiría tener clases de matemática en el laboratorio que en el aula", que es una pregunta de sí/no. Los elementos que están marcados con un asterisco (*) deben invertirse para calcular el índice de un resultado. Por ejemplo, las respuestas a la pregunta "Matemática es más difícil para mí que para muchos de mis compañeros de clase" se invierten, de modo que los valores más altos correspondan a autoconceptos en matemática superiores.

Cuadro A.2: Pruebas de robustez - Estimación de errores estándar

	Nivel sección		Nivel escuela			
	Cluster estándar	Bootstrapping	Cluster estándar	Bootstrapping	Bertrand et al. (2004)	Ibragimov y Muller (2010)
	(1)	(2)	(3)	(4)	(5)	(6)
Matemática	0,27 (0,04)***	0,27 (0,06)***	0,27 (0,06)***	0,27 (0,06)***	0,28 (0,06)***	0,27 (0,07)***
Lenguaje	-0,01 (0,04)	-0,01 (0,06)	-0,01 (0,06)	-0,01 (0,06)	0,00 (0,05)	0,03 (0,06)

Nota: Este cuadro presenta los efectos estimados de ConectaIdeas en los puntajes en matemática y lenguaje utilizando diferentes metodologías para calcular los errores estándar. Se utilizan datos del examen nacional estandarizado de 2017. Las etiquetas en las filas corresponden a las variables dependientes. Las columnas (1) a (4) presentan resultados de regresiones que utilizan nuestra especificación principal (diferencias ajustadas) controlando por efectos fijos por escuela y por el valor de la variable de resultado en la línea de base. Los errores estándar de las columnas (1) y (3) se calculan incluyendo clusters a nivel de sección y de escuela, respectivamente. Los errores estándar de las columnas (2) y (4) se calculan utilizando la metodología sugerida por Cameron y Miller 2015 y realizando bootstrapping a nivel de sección y de escuela, respectivamente. La columna (5), emplea la estrategia propuesta por Bertrand, Duflo y Mullainathan (2004), donde los puntajes (ajustados por los niveles en la línea de base) son promediados a nivel de sección, y luego se corre una regresión usando nuestra especificación preferida utilizando los datos agregados. Por último, la columna (6) sigue a Ibragimov y Muller (2010), donde el principal modelo se estima por separado para cada escuela, y luego llevamos a cabo una prueba-t sobre la distribución de los coeficientes de tratamiento estimados. La muestra utilizada para analizar los puntajes de la prueba de matemática incluye a los alumnos que participaron en el examen de la línea de base de matemática y en el examen nacional estandarizado de matemática de 2017. La muestra utilizada para analizar los puntajes de la prueba de lenguaje incluye a los alumnos que participaron en el examen de la línea de base de lenguaje y en el examen nacional estandarizado de lenguaje de 2017. Los puntajes de las pruebas han sido normalizados restando la media y dividiendo por la desviación estándar de la muestra, la cual incluye a todos los alumnos del país. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Cuadro A.3: Explorando efectos en las secciones de control

	(1)	(2)	(3)	(4)
Matemática	0.12 (0.09)	0.11 (0.10)	0.01 (0.10)	0.01 (0.10)
Lenguaje	-0.01 (0.08)	-0.03 (0.08)	-0.06 (0.08)	-0.06 (0.08)
Controles	N	S	N	S
Propensity score reweighting	N	N	S	S
<i>Número de alumnos</i>	<i>23,040</i>	<i>22,895</i>	<i>17,883</i>	<i>17,786</i>
<i>Número de escuelas</i>	<i>218</i>	<i>218</i>	<i>178</i>	<i>178</i>

Nota: Este cuadro presenta los efectos estimados de diferencia en diferencia de ConectaIdeas en los puntajes de matemática y lenguaje de los estudiantes de las secciones de control. Se utilizan datos de los exámenes nacionales estandarizados de 2016 y 2017. Cada celda corresponde a una regresión. Cada regresión incluye un indicador de tratamiento, características del alumno (edad, género, asistió a una preprimaria, madre completó la escuela secundaria), efectos fijos por escuela y efectos fijos por año. Las etiquetas en las filas corresponden a las variables dependientes. Las columnas (1) y (2) incluyen escuelas urbanas en la zona metropolitana de Santiago que se encuentran en las dos categorías inferiores (de las cinco) en términos de estatus socioeconómico y que tuvieron una o dos secciones en 2016. Las columnas (3) y (4) limitan aún más la muestra a escuelas para las que hay un solapamiento en los puntajes de propensión estimados con base en características de 2016. Los resultados de regresión presentados en las columnas (2) y (4) también incluyen controles que varían en el tiempo. Los puntajes de las pruebas han sido normalizados restando la media y dividiendo por la desviación estándar de la muestra, la cual incluye a todos los alumnos del país, para cada año. Los errores estándar, reportados en paréntesis, están estimados con clusters a nivel de escuela. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.

Cuadro A.4: Efectos en el rendimiento académico - Exámenes alternativos

	Tratamiento	Control	Diferencia	Diferencia ajustada	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Línea media - Exámenes del estudio					
Matemática	0,06	0,00	0,11 (0.06)*	0,18 (0.05)***	903
Lenguaje	-0,07	0,00	-0,05 (0.06)	-0,03 (0.05)	844
Panel B: Línea final - Exámenes del estudio					
Matemática	0,07	0,00	0,09 (0.05)	0,13 (0.05)***	923
Lenguaje	-0,02	0,00	-0,03 (0.05)	0,00 (0.04)	882

Nota: Este cuadro presenta los efectos estimados de ConectaIdeas en el rendimiento académico en matemática y lenguaje usando datos de exámenes implementados como parte de este estudio. El panel A reporta resultados generados usando datos del examen de línea media. El panel B reporta resultados generados usando datos del examen de línea final. Las etiquetas en las filas corresponden a las variables dependientes. Las columnas (1) y (2) presentan las medias para los grupos de tratamiento y control, respectivamente. La columna (3), presenta diferencias estimadas controlando por efectos fijos por escuela. La columna (4) presenta diferencias ajustadas controlando por efectos fijos por escuela y por el valor de la variable de resultado en la línea de base. La columna (5) presenta el número de alumnos en cada muestra. La muestra utilizada para analizar los puntajes de la prueba de matemática y las características de los alumnos incluye a los alumnos que participaron en el examen de la línea de base de matemática y en el examen nacional estandarizado de matemática de 2017. La muestra utilizada para analizar los puntajes de la prueba de lenguaje incluye a los alumnos que participaron en el examen de la línea de base de lenguaje y en el examen nacional estandarizado de lenguaje de 2017. Los puntajes de las pruebas han sido normalizados restando la media y dividiendo por la desviación estándar del grupo de control. Los errores estándar, reportados en paréntesis, están estimados con clusters a nivel de sección. La significancia estadística al 1, 5 y 10 por ciento es indicada por ***, ** y *, respectivamente.