

Lo que me pareció convincente en un nuevo ensayo aleatorio de Khan Academy en una intervención en matemáticas

24 de agosto de 2026 · Por Sal Khan, fundador y CEO de Khan Academy

La semana pasada, el equipo de investigación y efectividad me envió un nuevo documento de trabajo: "One Click Away: AI Tutoring with Khanmigo in a Two-Year School Experiment", de Philip Oreopoulos y Nina Low ("A un clic de distancia: Tutoría con IA mediante Khanmigo en un experimento escolar de dos años"). Un aviso de antemano: Khan Academy no estuvo involucrado en el diseño e implementación de este estudio. Es una investigación independiente y justamente por eso vale la pena compartirla. Esta es mi lectura sobre lo que es interesante del experimento.

Ajustando por el efecto implícito de un año completo de participación activa, un segundo análisis de los alumnos que tenían un rendimiento escolar inferior al del nivel de curso muestra que la mejora fue de 0,14 desviaciones estándar (DE). Ese es el número en el que me quiero enfocar.

El estudio resalta que las 0,14 DE están dentro del rango de evidencia experimental para un software de práctica de matemáticas (el cual es un rango entre 0,05 y 0,20 DE). Este resultado es inferior al [del tamaño del gran impacto de 0,45 que vimos en un estudio realizado por uno de los mismos investigadores en la India](#), pero acá el contexto fue muy diferente.

Primero, la población de estudiantes de este experimento tiene un rendimiento inferior al del nivel de curso en matemáticas. Segundo, el grupo de control incluía el uso de otras herramientas tecnológicas educativas. Por lo tanto, la desviación estándar de 0,14 puede

interpretarse como un beneficio de Khan Academy y Khanmigo por sobre y más allá de las herramientas tecnológicas educativas alternativas, y como ayuda para los estudiantes que comienzan el año con un rendimiento inferior al del nivel.

La configuración

Los investigadores de la Universidad de Toronto y la consultora Charles River Associates llevaron a cabo este estudio como un verdadero ensayo controlado aleatorizado (ECA) en 18 middle schools (en EE. UU., esto corresponde a los grados 6-8) durante 2 años (años escolares 2024-2025 y 2025-2026) en el condado de Hamilton, Tennessee.

Un ECA suele considerarse el estándar de oro en evidencia. En vez de comparar escuelas o estudiantes que eligen participar, un ECA asigna aleatoriamente quién participa en el programa y quién no, lo cual es la mejor manera que hay para aislar causa y efecto. Los estudiantes del estudio tenían un rendimiento por debajo de su nivel y participaban diariamente en un programa de intervención de matemáticas. Más de un cuarto de la muestra del estudio tenía una clasificación educativa especial. Una crítica hacia las herramientas educativas tecnológicas es que casi nunca funcionan para este grupo de estudiantes. Esa es otra razón por la que me interesaba saber qué podría concluir este experimento. ¿Podemos ayudar a que este grupo de estudiantes logre avances?

Por qué el grupo de control importa tanto acá

Este es el detalle que más importa al interpretar el número 0,14 DE: el grupo de control fue denominado por los investigadores “como si nada ocurriera”. Esto significa que el grupo de control continuó usando lo que había antes del estudio. No es que se compare Khan Academy con “ninguna tecnología” o con una sala de clases tradicional. Para la mayoría de estas escuelas, “como si nada ocurriera” significaba usar otras herramientas en línea como Waggle, IXL, Zearn y DeltaMath. De las 18 escuelas, 14 usaron al menos una de estas herramientas en su condición de control.

Entonces, este estudio no está cuestionando “si Khan Academy le gana a hacer nada”; está comparando Khan Academy con lo que ya estaba implementado en el distrito durante el periodo de la intervención, que, en la mayoría de las escuelas, incluía otros productos de tecnología educativa. En mi opinión, esa es una barra mucho más alta y la

que realmente importa para un distrito al momento de decidir en qué gastar su presupuesto. Escrito de otra forma, no se trata de un beneficio completo de Khan Academy frente a nada; se trata de sopesar el beneficio progresivo de Khan Academy por sobre las intervenciones existentes.

La cifra principal, y la que creo que es la real

Usando la comparación estándar y más conservadora (Intención de tratamiento: todos los asignados al grupo de control frente a todos los asignados a Khan Academy, independientemente de cuánto tiempo lo usaran realmente), el efecto combinado de dos años fue de aproximadamente 0,06 DE, con un 0,08 DE solo el segundo año una vez que la implementación estaba establecida. El primer año no mostró un efecto confiable, atribuido a un inicio lento: demoras en el registro de los estudiantes y un cambio en el sistema de evaluación a mitad de año.

0,14 DE

Pero esta es la cifra que creo que es la lectura más importante del impacto real del programa: 0,14 DE.

El estudio se basa en la regla de "una vez dentro, siempre dentro". Una vez que un estudiante participaba en el estudio, permanecía en él para siempre. Sin embargo, aproximadamente dos de cada cinco estudiantes completaron el programa de intervención durante el estudio. Aun así, se incluyeron en el cálculo, aunque ya no fueran parte. Cuando los investigadores aislaron solo a los estudiantes que permanecieron en el programa de intervención todo el segundo año, el grupo de tratamiento que utilizó Khan Academy obtuvo un progreso de 0,14 DE en comparación con el grupo de control. Yo diría que podría haber sido incluso mayor si los 2/5 de los estudiantes que más avanzaron y que completaron el programa se hubieran quedado en él y hubieran mantenido ese progreso.

¿Con qué concluye el estudio?

Debo aclarar que este hallazgo no es el principal que los autores destacan. El título del estudio es "One Click Away: AI Tutoring with Khanmigo in a Two-Year School Experiment" ("A un clic de distancia: Tutoría con IA mediante Khanmigo en un experimento escolar de dos años"). El estudio pretendía examinar los efectos de Khanmigo. Sin embargo, los estudiantes lo usaron con poca frecuencia, lo cual coincide con lo que hemos estado diciendo públicamente desde hace tiempo. El estudio atribuye el bajo uso de Khanmigo a barreras conductuales relacionadas con la búsqueda de ayuda, agravadas por una interfaz poco proactiva, en vez de alguna limitación de lo que Khanmigo en sí mismo podía hacer. Nosotros también hemos observado esto y es precisamente por eso que rediseñamos la experiencia en la sala de clases de manera que Khanmigo forme parte de la práctica en lugar de que los estudiantes tengan que abrirlo. Todos nuestros distritos escolares asociados usarán la experiencia rediseñada este año escolar.

¿Por qué estoy compartiendo esto?

Este es uno de los [muchos](#) estudios sobre Khan Academy y, al igual que en varios de ellos, no participamos en su diseño ni en su ejecución. Un riguroso ECA que muestra un efecto de 0,14 desviaciones estándar para estudiantes con un rendimiento inferior al del nivel y que tuvieron acceso continuo, medido en comparación con lo que el distrito ya había implementado en lugar de nada, en una población en la que es históricamente difícil producir un cambio, es un resultado realmente potente. Creo que las cifras principales de 0,06 y 0,08 en realidad le restan importancia.

Si trabajas en políticas educativas, tecnología educativa o simplemente quieres ver cómo es la evidencia rigurosa sobre tecnología educativa en la actualidad, puedes consultar el artículo completo aquí: <https://www.nber.org/papers/w35620>

