

**Recibido:** 2026-04-24

**Aceptado:** 2026-05-08

**Publicado:** 2026-05-22

**Efectos cognitivos y actitudinales de un taller de nivelación de  
Matemática General mediado por IA generativa con prompt de tutor  
socrático**

**Cognitive and attitudinal effects of a General Mathematics leveling  
workshop mediated by generative AI with socratic tutor prompt**

**Autor(s):**

**Melvin Ramírez Bogantes <sup>1</sup>**

Escuela de Matemática

[mramirez@itcr.ac.cr](mailto:mramirez@itcr.ac.cr)

<https://orcid.org/0000-0001-5516-0085>

**Instituto Tecnológico de Costa Rica**

Cartago – Costa Rica

**Esteban Quesada Ávila <sup>2</sup>**

Licenciatura en la Enseñanza de la Matemática

[teo021097@estudiantec.cr](mailto:teo021097@estudiantec.cr)

<https://orcid.org/0009-0009-3847-036X>

**Instituto Tecnológico de Costa Rica**

Cartago – Costa Rica

**Francine Araya López <sup>3</sup>**

Licenciatura en la Enseñanza de la Matemática

[caaraya@estudiantecr.onmicrosoft.com](mailto:caaraya@estudiantecr.onmicrosoft.com)

<https://orcid.org/0009-0008-3789-5510>

**Instituto Tecnológico de Costa Rica**

Cartago – Costa Rica

## Resumen

Las brechas matemáticas con que los estudiantes ingresan a la universidad, particularmente en el razonamiento abstracto sobre el concepto de función, constituyeron el problema cognitivo de partida de este estudio, cuya mitigación exige dispositivos formativos breves de alto impacto. El objetivo fue evaluar los efectos cognitivos y actitudinales de un taller de nivelación de Matemática General mediado por inteligencia artificial generativa (IAGen) con prompt de tutor socrático, implementado en el Instituto Tecnológico de Costa Rica. Se aplicó un diseño cuantitativo pre-experimental de grupo único con mediciones pre y post, correspondiente al componente cuantitativo de un estudio mixto convergente paralelo cuyo componente cualitativo se reporta en publicación complementaria. Veintidós estudiantes de primer ingreso completaron una prueba isomórfica de 15 ítems sobre el concepto de función y una escala actitudinal de 18 ítems Likert. Los datos se analizaron mediante estadística descriptiva, prueba t de Student para muestras relacionadas, prueba de Wilcoxon, prueba exacta de McNemar, prueba U de Mann-Whitney y el tamaño del efecto d de Cohen. La intervención produjo una mejora media de 24,85 puntos ( $t(21) = 6,14$ ;  $p < 0,001$ ; d de Cohen = 1,31), con 86,4 % de estudiantes que mejoraron. La ganancia fue mayor en la dimensión algebraica (+34,1 pp;  $d = 1,63$ ) que en la conceptual (+18,7 pp;  $d = 0,69$ ). La valoración actitudinal resultó marcadamente positiva (Disposición M = 4,49;  $\alpha$  global = 0,921;  $\omega = 0,953$ ), con una brecha significativa por experiencia previa con IA en utilidad ( $p = 0,035$ ) y disposición ( $p = 0,022$ ). La triangulación interna entre instrumentos identificó un patrón asimétrico: la mediación con IAGen bajo tutoría socrática generó ganancias procedimentales sustanciales y aceptación tecnológica robusta, pero encontró un límite en el registro gráfico-simbólico del concepto de función. Se concluye que la IAGen con prompt de tutor socrático constituye un dispositivo eficaz para mitigar el rezago cognitivo en contenidos procedimentales dentro de talleres breves de nivelación universitaria, pero no sustituye el tratamiento didáctico explícito de los obstáculos epistemológicos asociados a la articulación entre registros de representación.

**Palabras clave:** enseñanza de las matemáticas; inteligencia artificial; educación superior; aprendizaje activo; actitud del estudiante.

---

## Abstract

The mathematical gaps with which students enter university, particularly in abstract reasoning about the function concept, constituted the initial cognitive problem of this study, whose mitigation requires brief high-impact formative devices. The objective was to evaluate the cognitive and attitudinal effects of a General Mathematics leveling workshop mediated by generative artificial intelligence (GenAI) with a Socratic tutor prompt, implemented at the Instituto Tecnológico de Costa Rica. A quantitative single-group pre-experimental design with pre- and post-measurements was applied, corresponding to the quantitative component of a convergent parallel mixed-methods study whose qualitative component is reported in a complementary publication. Twenty-two first-year students completed a 15-item isomorphic test on the function concept and an 18-item Likert attitudinal scale. Data were analyzed using descriptive statistics, Student's t-test for related samples, Wilcoxon test, exact McNemar test, Mann-Whitney U test, and Cohen's d effect size. The intervention produced a mean improvement of 24.85 points ( $t(21) = 6.14$ ;  $p < .001$ ; Cohen's  $d = 1.31$ ), with 86.4% of students improving. The gain was greater in the algebraic dimension (+34.1 pp;  $d = 1.63$ ) than in the conceptual one (+18.7 pp;  $d = 0.69$ ). The attitudinal assessment was markedly positive (Disposition  $M = 4.49$ ; global  $\alpha = .921$ ;  $\omega = .953$ ), with a significant gap by prior AI experience in utility ( $p = .035$ ) and disposition ( $p = .022$ ). Internal triangulation between instruments identified an asymmetric pattern: GenAI mediation under Socratic tutoring generated substantial procedural gains and robust technological acceptance, but encountered a limit in the graphic-symbolic register of the function concept. It is concluded that GenAI with a Socratic tutor prompt constitutes an effective device for mitigating cognitive lag in procedural content within brief university leveling workshops, but does not replace the explicit didactic treatment of epistemological obstacles associated with articulation between representation registers.

**Keywords:** *mathematics education; artificial intelligence; higher education; active learning; student attitudes.*

## Introducción

El rezago en razonamiento matemático abstracto con que los estudiantes ingresan a la universidad, particularmente en el dominio del concepto de función y de sus articulaciones algebraico-gráficas, constituye un problema cognitivo persistente y ampliamente documentado en el contexto latinoamericano. Estudios sucesivos sobre el desempeño matemático al ingreso a la educación superior reportan deficiencias sistemáticas en aritmética, álgebra elemental y comprensión del concepto de función, deficiencias que se traducen posteriormente en altas tasas de reprobación de los primeros cursos universitarios y en abandono temprano de las carreras de ingeniería. Este rezago cognitivo, y no la ausencia de una tecnología específica, constituye el problema científico de fondo del presente estudio. Frente a él, las instituciones de educación superior han desarrollado dispositivos de nivelación previos al inicio del semestre, con la finalidad de mitigar la brecha cognitiva antes del primer curso. Estos dispositivos operan, sin embargo, bajo condiciones de fuerte presión temporal: pretenden cubrir contenidos del primer curso en una semana intensiva o en pocas sesiones distribuidas, lo que limita el alcance de las estrategias instruccionales centradas en la transmisión directa y la práctica masiva frente a las exigencias de elaboración cognitiva activa que la literatura psicoeducativa identifica como condición del aprendizaje matemático profundo (Sweller et al., 2019).

La irrupción de la inteligencia artificial generativa (IAGen) en los entornos formativos universitarios introduce, en este escenario, *un dispositivo tecnológico-pedagógico con potencial para mediar la reducción del rezago cognitivo descrito*. Las herramientas de IAGen permiten personalizar la atención, proporcionar retroalimentación inmediata, generar múltiples representaciones de un mismo contenido y sostener interacciones individualizadas a un costo marginal bajo (Kasneci et al., 2023). Estas características resultan particularmente atractivas para contextos de nivelación, donde la heterogeneidad cognitiva del estudiantado y la presión temporal hacen difícil la atención individualizada por parte del docente. La integración pedagógica de la IAGen, no obstante, requiere decisiones didácticas explícitas que orienten el uso de la herramienta hacia la elaboración cognitiva activa y eviten su reducción a un dispositivo de entrega de respuestas.

Una propuesta consolidada para canalizar productivamente la mediación con IAGen consiste en estructurar la interacción mediante el llamado prompt de tutor socrático: una

instrucción que convierte a la herramienta en un guía que acompaña el razonamiento mediante preguntas, sin proveer la solución directa. El supuesto psicoeducativo subyacente sostiene que esta forma de interacción, además de evitar el aprendizaje superficial, genera condiciones para una elaboración cognitiva más profunda. Hasta qué punto esta forma de mediación produce efectos cognitivos y actitudinales medibles, en qué dimensiones del aprendizaje matemático se concentran esos efectos y cómo se distribuyen entre subgrupos de estudiantes son preguntas que requieren evidencia empírica sistemática y que la literatura emergente sobre IAGen en educación matemática aún no ha respondido para el contexto universitario latinoamericano.

### **Conocimiento conceptual y conocimiento procedimental**

La distinción entre conocimiento conceptual y conocimiento procedimental constituye uno de los marcos clásicos para el análisis del aprendizaje matemático. Rittle-Johnson y Schneider (2015), en una revisión sistemática que sintetiza dos décadas de investigación, definen el conocimiento conceptual como la comprensión de los principios que rigen un dominio y de las relaciones entre las unidades de información, mientras que el conocimiento procedimental refiere a la capacidad de ejecutar secuencias de pasos para resolver problemas. Aunque ambos componentes se desarrollan en interacción, el avance en uno suele facilitar el avance en el otro, su evaluación requiere instrumentos sensibles a la naturaleza específica de cada dimensión. Este marco resulta operativo para la evaluación de intervenciones formativas porque permite identificar patrones diferenciales: una intervención puede generar ganancias sustanciales en lo procedimental sin que ello implique avances equivalentes en lo conceptual. La distinción es particularmente relevante en intervenciones breves, donde las reorganizaciones cognitivas profundas que exige el conocimiento conceptual pueden no alcanzar el tiempo de maduración necesario, mientras que la práctica estructurada del conocimiento procedimental sí encuentra condiciones favorables.

### **Obstáculos epistemológicos y registros del concepto de función**

El concepto de función ha sido objeto de extensa investigación didáctica por su centralidad en el currículo matemático y por las dificultades sistemáticas que su aprendizaje presenta. Brousseau (1998), en su teoría de las situaciones didácticas, conceptualiza ciertas dificultades del aprendizaje como obstáculos epistemológicos:

configuraciones cognitivas que, una vez instaladas, resisten su transformación porque están enraizadas en concepciones más amplias del conocimiento matemático. La noción de función como correspondencia entre conjuntos, la distinción entre imagen y preimagen, y la lectura simultánea del registro algebraico y el registro gráfico son puntos donde estos obstáculos se manifiestan con particular insistencia. En articulación con esta perspectiva, y atendiendo precisamente al núcleo gráfico-simbólico señalado por Brousseau, Duval (2006) desarrolla la noción complementaria de articulación entre registros de representación. Para Duval, la comprensión matemática auténtica exige la capacidad de transitar entre registros distintos: algebraico, gráfico, tabular, lenguaje natural, y de reconocer en ellos un mismo objeto matemático. La fragilidad en esta articulación produce errores característicos: el estudiante puede ejecutar correctamente operaciones en el registro algebraico y, simultáneamente, no reconocer su correlato gráfico. Investigaciones recientes (Artigue y Winsløw, 2010) confirmaron la persistencia de estas dificultades en estudiantes universitarios, especialmente en lo relativo a la lectura de intersecciones con los ejes y a la interpretación de la notación funcional sobre representaciones gráficas; ambos contenidos resultarán centrales en la interpretación de los resultados del presente estudio.

### **Modelo de aceptación tecnológica y aplicación a la IA en educación**

El Modelo de Aceptación Tecnológica (Technology Acceptance Model, TAM) propuesto por Davis (1989) ofrece un marco consolidado para analizar la actitud de los usuarios hacia las tecnologías. El modelo postula que dos constructos predicen la intención de uso: la utilidad percibida: el grado en que la persona considera que la tecnología mejorará su desempeño, y la facilidad de uso percibida. La utilidad percibida ha sido reiteradamente confirmada como predictor robusto de la intención de uso en contextos educativos. La emergencia de la IAGen ha motivado aplicaciones recientes del marco a la integración de estas herramientas en educación superior. Almassaad et al. (2024), Bernabei et al. (2023) y Strzelecki (2024), entre otros, documentaron la replicación del patrón clásico del TAM en estudios sobre aceptación de IAGen: la percepción de utilidad predice fuertemente la disposición a incorporar la herramienta a las prácticas de estudio. Junto con la utilidad y la disposición, la motivación intrínseca hacia el dominio disciplinar (Pintrich et al., 1991) constituye una dimensión actitudinal complementaria. Estos tres constructos: utilidad,

motivación y disposición; articulan, junto con una cuarta dimensión de uso crítico, el espacio actitudinal evaluado en el presente estudio.

### **Mediación con IA generativa en educación matemática**

La integración de IAGen en la enseñanza de las matemáticas constituye un campo de investigación emergente y de rápida evolución. Kasneci et al. (2023), en una de las primeras revisiones sistemáticas sobre el tema, identificaron tanto las potencialidades como los riesgos de la mediación: la herramienta puede ofrecer explicaciones personalizadas, ejemplos múltiples y retroalimentación inmediata, pero también puede inducir aprendizaje superficial cuando se reduce a la entrega de respuestas. Revisiones posteriores ampliaron este diagnóstico inicial. Yan et al. (2024), en una revisión sistemática de alcance sobre desafíos prácticos y éticos de los modelos de lenguaje en educación, advirtieron sobre la necesidad de diseños didácticos que mitiguen el riesgo de delegación cognitiva.

### **Aporte del estudio y objetivos**

El aporte del presente estudio al área del conocimiento se sitúa en la intersección de los marcos anteriores. Primero, ofrece evidencia empírica cuantitativa, basada en dos instrumentos validados, sobre los efectos cognitivos diferenciales de la mediación con IAGen y tutor socrático en contenidos procedimentales versus conceptuales del concepto de función, dimensión sobre la cual la literatura disponible aún es escasa. Segundo, articula una triangulación interna entre dos instrumentos cuantitativos: prueba cognitiva isomórfica y escala actitudinal, que permite identificar patrones diferenciales del aprendizaje difíciles de captar desde un solo plano de análisis. Tercero, documenta una brecha por experiencia previa con IA que vincula la integración pedagógica de la herramienta con la alfabetización digital previa del estudiantado, hallazgo con implicaciones de equidad relevantes para el diseño de políticas institucionales de nivelación matemática.

El objetivo general del estudio fue evaluar los efectos cognitivos y actitudinales de un taller de nivelación de Matemática General mediado por IAGen con prompt de tutor socrático, implementado en el Instituto Tecnológico de Costa Rica. Tres objetivos específicos lo desagregaron: (a) determinar la magnitud del cambio pre/post en el desempeño cognitivo sobre el concepto de función, así como su distribución por

dimensión (conceptual y algebraica) y por subtema; (b) caracterizar la valoración actitudinal del estudiantado en cuatro subescalas teóricamente fundamentadas: utilidad, motivación, uso crítico y disposición, y la estructura relacional entre ellas; y (c) identificar, mediante triangulación interna entre ambos instrumentos cuantitativos, patrones diferenciales del aprendizaje que orientaran el rediseño didáctico de futuras iteraciones. El presente artículo reporta el componente cuantitativo del estudio mayor, de diseño mixto convergente paralelo; el componente cualitativo, que examina la apropiación pedagógica del prompt y el modelo híbrido docente, se reporta en publicación complementaria.

### Hipótesis de investigación

Sobre la base del marco teórico anterior, el estudio contrastó la siguiente hipótesis diferencial sobre el efecto cognitivo de la intervención. Sea  $\mu_{pre}$  la media poblacional de desempeño cognitivo en el pretest y  $\mu_{post}$  la media poblacional de desempeño cognitivo en el post-test sobre el concepto de función.

**Hipótesis nula ( $H_0$ ):** No existió diferencia estadísticamente significativa entre el desempeño cognitivo pre y post sobre el concepto de función tras la intervención mediada por IAGen con prompt de tutor socrático.

$$H_0: \mu_{post} - \mu_{pre} = 0$$

**Hipótesis de investigación ( $H_1$ ):** El desempeño cognitivo sobre el concepto de función fue significativamente mayor en el post-test que en el pretest tras la intervención mediada por IAGen con prompt de tutor socrático, con un tamaño del efecto al menos grande.

$$H_1: \mu_{post} - \mu_{pre} > 0; |d| \geq 0,80$$

## Material y métodos

### Material

El material formativo central del estudio fue una unidad didáctica completa diseñada específicamente para el taller, que se distribuyó en cinco sesiones presenciales consecutivas de 2 horas y 30 minutos cada una. La unidad se organizó en torno a ocho

ejercicios estructurados que cubren el contenido completo del concepto de función previsto en la guía del Programa Éxito Académico: identificación de funciones, dominio y restricciones, imágenes y preimágenes, intersecciones con los ejes, razón de cambio promedio, transformaciones, composición, inversa, valor absoluto e inyectividad.

Cada ejercicio siguió la secuencia metodológica PRVR (Planear–Resolver–Verificar–Reflexionar), en la que el estudiante: (a) planeó la ruta de resolución antes de operar, (b) resolvió manualmente el ejercicio, (c) contrastó su resolución con representaciones gráficas en GeoGebra o Desmos y verificó el razonamiento mediante interacción con IAGen bajo prompt de tutor socrático, y (d) registró en una bitácora reflexiva el proceso seguido, los errores detectados y los aprendizajes derivados. La triangulación representacional: algebraica, gráfica y mediada por IA, constituyó el eje metodológico transversal de la unidad.

El prompt de tutor socrático, modelado en la sesión inicial, consistió en una instrucción que reconfiguró el comportamiento del modelo de lenguaje para que (a) no proporcionara la solución directa al ejercicio planteado, (b) acompañara el razonamiento del estudiante mediante preguntas orientadoras escalonadas, (c) ofreciera retroalimentación específica sobre el procedimiento más que sobre el resultado, y (d) promoviera la verificación independiente de las respuestas obtenidas. El prompt base fue adaptado por cada estudiante a las particularidades de cada ejercicio. La intervención no prescribió una plataforma específica de IAGen: cada estudiante utilizó el modelo de lenguaje con el que tenía familiaridad previa, con el propósito explícito de preservar la validez ecológica de la mediación y evitar que la curva de familiarización con una plataforma nueva interfiriera con el aprendizaje matemático. La observación sistemática de las sesiones documentó que la mayoría de los estudiantes utilizó ChatGPT (OpenAI), con utilización minoritaria de otras herramientas. La unidad didáctica incorporó, adicionalmente, un marco ético explícito, la “Regla de Oro”, que estableció los límites del uso académicamente honesto de la IAGen (Chan y Hu, 2023).

Se aplicaron dos instrumentos cuantitativos de recolección de datos. El primero, una prueba cognitiva isomórfica de 15 ítems de selección única sobre el concepto de función, administrada en versión pre y versión post. Cada ítem presenta cuatro opciones de respuesta con una única alternativa válida; la puntuación es dicotómica (1 = correcta; 0 = incorrecta) y el puntaje bruto se transforma a escala porcentual de 0 a 100. Los 15 ítems

se clasifican en dos dimensiones cognitivas: 9 conceptuales y 6 algebraicos. La cobertura por subtemas incluye identificación de funciones, dominio y restricciones, imágenes y preimágenes, intersecciones con los ejes, razón de cambio promedio, transformaciones, composición, inversa, valor absoluto e inyectividad.

La validación de contenido de ambas versiones de la prueba se realizó mediante el método de juicio de expertos siguiendo el procedimiento de consenso descrito por Escobar-Pérez y Cuervo-Martínez (2008). Tres expertos en didáctica de la matemática e integración tecnológica evaluaron, ítem a ítem, los criterios de (a) suficiencia, (b) claridad, (c) coherencia y (d) relevancia, así como la equivalencia conceptual entre las versiones pre y post. El procedimiento alcanzó consenso unánime de los tres expertos sobre la totalidad de los ítems en los cuatro criterios, lo que sustenta la validez de contenido del instrumento. Por tratarse de un procedimiento de consenso cualitativo, no se reportan coeficientes numéricos del tipo V de Aiken o Razón de Validez de Contenido (CVR); esta decisión metodológica se retoma en el apartado de limitaciones.

El segundo instrumento es una escala actitudinal sobre el uso de IAGen, compuesta por 18 ítems tipo Likert de cinco puntos (1 = Muy en desacuerdo; 5 = Muy de acuerdo). Los ítems se organizan teóricamente en cuatro subescalas: Utilidad (5 ítems, fundamentada en TAM), Motivación (2 ítems), Uso crítico (5 ítems, con un ítem inverso) y Disposición (6 ítems). La escala incluye, además, tres preguntas abiertas al final, sobre aprendizajes obtenidos, aspectos valorados y sugerencias de mejora. El cuestionario incluye también una sección sociodemográfica y dos ítems sobre experiencia previa con IA: uso durante la secundaria y frecuencia de uso actual fuera del aula. La escala actitudinal fue validada por el mismo panel de tres expertos siguiendo el procedimiento descrito.

## **Métodos**

El estudio adoptó un diseño cuantitativo pre-experimental de grupo único con mediciones pre y post, ubicado en el componente cuantitativo de un diseño mixto convergente paralelo (Creswell y Plano Clark, 2018). La elección del diseño pre-experimental respondió a las restricciones del contexto institucional: el taller se ofrece a la totalidad de los estudiantes que cumplen los criterios de admisión al Programa Éxito Académico, lo que impidió la conformación de un grupo control con asignación aleatoria. Esta limitación

se aborda explícitamente en el apartado de limitaciones y se mitigó mediante la triangulación interna entre los dos instrumentos cuantitativos.

El procedimiento de selección de la muestra correspondió a un muestreo no probabilístico por conveniencia, también denominado muestreo intencional (Hernández-Sampieri y Mendoza, 2018), dado que estuvo constituida por la totalidad de estudiantes matriculados en la sección presencial del taller que cumplieran con los criterios de inclusión del programa, sin que mediara procedimiento aleatorio de asignación.

Las hipótesis formuladas en la introducción se contrastaron mediante las siguientes pruebas, todas al nivel de significancia  $\alpha = 0,05$ . Para el cambio cognitivo global pre/post se calcularon medias, desviaciones estándar, percentiles y el intervalo de confianza al 95 % para la diferencia media. La normalidad de las diferencias se verificó mediante la prueba de Shapiro-Wilk (Shapiro y Wilk, 1965); verificado el supuesto, se aplicó la prueba t de Student para muestras relacionadas como prueba principal y la prueba de los rangos con signo de Wilcoxon como contraste de robustez no paramétrico. El tamaño del efecto se estimó mediante la d de Cohen (Cohen, 1988) para diferencias en muestras relacionadas. Para el análisis ítem por ítem se aplicó la prueba exacta de McNemar. La confiabilidad interna de la prueba cognitiva se estimó mediante el coeficiente KR-20. La relación entre desempeño inicial y final se examinó con el coeficiente de correlación de Spearman. Para los análisis por subgrupos se utilizó la prueba U de Mann-Whitney, reportándose el estadístico U, el estadístico Z asociado y el valor p exacto.

El análisis de la escala actitudinal contempló estadística descriptiva por ítem y por subescala, evaluación de la consistencia interna mediante el coeficiente alfa de Cronbach y el coeficiente omega de McDonald (Dunn et al., 2014), correlación ítem-total corregida, análisis de «alfa si el ítem se elimina», correlaciones de Pearson entre subescalas y comparación de subgrupos mediante la prueba U de Mann-Whitney. Las respuestas a las preguntas abiertas se codificaron temáticamente mediante un procedimiento inductivo de análisis temático reflexivo (Braun y Clarke, 2006), identificando categorías emergentes; esta codificación, de naturaleza cualitativa, se incorporó únicamente con fines de triangulación interna. El estudio contó con consentimiento informado escrito de todos los participantes y los datos fueron anonimizados con códigos alfanuméricos.

## Resultados

### Descripción de la muestra

La intervención se desarrolló durante la semana del 9 al 13 de febrero de 2026, en el marco del taller de nivelación de Matemática General del Programa Éxito Académico del Instituto Tecnológico de Costa Rica, en el Campus Tecnológico Central de Cartago. La convocatoria reunió a 32 estudiantes de primer ingreso matriculados, de los cuales 22 completaron tanto el pretest como el post-test y conforman la muestra efectiva del análisis. La atrición del 31,3 % es consistente con la reportada en intervenciones breves de nivelación previas al inicio del semestre. El grupo se compuso mayoritariamente por estudiantes de carreras de ingeniería del Campus Tecnológico Central, con concentración en Ingeniería en Computación, Construcción, Electrónica, Mantenimiento Industrial, Agronegocios, Producción Industrial, Agrícola y otras, además de Administración de Tecnologías de Información. La asistencia fue alta: 86,4 % de los estudiantes asistieron a las cinco sesiones del taller. La Tabla 1 sintetiza la caracterización de la muestra efectiva.

**Tabla 1** *Caracterización de la muestra efectiva (n = 22)*

| Variable         | Categoría                     | Frecuencia (%) |
|------------------|-------------------------------|----------------|
| Muestra inicial  | Matriculados al taller        | 32 (100,0)     |
| Muestra efectiva | Pretest y post-test completos | 22 (68,7)      |
| Atrición         | Sin una o ambas pruebas       | 10 (31,3)      |
| Sexo             | Masculino                     | 14 (63,6)      |
|                  | Femenino                      | 8 (36,4)       |
| Asistencia       | 5 / 5 sesiones                | 19 (86,4)      |
|                  | 4 / 5 sesiones                | 3 (13,6)       |

**Nota.** Los porcentajes de la muestra efectiva se calculan sobre n = 22; los de muestra inicial y atrición, sobre N = 32. Elaboración propia.

## Análisis de los Resultados

### *Cambio cognitivo global*

El análisis comparativo entre las puntuaciones pretest y post-test evidenció una mejora sustancial y estadísticamente significativa en el desempeño global del grupo, sintetizada en la Tabla 2.

**Tabla 2** Estadística descriptiva e inferencial del cambio pre/post ( $n = 22$ )

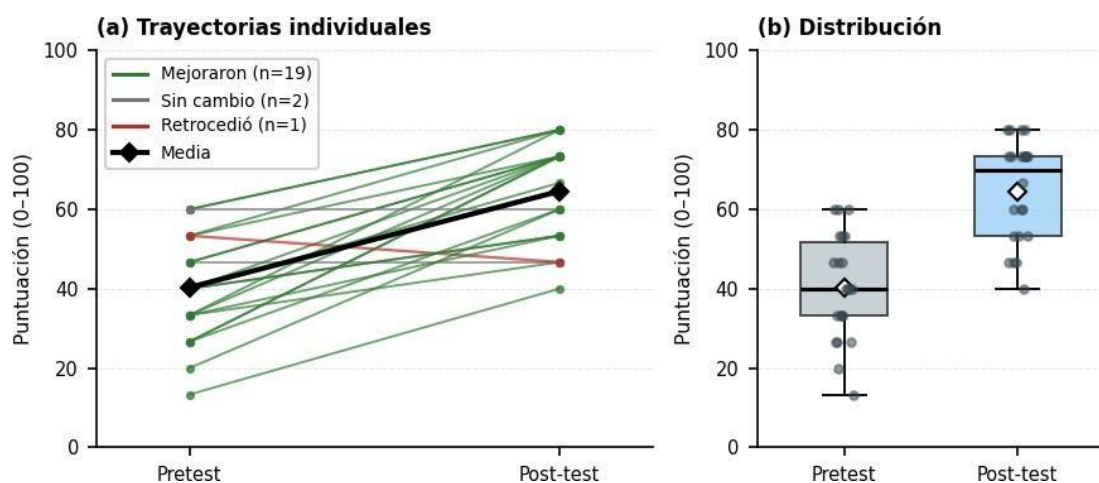
| Estadístico                           | Pretest                   | Post-test |
|---------------------------------------|---------------------------|-----------|
| Media                                 | 40,00                     | 64,85     |
| Desviación estándar                   | 13,65                     | 12,55     |
| Mediana                               | 40,00                     | 66,67     |
| Mínimo                                | 6,67                      | 26,67     |
| Máximo                                | 60,00                     | 80,00     |
| <b>Diferencia media (post – pre)</b>  | <b>+24,85</b>             |           |
| IC 95 % para la diferencia media      | (16,91; 32,78)            |           |
| Prueba t pareada                      | $t(21) = 6,14; p < 0,001$ |           |
| <b>d de Cohen (tamaño del efecto)</b> | <b>1,31 (muy grande)</b>  |           |

**Nota.** Elaboración propia.

La media ascendió de 40,00 puntos en el pretest (DE = 13,65) a 64,85 puntos en el post-test (DE = 12,55), con un incremento absoluto de 24,85 puntos (IC 95 %: 16,91–32,78). La distribución del cambio resultó compatible con la normalidad según la prueba de Shapiro-Wilk ( $W = 0,959; p = 0,471$ ), habilitando la aplicación de la prueba t de Student para muestras relacionadas:  $t(21) = 6,14; p < 0,001$ . El contraste de robustez mediante Wilcoxon confirmó el resultado ( $W = 1,00; p < 0,001$ ). La d de Cohen alcanzó 1,31, valor que corresponde a un tamaño del efecto muy grande. La mejora fue prácticamente universal: 19 de los 22 estudiantes (86,4 %) aumentaron su puntaje, dos (9,1 %) mantuvieron el desempeño y uno (4,5 %) retrocedió. La correlación entre nota inicial y final fue prácticamente nula ( $\rho$  de Spearman =  $-0,072; p = 0,751$ ), indicando que la ganancia no dependió del nivel de partida.

La consistencia interna del instrumento, estimada mediante el coeficiente KR-20, alcanzó valores de 0,324 en el pretest y 0,261 en el post-test, ambos por debajo del umbral convencional de 0,70 utilizado para fines diagnósticos. Estos valores se interpretan en el apartado de limitaciones (sección Conclusiones) a la luz de las características multidimensionales del instrumento, del tamaño muestral y de la distribución de los ítems.

**Figura 1** Trayectorias individuales pre-post y distribución de las puntuaciones globales



**Fuente:** Elaboración propia (n = 22).

### **Cambio por dimensión cognitiva**

El análisis desagregado por dimensión cognitiva ofrece una lectura más fina del efecto observado. La proporción media de aciertos en la dimensión de comprensión conceptual pasó de 0,434 en el pretest a 0,621 en el post-test, lo que equivale a un incremento de 18,7 puntos porcentuales ( $d = 0,69$ , efecto grande). En la dimensión de desempeño algebraico, la proporción aumentó de 0,348 a 0,689, un incremento de 34,1 puntos porcentuales ( $d = 1,63$ , efecto muy grande). Ambos cambios son estadísticamente significativos, pero la magnitud del efecto en la dimensión algebraica casi duplica la registrada en la conceptual. La Tabla 3 sintetiza el contraste.

**Tabla 3** Comparación del desempeño por dimensión cognitiva ( $n = 22$ )

| Dimensión              | # ítems | $\hat{p}$ pre | $\hat{p}$ post | $\Delta$ (pp) | d    |
|------------------------|---------|---------------|----------------|---------------|------|
| Comprensión conceptual | 9       | 0,434         | 0,621          | +18,7         | 0,69 |
| Desempeño algebraico   | 6       | 0,348         | 0,689          | +34,1         | 1,63 |

**Nota.**  $\hat{p}$  representa la proporción media de aciertos; pp = puntos porcentuales; d = d de Cohen para diferencias. Elaboración propia.

El patrón asimétrico identificado constituye uno de los hallazgos centrales del análisis cognitivo y será retomado en la triangulación interna y en la discusión.

#### **Cambio por subtema y prueba de McNemar**

El análisis ítem por ítem mediante la prueba exacta de McNemar identificó seis ítems con mejora estadísticamente significativa: cuatro al nivel  $p < 0,01$  (ítems 1, 2, 3 y 10) y dos al nivel  $p < 0,05$  (ítems 9 y 13). La Tabla 4 sintetiza los principales resultados.

**Tabla 4** Cambio por subtema seleccionado y resultados de la prueba de McNemar ( $n = 22$ )

| Subtema                                     | $\hat{p}$ pre | $\hat{p}$ post | $\Delta$ (pp) | p McNemar    |
|---------------------------------------------|---------------|----------------|---------------|--------------|
| Dominio / restricciones (ítem 2)            | 0,318         | 0,864          | +54,5         | 0,004 **     |
| Composición de funciones (ítem 10)          | 0,227         | 0,773          | +54,5         | 0,002 **     |
| Identificación de función (ítem 1)          | 0,591         | 1,000          | +40,9         | 0,004 **     |
| Ecuaciones con valor absoluto (ítem 13)     | 0,227         | 0,636          | +40,9         | 0,012 *      |
| Transformación horizontal (ítem 9)          | 0,227         | 0,682          | +45,5         | 0,013 *      |
| <b>Imágenes y preimágenes (ítem 5)</b>      | <b>0,591</b>  | <b>0,500</b>   | <b>-9,1</b>   | <b>0,688</b> |
| <b>Intersecciones con los ejes (ítem 6)</b> | <b>0,636</b>  | <b>0,409</b>   | <b>-22,7</b>  | <b>0,267</b> |
| Función inyectiva (ítem 15)                 | 0,000         | 0,182          | +18,2         | 0,125        |

**Nota.** Selección de ítems con mayor magnitud de cambio. \*\*  $p < 0,01$ ; \*  $p < 0,05$ . Las filas resaltadas corresponden a subtemas con retroceso. Elaboración propia.

Los ítems con avance significativo cubrieron tanto contenidos conceptuales (identificación de funciones, transformaciones horizontales) como algebraicos (dominio de funciones racionales, composición de funciones, ecuaciones con valor absoluto). Los

mayores incrementos por subtema se concentraron en contenidos que combinan reglas operatorias con marcos conceptuales accesibles: dominio y restricciones (+54,5 pp), composición de funciones (+54,5 pp), ecuaciones con valor absoluto (+40,9 pp), transformaciones (+38,6 pp) e identificación de funciones (+30,3 pp). En contraste, dos ítems presentaron retroceso: el ítem 5 (imágenes y preimágenes), con -9,1 pp no significativos, y el ítem 6 (intersecciones con los ejes), con -22,7 pp. El pretest evaluó la condición que caracteriza los puntos de corte con el eje x y el post-test la condición análoga para el eje y, ambas validadas por juicio experto como equivalentes; el retroceso sugiere que una proporción del grupo construyó un esquema conceptual asimétrico: dominó la lectura del corte con x, pero no la transfirió al corte con y. El ítem 15 (función inyectiva) partió de un efecto piso absoluto (0,000) y alcanzó solo 0,182 en el post-test. Estos tres puntos remitieron al núcleo gráfico-simbólico del concepto de función.

### ***Confiabilidad y comportamiento psicométrico de la escala de actitudes***

La escala actitudinal mostró un comportamiento psicométrico globalmente robusto. El coeficiente alfa de Cronbach del conjunto de los 18 ítems, con el ítem inverso C5 recodificado, fue de 0,921, y el coeficiente omega de McDonald alcanzó 0,953, ambos por encima del umbral convencional de 0,70. Tres de las cuatro subescalas presentan consistencia interna alta o excelente, con todos los ítems mostrando correlaciones ítem-total positivas y de magnitud apropiada (entre 0,66 y 0,88). La Tabla 5 sintetiza los resultados de confiabilidad.

**Tabla 5** *Consistencia interna de la escala global y por subescala (n = 22)*

| Escala / Subescala   | # ítems   | $\alpha$ Cronbach | $\omega$ McDonald |
|----------------------|-----------|-------------------|-------------------|
| <b>Escala global</b> | <b>18</b> | <b>0,921</b>      | <b>0,953</b>      |
| Utilidad (TAM)       | 5         | 0,916             | 0,941             |
| Motivación           | 2         | 0,832             | 0,926             |
| Uso crítico          | 5         | 0,030             | 0,077             |
| Disposición          | 6         | 0,921             | 0,948             |

**Nota.** Interpretación de alfa:  $\geq 0,90$  excelente; 0,80–0,89 buena; 0,70–0,79 aceptable;  $< 0,70$  cuestionable. Elaboración propia.

La subescala de Uso crítico, con un alfa de 0,030 y un omega de 0,077, presenta un comportamiento psicométrico atípico que requiere explicación. La causa principal es un efecto de techo muy marcado: los ítems C2 («Comprender lo que responde la IA antes de usarlo en mi trabajo académico») y C3 («Usé la IA como apoyo, sin depender completamente de ella») recibieron el respaldo del 100 % del grupo, y el ítem C4 («Reflexioné sobre los límites y riesgos del uso de IA») alcanzó el 95,5 %. En estos tres ítems, todos los estudiantes respondieron 4 o 5, lo que reduce la desviación estándar a valores entre 0,35 y 0,58 y deja virtualmente sin varianza a la subescala.

### ***Valoración actitudinal por subescala***

Las cuatro subescalas se sitúan por encima del punto medio de la escala (3), lo que indica una valoración global positiva. El ordenamiento de mayor a menor es: Disposición ( $M = 4,49$ ;  $DE = 0,83$ ), Uso crítico ( $M = 4,31$ ;  $DE = 0,37$ ), Utilidad ( $M = 4,21$ ;  $DE = 1,01$ ) y Motivación ( $M = 3,45$ ;  $DE = 1,29$ ). La Tabla 6 sintetiza los descriptivos.

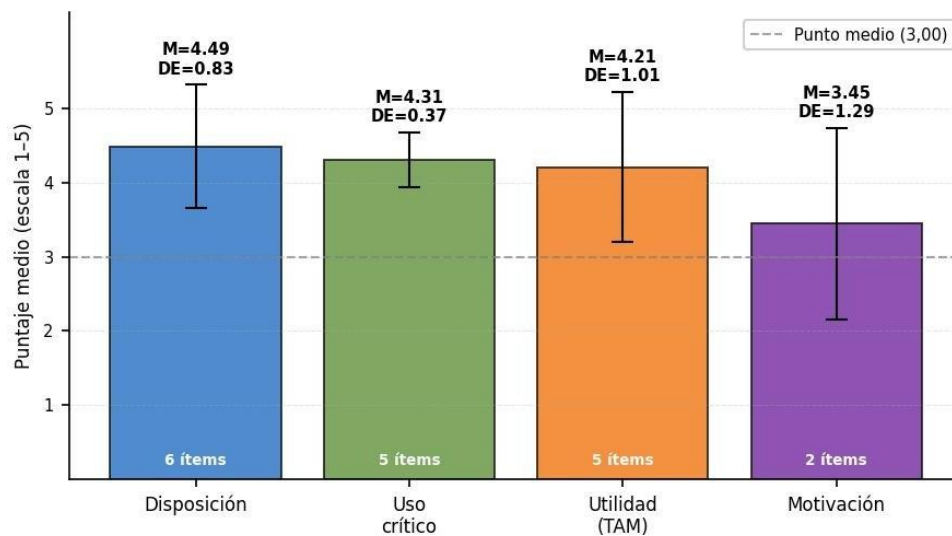
**Tabla 6** Estadísticos descriptivos por subescala ( $n = 22$ )

| Subescala      | # ítems | Media | DE   | Mín  | Máx  |
|----------------|---------|-------|------|------|------|
| Disposición    | 6       | 4,49  | 0,83 | 1,33 | 5,00 |
| Uso crítico    | 5       | 4,31  | 0,37 | 3,60 | 4,80 |
| Utilidad (TAM) | 5       | 4,21  | 1,01 | 1,40 | 5,00 |
| Motivación     | 2       | 3,45  | 1,29 | 1,00 | 5,00 |

**Nota.** Elaboración propia.

El hallazgo más contundente es la disposición a continuar utilizando IAGen: el 95,5 % de los estudiantes considera que la herramienta será útil en sus cursos universitarios futuros, el 90,9 % desea su incorporación en el curso ordinario de Matemática General y el 86,4 % la recomendaría a otros compañeros. En contraste, la subescala de Motivación es la única con un respaldo solo moderado: 54,5 % del grupo está de acuerdo en que el uso de IAGen motiva su interés por aprender matemáticas, y solo 36,4 % en que la presencia de la herramienta genera un ambiente más dinámico y participativo en clase. La IAGen es percibida, por tanto, como recurso útil y deseable, pero no necesariamente como dispositivo transformador del interés por la disciplina.

**Figura 2** Puntaje medio por subescala con barras de error



Fuente:

Elaboración propia. Las barras de error representan una desviación estándar; la línea discontinua marca el punto medio de la escala ( $n = 22$ ).

### Estructura relacional entre subescalas

Las correlaciones de Pearson entre las cuatro subescalas revelaron tres hallazgos. Primero, las subescalas de Utilidad, Motivación y Disposición correlacionaron muy fuertemente entre sí ( $r$  entre 0,747 y 0,881), configurando un macro-factor coherente de valoración positiva hacia la IAGen. Segundo, la subescala de Uso crítico apareció prácticamente desacoplada del resto ( $r$  entre  $-0,009$  y 0,177), funcionando como dimensión independiente. Tercero, la ausencia de correlación negativa entre Uso crítico y las otras subescalas indicó que la valoración positiva de la IAGen no resultó incompatible con mantener una postura crítica frente a ella. La Tabla 7 presenta la matriz completa.

**Tabla 7** Correlaciones de Pearson entre las cuatro subescalas ( $n = 22$ )

|             | Utilidad | Motivación | Uso crítico | Disposición |
|-------------|----------|------------|-------------|-------------|
| Utilidad    | 1,000    | 0,855      | 0,177       | 0,881       |
| Motivación  | 0,855    | 1,000      | $-0,009$    | 0,747       |
| Uso crítico | 0,177    | $-0,009$   | 1,000       | 0,127       |
| Disposición | 0,881    | 0,747      | 0,127       | 1,000       |

**Nota.** Elaboración propia.

### *Análisis exploratorio por subgrupos*

El análisis comparativo por subgrupos permitió explorar efectos diferenciales. La comparación por sexo no encontró diferencias significativas en las ganancias del pre/post ( $U = 62,50$ ;  $Z = 0,412$ ;  $p = 0,681$ ): hombres y mujeres mejoraron en proporciones casi idénticas (+24,76 y +25,00 puntos respectivamente).

El análisis por experiencia previa con IAGen contrastó a estudiantes con uso frecuente o muy frecuente durante la secundaria ( $n = 13$ ) frente a quienes reportaron uso escaso o nulo ( $n = 9$ ). Los estudiantes con alta experiencia valoraron significativamente mejor la utilidad de la herramienta y mostraron mayor disposición a su uso académico. La Tabla 8 presenta los estadísticos completos.

**Tabla 8** Comparación por experiencia previa con IA en secundaria mediante prueba U de Mann-Whitney ( $n = 22$ )

| Subescala   | Alta exp.<br>( $n=13$ ) | Baja exp.<br>( $n=9$ ) | Diferencia | U    | Z    | p       |
|-------------|-------------------------|------------------------|------------|------|------|---------|
| Utilidad    | 4,58                    | 3,69                   | +0,89      | 90,0 | 2,10 | 0,035 * |
| Motivación  | 3,77                    | 3,00                   | +0,77      | 79,0 | 1,37 | 0,165   |
| Uso crítico | 4,32                    | 4,29                   | +0,03      | 61,0 | 0,19 | 0,853   |
| Disposición | 4,82                    | 4,02                   | +0,80      | 93,0 | 2,30 | 0,022 * |

**Nota.** Alta experiencia: uso frecuente o muy frecuente en secundaria. Baja experiencia: uso escaso o nulo. U = estadístico de Mann-Whitney; Z = estadístico tipificado asociado ( $\mu_U = 58,5$ ;  $\sigma_U = 14,98$  para  $n_1 = 13$ ,  $n_2 = 9$ ). \*  $p < 0,05$ . Elaboración propia.

De manera similar, la frecuencia de uso actual fuera del aula se asoció significativamente con la disposición ( $U = 87$ ;  $Z = 1,90$ ;  $p = 0,010$ ) y, en el límite del umbral convencional, con la motivación ( $U = 76$ ;  $Z = 1,17$ ;  $p = 0,062$ ). Estos resultados sugirieron la presencia de una brecha por experiencia previa: la apropiación de la IAGen como recurso académico se sostuvo, en parte, sobre una alfabetización tecnológica adquirida con anterioridad a la intervención.

### ***Triangulación interna entre los dos instrumentos cuantitativos***

El aporte interpretativo distintivo del análisis surgió de la triangulación interna entre los dos instrumentos cuantitativos. La asimetría entre dimensiones cognitivas detectada en la prueba pre/post: ganancia de 34,1 pp en lo algebraico frente a 18,7 pp en lo conceptual, encontró correlato en la escala actitudinal. El ítem U2 («La IA facilitó el análisis de las transformaciones de funciones de forma visual y conceptual») presentó la media más baja de la subescala de Utilidad (3,95) y la mayor dispersión ( $DE = 1,40$ ), con cuatro respuestas en valores 1 o 2. Es decir, el ítem que indagó específicamente por el apoyo de la IA en el registro visual-conceptual de las transformaciones resultó el menos respaldado del conjunto. Esta convergencia entre menor avance objetivo en la dimensión conceptual y menor satisfacción autoinformada con el apoyo en el registro visual-conceptual se reforzó con la lectura por subtema: los retrocesos en intersecciones con los ejes ( $-22,7$  pp) y en imágenes y preimágenes ( $-9,1$  pp) pertenecieron al mismo núcleo gráfico-simbólico que el ítem U2 problematizó.

La codificación cualitativa de las respuestas abiertas del Likert, realizada mediante análisis temático inductivo (Braun y Clarke, 2006) e integrada con fines exclusivos de triangulación, añadió un tercer plano de evidencia: los estudiantes valoraron explícitamente la capacidad de la IAGen para proveer múltiples representaciones (siete menciones) y la metodología socrática (al menos seis menciones), pero ninguno de los 22 testimonios analizados mencionó avances específicos en lectura de intersecciones con los ejes ni en interpretación de la notación funcional sobre representaciones gráficas. La codificación cualitativa también reveló una demanda explícita y recurrente: la formación en ingeniería de prompts. Los estudiantes solicitaron, en sus propias palabras, «que nos enseñen a realizar bien los prompts» o el desarrollo de «una tabla fija de prompts para situaciones comunes».

## **Discusión**

### ***Cambio cognitivo sobre el concepto de función***

El primer objetivo específico: determinar la magnitud del cambio pre/post en el desempeño cognitivo sobre el concepto de función y su distribución por dimensión y subtema, encontró respuesta empírica en una mejora media de 24,85 puntos con un

tamaño del efecto muy grande ( $d$  de Cohen = 1,31). La magnitud cobra particular relevancia al considerarse que se alcanzó en una intervención de apenas cinco sesiones presenciales (12,5 horas totales), parámetro temporal sustancialmente más breve que el de la mayoría de intervenciones formativas con efectos comparables. La mejora prácticamente universal a nivel individual (86,4 % del grupo) y la independencia de la ganancia respecto del nivel de partida ( $\rho \approx 0$ ) reforzaron la consistencia del efecto: la intervención benefició al conjunto del grupo y no solo a un subgrupo con condiciones particulares de partida. Estos hallazgos son congruentes con la literatura emergente sobre mediación con IAGen en matemáticas (Yan et al., 2024).

El efecto fue, sin embargo, claramente asimétrico entre dimensiones: la dimensión algebraica casi duplicó el avance de la dimensión conceptual (+34,1 pp frente a +18,7 pp). Esta asimetría es interpretable a la luz de la distinción de Rittle-Johnson y Schneider (2015). El conocimiento procedimental tiende a beneficiarse de la práctica estructurada con retroalimentación inmediata, condición que la mediación con IAGen ofrece con particular eficiencia. El conocimiento conceptual, en cambio, exige reorganizaciones cognitivas más profundas vinculadas con la articulación entre representaciones (Duval, 2006), que requieren tiempos de maduración que probablemente excedieron las 12,5 horas de la intervención piloto y diseños didácticos que articulen explícitamente los registros, no que los presenten en secuencia.

Los dos subtemas con retroceso pre/post, intersecciones con los ejes (-22,7 pp) e imágenes y preimágenes (-9,1 pp), no constituyeron ruido estadístico; ambos remitieron al núcleo gráfico-simbólico del concepto de función y resultaron consistentes con los obstáculos epistemológicos documentados por Brousseau (1998) y refinados por Duval (2006) y Artigue y Winsløw (2010). El caso de las intersecciones con los ejes es particularmente revelador: el retroceso sugirió que una proporción del grupo construyó un esquema conceptual asimétrico: dominó la lectura del corte con el eje  $x$  pero presentó dificultades para transferir el razonamiento al corte con el eje  $y$ . La asimetría refleja precisamente la dificultad de articulación entre registros algebraico y gráfico que la teoría duvaliana identifica como núcleo del concepto.

### ***Valoración actitudinal y brecha por experiencia previa***

El segundo objetivo específico, caracterizar la valoración actitudinal del estudiantado y la estructura relacional entre subescalas, encontró respuesta en una valoración

marcadamente positiva del grupo, con Disposición ( $M = 4,49$ ) y Utilidad ( $M = 4,21$ ) como subescalas mejor puntuadas. Las correlaciones internas entre Utilidad, Motivación y Disposición ( $r$  entre 0,747 y 0,881) configuraron un macro-factor coherente que replicó el patrón clásico del Modelo de Aceptación Tecnológica de Davis (1989), según el cual la percepción de utilidad opera como predictor robusto de la intención de uso. Este resultado se alineó con observaciones recientes sobre aceptación de IAGen en educación superior (Almassaad et al., 2024; Bernabei et al., 2023; Strzelecki, 2024).

La subescala de Motivación se situó, sin embargo, por debajo del resto, con respaldo solo moderado: 54,5 % del grupo coincidió en que la IAGen motiva su interés por aprender matemáticas y solo 36,4 % en que genera un ambiente de aula más dinámico. Esta asimetría entre alta valoración instrumental y baja transformación motivacional es un hallazgo conceptualmente relevante que merece explicación. La literatura psicoeducativa contemporánea sobre motivación académica (Eccles y Wigfield, 2002) sostiene que la motivación intrínseca hacia una disciplina es un constructo identitario y de valor de tarea, que se construye en plazos largos a partir de experiencias acumuladas con el contenido, vínculos afectivos con la disciplina y percepciones de autoeficacia consolidadas durante la trayectoria escolar previa. No es, por tanto, una variable que pueda transformarse por la incorporación de una herramienta tecnológica, por útil que esta resulte. Los estudiantes percibieron la IAGen como un dispositivo instrumental valioso para resolver ejercicios concretos, pero esa valoración instrumental no se trasladó a una reconfiguración de su relación afectiva con las matemáticas. Una intervención de 12,5 horas puede transformar el desempeño cognitivo sobre un contenido específico y la valoración tecnológica de una herramienta; difícilmente puede transformar disposiciones motivacionales de fondo construidas a lo largo de la trayectoria escolar previa.

La subescala de Uso crítico exhibió un comportamiento psicométrico atípico ( $\alpha = 0,030$ ;  $\omega = 0,077$ ) atribuible a un efecto de techo severo: tres ítems concentraron el respaldo del 95–100 % del grupo. Más allá de la limitación psicométrica, este patrón abrió interrogantes sobre la posible influencia de sesgos de deseabilidad social en ítems que evalúan prácticas académicamente valoradas; futuras iteraciones podrían sustituir ítems de acuerdo por ítems de frecuencia conductual.

Particular atención merece la brecha por experiencia previa con IAGen. Los estudiantes con uso frecuente de la herramienta durante la secundaria valoraron significativamente

mejor su utilidad académica ( $p = 0,035$ ) y mostraron mayor disposición a su uso ( $p = 0,022$ ) que sus pares sin esa experiencia. La brecha también se expresó en la frecuencia de uso fuera del aula como predictor de la disposición ( $p = 0,010$ ). El hallazgo plantea una cuestión estructural de equidad: la integración pedagógica de la IAGen puede estar parcialmente sostenida sobre una alfabetización tecnológica previa que el sistema educativo aún no garantiza de manera homogénea. La demanda explícita de los estudiantes, documentada en las respuestas abiertas, de formación específica en ingeniería de prompts es coherente con esta lectura.

### ***Triangulación interna y patrón diferencial del aprendizaje***

El tercer objetivo específico: identificar, mediante triangulación interna entre ambos instrumentos cuantitativos, patrones diferenciales del aprendizaje que orientaran el rediseño didáctico, encontró respuesta en una triple convergencia que articuló los hallazgos previos sobre un mismo núcleo conceptual. Primero, el dato cognitivo objetivo evidenció una asimetría a favor del avance procedimental sobre el conceptual y dos retrocesos localizados en el núcleo gráfico-simbólico (intersecciones con los ejes e imágenes y preimágenes). Segundo, el autoinforme cerrado de la escala Likert identificó el ítem U2, el único que indagó específicamente por el apoyo de la IA en el análisis visual-conceptual de transformaciones, como el menos respaldado de la subescala de Utilidad. Tercero, la codificación temática inductiva de las respuestas abiertas reveló ausencia total de menciones a avances en lectura gráfica, frente a múltiples menciones del valor instrumental, las representaciones múltiples y la metodología socrática.

Esta triple convergencia sostiene la interpretación principal del estudio: la mediación con IAGen bajo prompt de tutor socrático produjo efectos sustanciales sobre los contenidos procedimentales y sobre la valoración global de la herramienta, pero encontró un límite identificable en el registro gráfico-simbólico del concepto de función, registro que requiere tratamiento didáctico específico que la herramienta por sí sola no logra cubrir. El patrón diferencial identificado es genuinamente pedagógico, no atribuible al instrumento ni al diseño, en tanto los tres planos de evidencia convergen sobre el mismo núcleo conceptual, validado por juicio experto. Esta convergencia metodológica responde a la lógica de triangulación interna entre componentes cuantitativos del diseño mixto (Creswell y Plano Clark, 2018).

---

## Conclusiones

### Limitaciones del estudio

Cuatro limitaciones acotaron el alcance de los hallazgos. Primero, el diseño pre-experimental con grupo único y sin comparación impidió atribuir la mejora observada exclusivamente a la mediación con IAGen; factores como la maduración natural, el efecto de la exposición previa al pretest, la motivación asociada al inicio del semestre y la presencia del docente-tutor concurrente pudieron explicar parte del cambio. Segundo, el tamaño muestral efectivo ( $n = 22$ ) restringió la potencia estadística de los contrastes por ítem y por subgrupo. Tercero, los coeficientes KR-20 de la prueba cognitiva (0,324 en pretest; 0,261 en post-test) resultaron notablemente inferiores al umbral convencional de 0,70; esta condición no constituyó un defecto del instrumento sino consecuencia directa de la naturaleza multidimensional del instrumento de diagnóstico (Sijtsma, 2009), del tamaño muestral pequeño y de la presencia de ítems con efectos extremos de piso y techo, sin comprometer la validez de contenido del instrumento, validada independientemente por juicio experto. Cuarto, la subescala de Uso crítico presentó limitaciones psicométricas significativas atribuibles al efecto de techo documentado, que configuraron una oportunidad de rediseño del instrumento.

### Conclusiones principales

El taller de nivelación mediado por IAGen con prompt de tutor socrático produce un cambio cognitivo de gran magnitud sobre el concepto de función en una intervención de 12,5 horas, con tamaño del efecto muy grande ( $d = 1,31$ ) y mejora prácticamente universal a nivel individual (86,4 %). El cambio se distribuye de manera asimétrica entre dimensiones, con ganancias casi duplicadas en la dimensión procedimental-algebraica frente a la conceptual, y presenta retrocesos localizados en el núcleo gráfico-simbólico que constituyen señales pedagógicas precisas sobre los límites de la mediación.

Por otra parte, la valoración actitudinal resulta marcadamente positiva, con disposición destacada ( $M = 4,49$ ) a continuar utilizando IAGen en cursos universitarios futuros y percepción de utilidad alta ( $M = 4,21$ ). La estructura relacional reproduce el patrón clásico del Modelo de Aceptación Tecnológica. La subescala de Motivación presenta, sin embargo, respaldo solo moderado ( $M = 3,45$ ), lo que indica que la aceptación instrumental de la herramienta no se traslada automáticamente a una transformación

motivacional hacia la disciplina en una intervención breve. Una brecha estadísticamente significativa por experiencia previa con IAGen en utilidad ( $p = 0,035$ ) y disposición ( $p = 0,022$ ) configura una cuestión estructural de equidad relevante para el diseño didáctico futuro.

Además, la triple convergencia entre dato cognitivo objetivo, autoinforme cerrado en el ítem U2 y codificación cualitativa de respuestas abiertas sostiene un patrón diferencial consistente: la mediación con IAGen bajo prompt de tutor socrático resulta eficaz para contenidos procedimentales y para la aceptación tecnológica general, pero encuentra un límite identificable en el registro gráfico-simbólico del concepto de función, registro que requiere tratamiento didáctico específico no sustituible por la herramienta.

La unidad didáctica completa que articula ocho ejercicios estructurados, la secuencia PRVR (Planear–Resolver–Verificar–Reflexionar), el prompt de tutor socrático, la triangulación representacional con GeoGebra y Desmos, la bitácora reflexiva y el marco ético de la Regla de Oro constituye un dispositivo formativo eficaz para mitigar el rezago cognitivo en contenidos procedimentales del concepto de función dentro de talleres breves de nivelación universitaria. La eficacia del dispositivo es selectiva: opera robustamente sobre el aprendizaje procedimental y sobre la aceptación tecnológica, pero no sustituye el tratamiento didáctico explícito de los obstáculos epistemológicos asociados a la articulación entre registros de representación. Este hallazgo aporta evidencia empírica relevante a la literatura emergente sobre integración pedagógica de la IA generativa en educación matemática universitaria latinoamericana.

### **Recomendaciones e implicaciones**

Los hallazgos sugieren cuatro recomendaciones para el diseño de intervenciones formativas mediadas por IAGen en educación matemática universitaria.

Primera, el rediseño didáctico de futuras iteraciones del taller debería contemplar tratamientos específicos para la articulación entre el registro gráfico y el simbólico del concepto de función, con énfasis explícito en la enseñanza simétrica de los cortes con ambos ejes coordenados y en la lectura de la notación funcional sobre representaciones gráficas. Segunda, el instrumento de Uso crítico podría rediseñarse sustituyendo los ítems de acuerdo por ítems de frecuencia conductual, lo que reduciría el efecto de techo, mitigaría el sesgo de deseabilidad social y permitiría una estimación más estable de la

consistencia interna. Tercera, conviene incorporar un módulo introductorio sobre alfabetización en IAGen e ingeniería de prompts al inicio del taller, atendiendo la demanda explícita del estudiantado y reduciendo la brecha significativa por experiencia previa identificada. Cuarta, la presente investigación abre una veta prometedora de análisis correlacional avanzado entre las ganancias cognitivas objetivas y las valoraciones actitudinales a nivel individual: la incorporación de un identificador anonimizado común entre ambos instrumentos habilitaría una agenda metodológicamente rica que incluye análisis de moderación por experiencia previa, modelos de mediación con la motivación como variable intermedia, e identificación de perfiles latentes mediante técnicas multivariadas como conglomerados k-means o clases latentes. Esta línea constituye una extensión metodológica natural del estudio piloto.

La continuación de esta línea de investigación con muestras ampliadas, diseños cuasi-experimentales y mediciones intermedias que permitan documentar la trayectoria del aprendizaje a lo largo de la intervención permitirá confirmar, refinar o matizar los hallazgos del presente estudio piloto. La evidencia aquí reportada constituye, en cualquier caso, una base empírica relevante para el diseño de futuros dispositivos de nivelación matemática mediados por inteligencia artificial generativa.

### **Referencias bibliográficas**

- Almassaad, A., Alajlan, H., y Alebaikan, R. (2024). Student perceptions of generative artificial intelligence: Investigating utilization, benefits, and challenges in higher education. <https://doi.org/10.3390/systems12100385>
- Artigue, M., y Winsløw, C. (2010). International comparative studies on mathematics education: A viewpoint from the anthropological theory of the didactic. *Recherches en Didactique des Mathématiques*, 30(1), 47–82.
- Bernabei, M., Colabianchi, S., Falegnami, A., y Costantino, F. (2023). Students' use of large language models in engineering education: A case study on technology acceptance, perceptions, efficacy, and detection chances. *Computers and Education: Artificial Intelligence*, 5, 100172. <https://doi.org/10.1016/j.caeai.2023.100172>
- Braun, V., y Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brousseau, G. (1998). *Théorie des situations didactiques*. La Pensée Sauvage.

- Chan, C. K. Y., y Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2.<sup>a</sup> ed.). Lawrence Erlbaum Associates.
- Creswell, J. W., y Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3.<sup>a</sup> ed.). SAGE Publications.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Dunn, T. J., Baguley, T., y Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics*, 61(1–2), 103–131. <https://doi.org/10.1007/s10649-006-0400-z>
- Eccles, J. S., y Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Escobar-Pérez, J., y Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6(1), 27–36.
- Hernández-Sampieri, R., y Mendoza, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Pintrich, P. R., Smith, D. A. F., García, T., y McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. National Center for Research to Improve Postsecondary Teaching and Learning. <https://eric.ed.gov/?id=ED338122>
- Rittle-Johnson, B., y Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. En R. Cohen Kadosh y A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 1118–1134). Oxford University Press. <https://doi.org/10.1093/oxfordhnb/9780199642342.013.014>
- Shapiro, S. S., y 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>

- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120. <https://doi.org/10.1007/s11336-008-9101-0>
- Strzelecki, A. (2024). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*, 32(9), 5142–5155. <https://doi.org/10.1080/10494820.2023.2209881>
- Sweller, J., van Merriënboer, J. J. G., y Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., y Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>

#### **Contribuciones de los autores:**

**Melvin Ramírez Bogantes:** Conceptualización del estudio, supervisión del proyecto, validación científica y revisión crítica del manuscrito.

**Esteban Quesada Ávila:** Diseño metodológico, validación de contenidos, revisión crítica, corrección académica y gestión del proyecto.

**Francine Araya López:** Investigación, búsqueda y recopilación bibliográfica, análisis de la literatura científica y redacción del borrador original.

#### **Conflicto de intereses:**

Los autores declaran que no existe conflicto de interés