

JMNJE -Vol.4. N1. 104

Modelado matemático y estadístico del rendimiento académico en el aprendizaje de la trigonometría a nivel preuniversitario: un análisis comparativo entre el uso de Khan Academy y los métodos tradicionales

Mathematical and statistical modeling of academic performance in trigonometry learning at the pre-university level: a comparative analysis between the use of Khan Academy and traditional methods

Autores:

José Oswaldo Briones Calvache
Universidad Nacional de Chimborazo
Riobamba – Ecuador
jose.briones@unach.edu.ec
<https://orcid.org/0009-0002-8070-3605>

Pablo Ignacio Páliz Larrea
Universidad Nacional de Chimborazo
Riobamba – Ecuador
pablo.paliz@unach.edu.ec
<https://orcid.org/0009-0000-9860-7465>

Autor de correspondencia: José Oswaldo Briones Calvache, jose.briones@unach.edu.ec

Enviado: 10-05-2026
Aceptado: 23-08-2026

Revisado: 24-06-2026
Publicado: 24-08-2026



Cómo citar este artículo:

Briones Calvache, J. O., & Páliz Larrea, P. I. (2026). Modelado matemático y estadístico del rendimiento académico en el aprendizaje de la trigonometría a nivel preuniversitario: un análisis comparativo entre el uso de Khan Academy y los métodos tradicionales. *Journal of Multidisciplinary Novel Journeys & Explorations*, 4(1), 1-19. <https://doi.org/10.63688/7wybgx82>

© 2026; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons (<https://creativecommons.org/licenses/by/4.0>) que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea correctamente citada.



RESUMEN

Introducción: Khan Academy ofrece práctica secuenciada y retroalimentación inmediata que puede complementar la enseñanza de trigonometría, aunque su efecto depende de la integración pedagógica y de las características del contexto educativo.

Metodología: Se realizó un estudio cuantitativo comparativo posttest con grupos no equivalentes. Se analizaron 120 estudiantes preuniversitarios de cuatro paralelos, 60 en enseñanza tradicional y 60 con apoyo de Khan Academy. La comparación incluyó t de Welch, g de Hedges, bootstrap y Mann–Whitney; en los usuarios de la plataforma se ajustó una regresión exploratoria con frecuencia de uso, tiempo de estudio, desempeño y autorregulación.

Resultados: La media posttest fue 6,78 en la condición tradicional y 8,43 en Khan Academy, con una diferencia de 1,647 puntos (IC 95 % [1,282; 2,012]; $p < 0,001$) y magnitud grande ($g = 1,622$). La autorregulación mantuvo una asociación positiva tras el ajuste por multiplicidad ($B = 0,324$; $q = 0,040$), pero el modelo no mostró capacidad predictiva estable en validación cruzada (R^2 medio = -0,027).

Conclusiones: Los resultados muestran una diferencia posttest amplia a favor de Khan Academy y una señal exploratoria sobre autorregulación. Sin pretest ni identificador de paralelo, los hallazgos deben interpretarse como comparativos y asociativos, no como evidencia causal.

Palabras clave: Khan Academy; trigonometría; rendimiento académico; educación digital; autorregulación.

ABSTRACT

Introduction: Khan Academy offers sequenced practice and immediate feedback that can complement the teaching of trigonometry, although its effect depends on pedagogical integration and the characteristics of the educational context.

Methodology: A quantitative comparative post-test study was carried out with non-equivalent groups. 120 pre-university students from four parallel courses were analyzed, 60 in traditional teaching and 60 with the support of Khan Academy. The comparison included Welch's t, Hedges' g, bootstrap and Mann–Whitney; In the users of the platform, an exploratory regression was adjusted with frequency of use, study time, performance and self-regulation.

Results: The post-test mean was 6.78 in the traditional condition and 8.43 in Khan Academy, with a difference of 1.647 points (95 % CI [1.282; 2.012]; $p < 0.001$) and large magnitude ($g = 1.622$). Self-regulation maintained a positive association after adjustment for multiplicity ($B = 0.324$; $q = 0.040$), but the model did not show stable predictive capacity in cross-validation (mean $R^2 = -0.027$).

Conclusions: The results show a wide post-test difference in favor of Khan Academy and an exploratory signal on self-regulation. Without pretest or parallel identifier, findings should be interpreted as comparative and associative, not causal evidence.

Keywords: Khan Academy; trigonometry; academic performance; digital education; self-regulation.



1. INTRODUCCIÓN

La trigonometría articula razonamiento geométrico, algebraico y funcional y constituye una base para contenidos posteriores de cálculo, física e ingeniería. Su aprendizaje exige coordinar definiciones, procedimientos y representaciones como el triángulo rectángulo, la circunferencia unitaria y las gráficas de funciones. La enseñanza centrada solo en reglas puede producir comprensiones limitadas, mientras que la conexión entre representaciones favorece el significado conceptual (Weber, 2005; Bedada & Machaba, 2022).

Las herramientas digitales pueden apoyar el aprendizaje mediante visualizaciones, práctica graduada y retroalimentación inmediata. La evidencia muestra efectos positivos en matemáticas, aunque variables según el recurso, la formación docente y su integración pedagógica; el uso complementario suele ser más favorable que la sustitución de la enseñanza habitual (Hillmayr et al., 2020; Higgins et al., 2019). En la misma línea, la UNESCO (2023) recomienda valorar la tecnología por sus resultados y por las condiciones concretas de implementación.

Khan Academy combina explicaciones audiovisuales, ejercicios secuenciados, retroalimentación y registros de progreso, por lo que puede utilizarse como apoyo a la enseñanza presencial. Sin embargo, la plataforma no garantiza por sí sola mejores resultados. Kelly y Rutherford (2017) no hallaron diferencias significativas en una intervención controlada de cuatro semanas, aunque las métricas de uso se relacionaron con los puntajes obtenidos.

Otros trabajos sugieren que el acompañamiento y la continuidad de la práctica son relevantes. Oreopoulos et al. (2024) reportaron mejoras de 0,12 a 0,22 desviaciones estándar cuando la plataforma se integró con orientación al profesorado, mientras que Traverso-Condori et al. (2024) observaron mejoras en competencias matemáticas en un estudio de un solo grupo. Estos antecedentes muestran que los resultados dependen del contexto, la duración y la forma de implementación.

Las trazas digitales también permiten estudiar asociaciones entre interacción y rendimiento, aunque la analítica del aprendizaje sigue enfrentando limitaciones para construir modelos generalizables (Ifenthaler & Yau, 2020; Paulsen & Lindsay, 2024). La autorregulación es especialmente relevante en entornos digitales porque implica organizar el tiempo, sostener la



práctica y supervisar el aprendizaje; su relación positiva con el rendimiento ha sido documentada en entornos en línea y combinados (Xu et al., 2023). Cuando se mide al final de una intervención, no obstante, debe interpretarse como asociación y no como antecedente causal.

El objetivo fue comparar el rendimiento posttest en trigonometría entre estudiantes preuniversitarios que utilizaron Khan Academy como apoyo y quienes continuaron con metodología tradicional. De forma complementaria, se examinó dentro de los usuarios de la plataforma la asociación del rendimiento con frecuencia de uso, tiempo de estudio, desempeño en actividades y autorregulación, así como la estabilidad fuera de muestra del modelo. Al no disponerse de pretest ni del identificador de paralelo en la base analítica, el alcance se limitó a una comparación posttest no causal y se incorporó un análisis de sensibilidad ante posible dependencia intraclase.

2. METODOLOGÍA

Enfoque, diseño y alcance

Se adoptó un enfoque cuantitativo y un diseño comparativo posttest con grupos no equivalentes. Se contrastó una condición de enseñanza tradicional con otra que incorporó Khan Academy como recurso de apoyo. El alcance fue comparativo y asociativo exploratorio, sin atribución causal.

La base analítica no conservó pretest ni identificador de curso; por ello no fue posible estimar cambio individual, verificar equivalencia inicial ni ajustar directamente la dependencia entre estudiantes del mismo paralelo. La unidad de análisis fue el estudiante y la posible agrupación se examinó mediante sensibilidad.

Contexto, población y muestra

La población accesible estuvo constituida por 840 estudiantes del curso de nivelación preuniversitaria del área técnica de una institución pública de educación superior de Quito, Ecuador, durante PAC1-2026. La muestra incluyó 120 estudiantes: 60 en metodología tradicional y 60 en Khan Academy.

Se utilizó muestreo no probabilístico por conveniencia a partir de grupos previamente constituidos. Se incluyeron registros con puntaje posttest y condición educativa claramente identificada; los 120 registros estaban completos y sin duplicados.



Precisión y potencia de la muestra

La muestra estuvo determinada por la disponibilidad de cuatro paralelos. Como referencia descriptiva, con $n = 120$, 90 % de confianza y máxima variabilidad, el margen de error teórico bajo muestreo aleatorio simple fue aproximadamente $\pm 7,5$ % (Cochran, 1977); este cálculo no implica representatividad probabilística.

$$e = Z_{pqn}(1)$$

Con 60 estudiantes por condición y $\alpha = 0,05$, la potencia aproximada fue 77,5 % para $d = 0,50$ y 99,1 % para $d = 0,80$; el efecto mínimo detectable con 80 % de potencia fue $d = 0,516$. Estas referencias suponen independencia individual (Faul et al., 2009).

Procedimiento de la intervención

La intervención comprendió cuatro paralelos previamente conformados, dos por condición. El grupo tradicional continuó con la metodología habitual del curso, mientras que Khan Academy se utilizó como recurso complementario de trigonometría. En esta condición se registraron frecuencia de acceso, tiempo de estudio y desempeño en actividades. Al finalizar el período se aplicó el postest de trigonometría a ambas condiciones.

Instrumentos de medición

El rendimiento académico se definió como el puntaje de la prueba de trigonometría en escala de 0 a 10. La autorregulación se registró mediante un índice de 0 a 10 derivado de un cuestionario aplicado al finalizar la intervención. Frecuencia de uso, tiempo de estudio y desempeño en actividades provinieron de los registros automáticos de Khan Academy. Estas variables de plataforma se analizaron únicamente en sus usuarios.

Consideraciones éticas

El análisis se realizó sobre registros académicos y de interacción. La copia analítica contenía las variables de estudio y los resultados se presentaron de forma agregada. La base original se preservó y el procesamiento se efectuó sobre una copia de trabajo, limitando el análisis a la información necesaria para los objetivos del estudio.

Fuente de datos y variables

La base incluyó condición educativa, rendimiento postest, uso de Khan Academy, frecuencia de uso, tiempo de estudio, desempeño en actividades y autorregulación. La condición se codificó como 0 = tradicional y 1 = Khan Academy. El rendimiento y la autorregulación se



expresaron de 0 a 10; las métricas de plataforma se reservaron para el análisis secundario de sus 60 usuarios.

Tabla 1.

Operacionalización de las variables analizadas

Variable	Definición operacional	Escala / unidad	Función analítica
Rendimiento posttest	Puntaje obtenido en la prueba académica de trigonometría.	Continua, 0–10	Desenlace principal
Grupo / uso de Khan Academy	Condición educativa: 0 = tradicional; 1 = Khan Academy.	Dicotómica nominal	Factor de comparación
Frecuencia de uso	Número de sesiones de interacción con la plataforma por semana.	Discreta; sesiones/semana	Secundaria; solo Khan
Tiempo de estudio	Tiempo acumulado de estudio registrado durante la intervención.	Continua; horas/semana	Secundaria; solo Khan
Desempeño en actividades	Promedio de aciertos en ejercicios desarrollados en la plataforma.	Continua, 0–10	Secundaria; solo Khan
Autorregulación	Índice estandarizado obtenido del cuestionario aplicado al final.	Ordinal cuantificada, 0–10	Secundaria; solo Khan

Nota. Las métricas generadas por Khan Academy no se utilizaron para caracterizar ni predecir a los estudiantes del grupo control.

Preparación y control de calidad

El procesamiento se efectuó sobre una copia analítica mediante un script reproducible. Se verificaron tipos de datos, rangos, valores faltantes, duplicados y concordancia entre grupo y uso de Khan Academy. No se imputaron valores ni se añadieron observaciones sintéticas; las exclusiones temporales por influencia se limitaron a análisis de sensibilidad.

Hipótesis y estimandos

La comparación principal evaluó $\Delta = \mu_1 - \mu_0$, donde μ_1 corresponde a Khan Academy y μ_0 a metodología tradicional. Se planteó una hipótesis bilateral de igualdad de medias y se estimó además g de Hedges como medida estandarizada de magnitud.



$$H_0: \mu_1 - \mu_0 = 0 \quad H_1: \mu_1 - \mu_0 \neq 0$$

Análisis estadístico

Todos los contrastes fueron bilaterales, con $\alpha = 0,05$ e IC 95 %. La interpretación consideró magnitud, precisión y estabilidad, además del valor p (Wasserstein & Lazar, 2016).

Análisis descriptivo y comparación principal

Se calcularon media, desviación estándar, mediana, intervalo intercuartílico y rango. La normalidad y homogeneidad de varianzas se examinaron con Shapiro-Wilk y Levene como diagnósticos (Levene, 1960; Shapiro & Wilk, 1965).

La diferencia principal se estimó con t de Welch e IC 95 %, acompañada de g de Hedges. Como sensibilidad se aplicaron Mann-Whitney y correlación biserial por rangos; para Δ y g se calcularon intervalos bootstrap estratificados con 20.000 remuestreos (Efron & Tibshirani, 1993; Hedges, 1981; Mann & Whitney, 1947; Welch, 1947).

$$Y_i = \beta_0 + \beta_1 G_i + \varepsilon_i(2)$$

Como verificación se ajustó una regresión lineal simple del posttest sobre la condición, con errores estándar HC3, IC 95 %, p y R^2 ; su interpretación fue asociativa (MacKinnon & White, 1985).

Análisis asociativo en usuarios de Khan Academy

En los 60 usuarios de Khan Academy se ajustó una regresión lineal múltiple del rendimiento sobre frecuencia de uso, tiempo de estudio, desempeño en actividades y autorregulación. Los coeficientes se interpretaron de forma exploratoria y no causal.

$$Y_i = \alpha_0 + \alpha_1 F_i + \alpha_2 T_i + \alpha_3 D_i + \alpha_4 A_i + \varepsilon_i(3)$$

Se informaron coeficientes, errores HC3, IC 95 %, p, R^2 y R^2 ajustado, con corrección Benjamini-Hochberg. Se evaluaron residuos, heterocedasticidad, multicolinealidad e influencia. La estabilidad se examinó con validación cruzada repetida de cinco pliegues y 200 repeticiones, resumiendo R^2 , RMSE y MAE fuera de muestra (Belsley et al., 1980; Benjamini & Hochberg, 1995; Breusch & Pagan, 1979; Cook, 1977; Kohavi, 1995).

Sensibilidad ante la posible agrupación por curso

Ante la ausencia del identificador de paralelo se realizó una simulación Monte Carlo para valorar sensibilidad a correlación intraclase. Se asumieron cuatro paralelos de 30 estudiantes, dos por condición, y se evaluaron ICC de 0,01; 0,03; 0,05 y 0,10 (Donner & Klar, 2000).

$$DEFF = 1 + (m - 1)ICC \quad neff = nDEFF(4)$$



Cada ICC se combinó con $d = 0; 0,30; 0,50$ y $0,80$ en 30.000 repeticiones. Se comparó el análisis individual con el basado en medias de paralelo mediante tasa de falsos positivos, potencia, sesgo y cobertura del IC 95 %. Los escenarios se interpretaron como sensibilidad y no como estimaciones de la ICC real.

Reproducibilidad y límites analíticos

Los análisis se implementaron en Python con semillas fijas para bootstrap, validación cruzada y simulación, favoreciendo la reproducibilidad del procesamiento.

Las inferencias se limitan al postest y a la muestra disponible. La ausencia de pretest, la falta del identificador de curso, la selección no probabilística y la medición final de autorregulación restringen la atribución causal y la generalización.

3. RESULTADOS

Conformación de la muestra y control de calidad

La muestra estuvo formada por 120 estudiantes de cuatro paralelos: dos tradicionales ($n = 60$) y dos con Khan Academy ($n = 60$). Todos contaron con postest, sin valores faltantes ni duplicados. Las métricas de uso se analizaron solo en los 60 usuarios de la plataforma.

Rendimiento postest por condición educativa

La media postest fue 6,78 ($DE = 1,09$) en la condición tradicional y 8,43 ($DE = 0,92$) en Khan Academy; las medianas fueron 6,60 y 8,48. La diferencia descriptiva fue 1,65 puntos y la dispersión fue ligeramente menor en Khan Academy.

Shapiro-Wilk no detectó desviaciones relevantes de normalidad en ninguno de los grupos y Levene no evidenció diferencias de varianza ($p = 0,188$), por lo que no se observaron anomalías importantes para la comparación prevista.

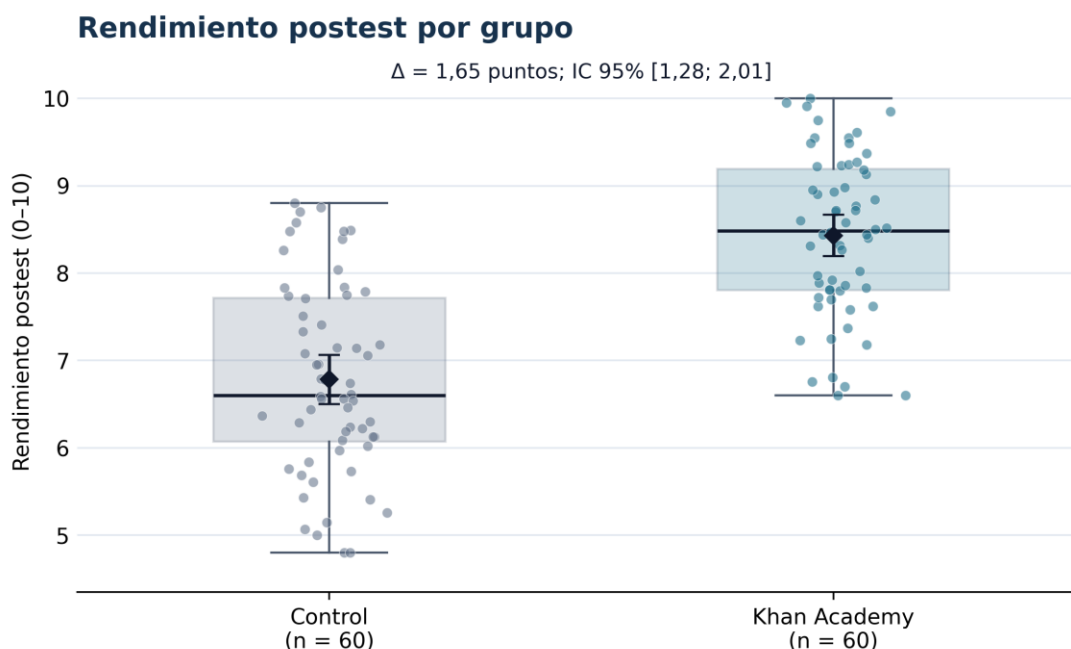
Tabla 2.

Estadísticos descriptivos del rendimiento postest según condición educativa

Condición	n	M	DE	Mediana	RIC	Rango
Tradicional	60	6,78	1,09	6,60	6,07–7,72	4,80–8,80
Khan Academy	60	8,43	0,92	8,48	7,81–9,19	6,60–10,00

Nota. M = media; DE = desviación estándar; RIC = intervalo intercuartílico (Q1–Q3). Escala del rendimiento: 0–10.



Figura 1.*Distribución del rendimiento posttest por condición educativa*

Nota. Los puntos representan estudiantes; las cajas muestran el RIC y la línea central la mediana. El rombo con barras representa la media y su IC 95 %.

Comparación principal entre condiciones

La diferencia media fue 1,647 puntos a favor de Khan Academy (IC 95 % [1,282; 2,012]), $t(114,51) = 8,939$; $p < 0,001$. La estimación muestra una separación clara entre condiciones, aunque el diseño posttest no permite atribuirle causalmente a la plataforma.

La magnitud fue grande (g de Hedges = 1,622; IC 95 % bootstrap [1,250; 2,095]). El bootstrap de la diferencia [1,291; 2,005] y Mann-Whitney ($U = 3131$; $p < 0,001$; $rrb = 0,739$) confirmaron que el resultado no dependió de un único procedimiento analítico.

La regresión simple reprodujo la diferencia no ajustada ($B = 1,647$; $EE\ HC3 = 0,186$; IC 95 % [1,279; 2,015]; $p < 0,001$) y explicó 40,4 % de la variabilidad posttest ($R^2 = 0,404$). Esta asociación no implica causalidad.

Tabla 3.*Comparación principal y análisis de sensibilidad distribucional*

Análisis	Estimación / estadístico	IC 95 %	p
t de Welch	$\Delta = 1,647$; $t = 8,939$	[1,282; 2,012]	< 0,001
g de Hedges	$g = 1,622$	[1,250; 2,095]	—
Bootstrap estratificado	$\Delta = 1,647$	[1,291; 2,005]	—
Mann–Whitney	$U = 3131$; $rrb = 0,739$	—	< 0,001
Regresión simple HC3	$B = 1,647$; $EE = 0,186$	[1,279; 2,015]	< 0,001

Nota. Δ = Khan Academy menos tradicional; rrb = correlación biserial por rangos; EE = error estándar HC3. Bootstrap: 20.000 remuestreos.

Análisis asociativo en los usuarios de Khan Academy

En los 60 usuarios de Khan Academy, frecuencia de uso, tiempo de estudio, desempeño y autorregulación explicaron 15,2 % de la variabilidad ($R^2 = 0,152$; R^2 ajustado = 0,091). El contraste global no alcanzó significación, $F(4,55) = 2,085$; $p = 0,095$.

La autorregulación fue el único predictor con asociación positiva tras el ajuste por multiplicidad ($B = 0,324$; EE HC3 = 0,121; IC 95 % [0,081; 0,567]; $p = 0,010$; $q = 0,040$). Frecuencia, tiempo y desempeño presentaron intervalos que incluyeron cero. Debido al tamaño muestral, al contraste global y a que la autorregulación se midió al final, este resultado se considera exploratorio.

Tabla 4.

Regresión lineal exploratoria en los usuarios de Khan Academy

Variable	B	EE HC3	IC 95 %	p	q BH
Intercepto	4,200	1,610	[0,974; 7,426]	= 0,012	—
Frecuencia de uso	0,075	0,083	[-0,090; 0,241]	= 0,366	0,370
Tiempo de estudio	0,112	0,078	[-0,044; 0,268]	= 0,156	0,312
Desempeño en actividades	0,095	0,105	[-0,116; 0,306]	= 0,370	0,370

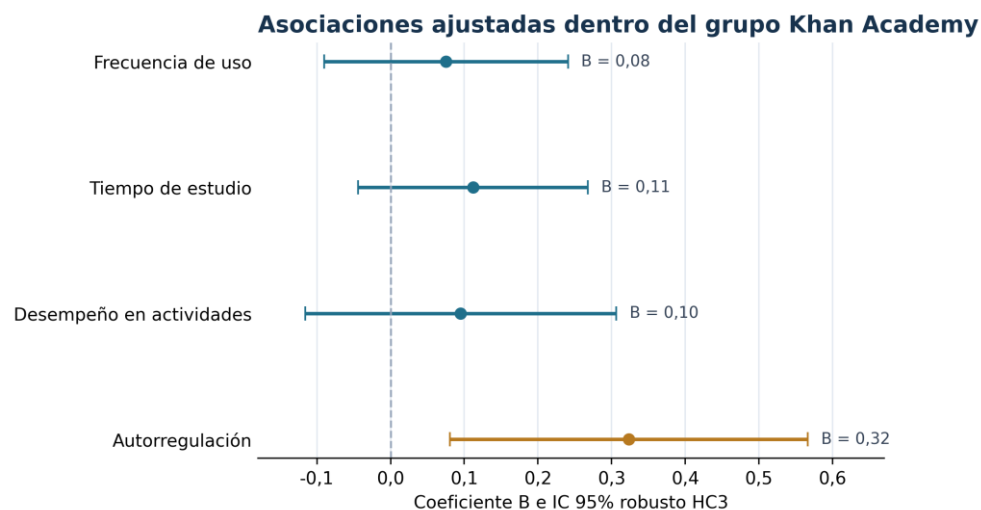


Variable	B	EE HC3	IC 95 %	p	q BH
Autorregulación	0,324	0,121	[0,081; 0,567]	= 0,010	0,040

Nota. $n = 60$. Variable dependiente: rendimiento postest. El ajuste de Benjamini–Hochberg se aplicó a los cuatro predictores; el intercepto no se incluyó.

Figura 2.

Coefficientes del análisis asociativo con intervalos robustos



Nota. Coeficientes no estandarizados e IC 95 % HC3; la figura no implica causalidad ni capacidad predictiva validada.

Diagnósticos y estabilidad del análisis asociativo

Los diagnósticos no mostraron problemas graves de normalidad de residuos, heterocedasticidad o multicolinealidad. Tres registros superaron el criterio de distancia de Cook; al excluirllos temporalmente, la asociación de autorregulación mantuvo el signo y aumentó ($B = 0,407$; $q = 0,002$), por lo que el análisis principal conservó todos los casos.

La validación cruzada mostró escasa estabilidad predictiva: R^2 medio = $-0,027$, RMSE = $0,921$ y MAE = $0,752$, valores muy próximos al modelo basal. Por ello, las asociaciones describen la muestra, pero no sustentan predicción individual para nuevos estudiantes.

Tabla 5.

Ajuste y estabilidad fuera de muestra del análisis secundario



Métrica	Resultado	Interpretación
R ² en muestra	0,152	Variabilidad explicada en los 60 usuarios
R ² ajustado	0,091	Ajuste por cuatro predictores
F global robusto	2,085; p = 0,095	No significativo a $\alpha = 0,05$
R ² validación cruzada	-0,027 [-0,152; 0,071]	Sin generalización estable
RMSE validación cruzada	0,921 [0,877; 0,976]	Modelo basal: 0,928
MAE validación cruzada	0,752 [0,714; 0,799]	Modelo basal: 0,772

Nota. Los intervalos corresponden a percentiles de 200 repeticiones de validación cruzada.

Sensibilidad de la inferencia ante la agrupación por curso

La simulación evaluó el impacto de una posible correlación intraclase que no pudo estimarse directamente. Con $d = 0$, la tasa de falsos positivos del análisis individual aumentó de 8,6 % con ICC = 0,01 a 32,9 % con ICC = 0,10, mientras que el análisis por medias de curso se mantuvo alrededor del 5 %.

La agrupación afectó principalmente la incertidumbre y no desplazó sistemáticamente la diferencia simulada. El tamaño muestral efectivo disminuyó de 93,0 a 30,8 observaciones equivalentes al aumentar la ICC de 0,01 a 0,10; con solo dos paralelos por condición, la potencia del análisis por curso fue limitada.

Estos escenarios no demuestran que la diferencia observada sea espuria, pero indican que los valores p e intervalos calculados a nivel individual pueden resultar optimistas si existe dependencia relevante entre estudiantes del mismo paralelo (Donner & Klar, 2000).

Tabla 6.

Comportamiento inferencial bajo agrupación hipotética por curso

ICC	n efectivo	Falso + individual	Falso + por curso	Potencia d = 0,30	Potencia d = 0,50	Potencia d = 0,80
0,01	93,0	8,6 %	4,9 %	14,2 %	28,4 %	54,0 %
0,03	64,2	15,2 %	5,1 %	11,7 %	21,8 %	42,4 %
0,05	49,0	21,2 %	5,0 %	10,0 %	18,1 %	34,9 %

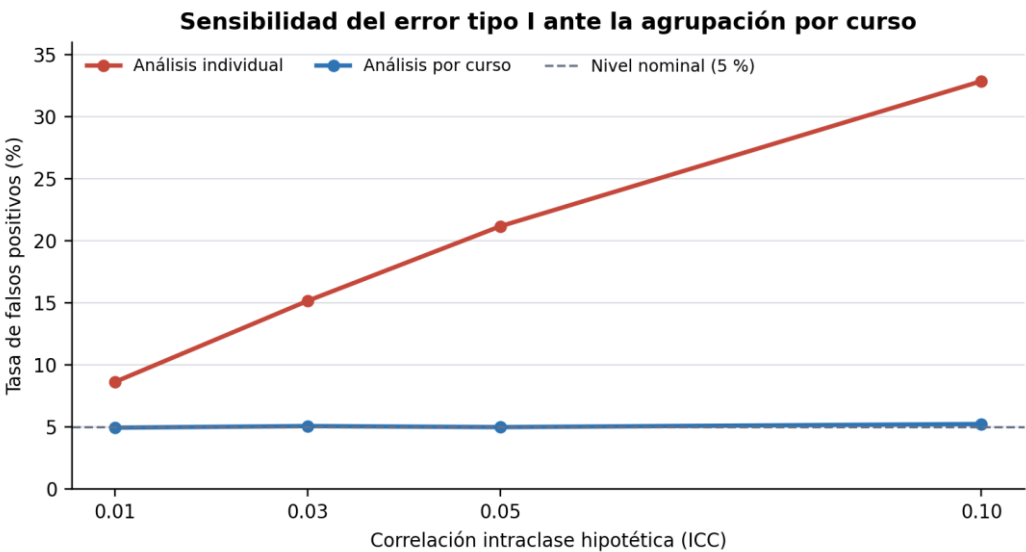


ICC	n efectivo	Falso + individual	Falso + por curso	Potencia d = 0,30	Potencia d = 0,50	Potencia d = 0,80
0,10	30,8	32,9 %	5,2 %	8,2 %	13,5 %	25,1 %

Nota. Cuatro cursos de 30 estudiantes, dos por condición; “Falso +” indica rechazos bajo $d = 0$.

Figura 3.

Sensibilidad del error tipo I ante la correlación intraclase hipotética



Nota. Comparación entre análisis individual y por medias de curso con $d = 0$ y 30.000 repeticiones por ICC.

4. DISCUSIÓN

La diferencia posttest de 1,65 puntos y $g = 1,62$ es compatible en dirección con la evidencia que atribuye a las tecnologías digitales un potencial favorable para el aprendizaje matemático cuando se integran como complemento pedagógico (Hillmayr et al., 2020; Higgins et al., 2019). Sin embargo, la magnitud observada es superior a la reportada habitualmente, por lo que debe interpretarse con cautela. Sin pretest no puede descartarse que parte de la separación final refleje diferencias iniciales entre paralelos u otros factores no controlados.

Los estudios específicos de Khan Academy muestran resultados heterogéneos. Kelly y Rutherford (2017) no encontraron diferencias significativas en cuatro semanas, mientras que Oreopoulos et al. (2024) informaron mejoras de 0,12 a 0,22 desviaciones estándar cuando existió orientación sostenida al profesorado. Traverso-Condori et al. (2024)



también observaron mejoras en competencias matemáticas, aunque con un diseño de un solo grupo. Las discrepancias pueden relacionarse con edad, contenido, duración, intensidad de uso, acompañamiento docente y composición de los grupos.

Desde el punto de vista pedagógico, Khan Academy puede ampliar oportunidades de práctica y retroalimentación, aspectos pertinentes en trigonometría por la necesidad de coordinar representaciones geométricas, algebraicas y gráficas (Weber, 2005; Bedada & Machaba, 2022). Esto coincide con la recomendación de la UNESCO (2023) de integrar la tecnología como apoyo y no asumir que su sola disponibilidad produce aprendizaje. En consecuencia, los resultados respaldan seguir evaluando la plataforma como recurso complementario, no como sustituto automático de la enseñanza.

La asociación positiva de la autorregulación coincide con la evidencia de Xu et al. (2023) sobre su relación con el rendimiento en entornos en línea y combinados. No obstante, aquí se midió al finalizar la intervención, por lo que puede reflejar disposiciones previas, habilidades desarrolladas durante el proceso o percepciones asociadas al rendimiento. Además, el modelo global no fue concluyente; por ello, la autorregulación debe considerarse una señal exploratoria que requiere mediciones basales y confirmación posterior.

La falta de capacidad predictiva estable evita confundir ajuste en la muestra con utilidad para nuevos estudiantes. El R^2 de validación cruzada fue negativo y el error apenas mejoró frente al modelo basal. Esto coincide con las advertencias de Ifenthaler y Yau (2020) y Paulsen y Lindsay (2024): las trazas digitales pueden describir patrones, pero requieren validación rigurosa antes de utilizarse para predicción o clasificación académica.

Las principales limitaciones fueron el muestreo no probabilístico, la participación de solo cuatro paralelos, la ausencia de pretest e identificador de curso y la medición final de autorregulación. La sensibilidad mostró que ignorar la agrupación puede sobreestimar la precisión inferencial. Como aporte, el estudio combina medidas de magnitud, análisis robustos, validación fuera de muestra y sensibilidad a la estructura por paralelos, diferenciando entre descripción, asociación, predicción y causalidad. Estudios prospectivos deberían conservar pretest e identificadores de curso e incluir más paralelos por condición.

Síntesis integrada de los resultados y su alcance

En conjunto, Khan Academy presentó puntajes posttest superiores a la enseñanza tradicional ($\Delta = 1,65$; IC 95 % [1,28; 2,01]; $g = 1,62$). La autorregulación mostró la asociación



ajustada más consistente dentro de sus usuarios, pero el modelo no generalizó adecuadamente fuera de muestra.

La evidencia debe interpretarse como comparativa y asociativa. La ausencia de equivalencia basal y la posible dependencia intraclase impiden atribuir causalidad exclusiva a la plataforma y obligan a moderar la precisión de la inferencia.

5. CONCLUSIÓN

Los estudiantes que utilizaron Khan Academy obtuvieron un promedio postest 1,65 puntos superior al de la condición tradicional, con una magnitud estandarizada grande y resultados consistentes entre procedimientos paramétricos, no paramétricos y bootstrap. La conclusión es descriptiva y asociativa; sin pretest no puede establecerse cuánto mejoró cada estudiante ni atribuir la diferencia exclusivamente a la plataforma.

Dentro de los usuarios de Khan Academy, la autorregulación fue la única variable que mantuvo una asociación positiva tras el ajuste por multiplicidad, mientras que el modelo conjunto explicó una proporción limitada de la variabilidad y no mostró capacidad predictiva estable en validación cruzada.

La posible agrupación por paralelo puede subestimar la incertidumbre cuando los registros se tratan como independientes. Por ello, futuros estudios deberían conservar pretest e identificadores de curso, incluir más paralelos por condición y validar los modelos en nuevas cohortes.

En términos educativos, los hallazgos respaldan continuar evaluando Khan Academy como apoyo a la enseñanza de trigonometría, integrada con planificación docente y estrategias de autorregulación, sin asumir que sustituye automáticamente a la enseñanza habitual.

REFERENCIAS BIBLIOGRÁFICAS

Bedada, T. B., & Machaba, F. (2022). The effect of GeoGebra on STEM students learning trigonometric functions. *Cogent Education*, 9(1), 2034240. <https://doi.org/10.1080/2331186X.2022.2034240>

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons. <https://doi.org/10.1002/0471725153>



- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287–1294. <https://doi.org/10.2307/1911963>
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18. <https://doi.org/10.1080/00401706.1977.10489493>
- Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. Arnold.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.3102/10769986006002107>
- Higgins, K., Huscroft-D'Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A meta-analysis. *Journal of Educational Computing Research*, 57(2), 283–319. <https://doi.org/10.1177/0735633117748416>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and*



Development, 68, 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>

Kelly, D. P., & Rutherford, T. (2017). Khan Academy as supplemental instruction: A controlled study of a computer-based mathematics intervention. *The International Review of Research in Open and Distributed Learning*, 18(4). <https://doi.org/10.19173/irrodl.v18i4.2984>

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (Vol. 2, pp. 1137–1143). Morgan Kaufmann.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3), 305–325. [https://doi.org/10.1016/0304-4076\(85\)90158-7](https://doi.org/10.1016/0304-4076(85)90158-7)

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <https://doi.org/10.1214/aoms/1177730491>

Oreopoulos, P., Gibbs, C., Jensen, M., & Price, J. (2024). Teaching teachers to use computer assisted learning effectively: Experimental and quasi-experimental evidence (NBER Working Paper No. 32388). National Bureau of Economic Research. <https://doi.org/10.3386/w32388>

Paulsen, L., & Lindsay, E. (2024). Learning analytics dashboards are increasingly becoming about learning and not just analytics: A systematic review. *Education and Information Technologies*, 29, 14279–14308. <https://doi.org/10.1007/s10639-023-12401-4>

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>

Traverso-Condori, L. C., Reyes-Arco, R. E., Reyes-Villalba, E., Torres-Gonzales, S. B., Maraza-Quispe, B., Tinco-Túpac, S. T., & Torres-Loayza, J. L. (2024). Evaluation of the impact of an online educational platform on the development of mathematical competencies in elementary school students. *IEEE Revista Iberoamericana*



de Tecnologías del Aprendizaje, 19, 397–406.
<https://doi.org/10.1109/RITA.2024.3465021>

UNESCO. (2023). Global education monitoring report 2023: Technology in education: A tool on whose terms? <https://doi.org/10.54676/UZQV8501>

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. The American Statistician, 70(2), 129–133.
<https://doi.org/10.1080/00031305.2016.1154108>

Weber, K. (2005). Students' understanding of trigonometric functions. Mathematics Education Research Journal, 17(3), 91–112. <https://doi.org/10.1007/BF03217423>

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. Biometrika, 34(1–2), 28–35.
<https://doi.org/10.1093/biomet/34.1-2.28>

Xu, Z., Zhao, Y., Zhang, B., Liew, J., & Kogut, A. (2023). A meta-analysis of the efficacy of self-regulated learning interventions on academic achievement in online and blended environments in K–12 and higher education. Behaviour & Information Technology, 42(16), 2911–2931. <https://doi.org/10.1080/0144929X.2022.2151935>

CONFLICTO DE INTERESES

No existe.

FINANCIAMIENTO

No aplica.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: José Oswaldo Briones Calvache, Pablo Ignacio Páliz Larrea.

Curación de datos: José Oswaldo Briones Calvache, Pablo Ignacio Páliz Larrea.

Análisis formal: José Oswaldo Briones Calvache, Pablo Ignacio Páliz Larrea.

Redacción – Borrador original: José Oswaldo Briones Calvache, Pablo Ignacio Páliz Larrea.

Redacción – Revisión y edición: José Oswaldo Briones Calvache, Pablo Ignacio Páliz Larrea.

