



Alfabetización en IA y experiencia subjetiva con un tutor virtual basado en GPT en estudiantes de Kinesiología: estudio transversal.

AI literacy and subjective experience with a GPT-based virtual tutor in undergraduate physiotherapy students: a cross-sectional study.

Italo Campos Montenegro^{1*}, Maximiliano Martínez Mena²

¹ Carrera de Kinesiología, Facultad de Ciencias de la Vida, Universidad Viña del Mar, Viña del Mar, Chile, italo.campos@uvm.cl, <https://orcid.org/0009-0007-0325-3344>

² Carrera de Kinesiología, Facultad de Ciencias de la Vida, Universidad Viña del Mar, Viña del Mar, Chile; maximiliano.martinez@uvm.cl; <https://orcid.org/0009-0006-6181-0324>

* Correspondencia: italo.campos@uvm.cl

Recibido: 7/5/26; Aceptado: 4/6/26; Publicado: 6/6/26

Resumen.

Objetivo: Analizar la asociación entre condiciones autorreportadas de uso de un tutor virtual basado en GPT y la experiencia subjetiva de aprendizaje en estudiantes de Kinesiología, considerando la motivación académica como resultado primario. **Métodos:** Estudio observacional analítico de corte transversal en una cohorte 2025 de estudiantes de Kinesiología expuestos a un tutor personalizado GPT implementado como apoyo complementario al aprendizaje. Se aplicó una encuesta estructurada para medir siete dimensiones subjetivas de experiencia: motivación, autoeficacia, facilitación del aprendizaje, reducción de ansiedad/estrés académico, calidad percibida, interacción/usabilidad y aceptación/valor percibido. Se evaluaron tres condiciones autorreportadas de uso: intensidad de uso del tutor, alfabetización en IA autorreportada mediante ítem único y conductas de verificación académica. Las asociaciones se analizaron mediante correlaciones bivariadas de Pearson. El resultado primario fue motivación académica; los resultados secundarios se ajustaron por multiplicidad mediante FDR de Benjamini-Hochberg. **Resultados:** Se incluyeron 23 estudiantes. Las tres condiciones autorreportadas de uso se asociaron positivamente con la motivación académica: intensidad de uso ($r = 0,703$; IC95% 0,409 a 0,865; $p < 0,001$), alfabetización en IA autorreportada mediante ítem único ($r = 0,767$; IC95% 0,519 a 0,896; $p < 0,001$) y conductas de verificación académica ($r = 0,773$; IC95% 0,530 a 0,899; $p < 0,001$). En los resultados secundarios, las asociaciones fueron positivas y mantuvieron significación estadística tras el ajuste FDR. La nota final del curso se describió únicamente como referencia contextual, sin inferencias de efectividad académica. **Conclusiones:** En esta cohorte pequeña, las condiciones autorreportadas de uso de un tutor virtual basado en GPT se asociaron positivamente con una experiencia subjetiva más favorable. Estos hallazgos deben interpretarse como asociaciones exploratorias, contextuales y basadas en autorreporte, sin inferir causalidad ni efectividad sobre el rendimiento académico. Futuros estudios deberán considerar muestras mayores, diseños longitudinales o controlados, instrumentos validados y registros objetivos de interacción.

Palabras clave: inteligencia artificial generativa; alfabetización en IA; tutor virtual; educación en salud; experiencia subjetiva; motivación académica.

Abstract.

Objective: To analyze the association between self-reported conditions of use of a GPT-based virtual tutor and subjective learning experience in undergraduate physiotherapy students, considering academic motivation as the primary outcome. **Methods:** A cross-sectional analytical observational study was conducted in a 2025 cohort of undergraduate physiotherapy students exposed to a customized GPT tutor implemented as a complementary learning resource. A structured survey was used to assess seven subjective experience dimensions: motivation, self-efficacy, perceived learning facilitation, reduction of academic anxiety/stress, perceived quality, interaction/usability, and acceptance/perceived value. Three self-reported conditions of use were evaluated: tutor usage intensity, self-reported AI literacy measured through a single item, and academic verification behaviors. Associations were analyzed using bivariate Pearson correlations. Academic motivation was considered the primary outcome; secondary outcomes were adjusted for multiplicity using the Benjamini-Hochberg false discovery rate. **Results:** Twenty-three students were included. The three self-reported conditions of use were positively associated with academic motivation: usage intensity ($r = 0.703$; 95% CI 0.409 to 0.865; $p < 0.001$), self-reported AI literacy measured through a single item ($r = 0.767$; 95% CI 0.519 to 0.896; $p < 0.001$), and academic verification behaviors ($r = 0.773$; 95% CI 0.530 to 0.899; $p < 0.001$). For secondary outcomes, associations were positive and remained statistically significant after FDR adjustment. Final course grade was described only as contextual information, without drawing inferences about academic effectiveness. **Conclusions:** In this small cohort, self-reported conditions of use of a GPT-based virtual tutor were positively associated with a more favorable subjective experience. These findings should be interpreted as exploratory, contextual, and self-reported associations, without inferring causality or effectiveness on academic performance. Future studies should consider larger samples, longitudinal or controlled designs, validated instruments, and objective interaction logs.

Keywords: generative artificial intelligence; AI literacy; virtual tutor; health professions education; subjective experience; academic motivation.

1. Introducción

La inteligencia artificial generativa (IAgen), y en particular los modelos de lenguaje de gran tamaño (LLM), se ha incorporado con rapidez a la educación superior como una herramienta capaz de ofrecer apoyo académico bajo demanda, personalizable y disponible de forma continua. En este escenario, plataformas como ChatGPT han comenzado a operar de facto como tutores virtuales para resolver dudas, organizar contenidos, retroalimentar tareas y apoyar el estudio autónomo. La literatura reciente sugiere que estas herramientas pueden aportar beneficios en percepción del aprendizaje, motivación y algunos indicadores de desempeño, aunque con una heterogeneidad importante entre contextos, diseños instruccionales y formas de uso (1-3).

En paralelo, la incorporación de LLM en educación superior ha sido discutida en contextos internacionales diversos, incluyendo experiencias y estudios desarrollados en Norteamérica, África, Europa y América Latina. Esta expansión ha intensificado preocupaciones sobre uso indebido, sustitución del esfuerzo cognitivo, integridad académica y dependencia de respuestas generadas automáticamente. Estas tensiones son especialmente relevantes cuando las herramientas generativas se integran en tareas evaluativas o en actividades donde el objetivo formativo requiere elaboración, razonamiento y juicio propio, más que simple acceso a información. Por ello, el debate actual ya no se centra solo en si estas herramientas están disponibles o son aceptadas por los/las estudiantes, sino en cómo se utilizan y bajo qué condiciones ese uso resulta formativamente valioso (4-6). Desde esta perspectiva, variables como la capacidad para formular prompts útiles, verificar información y

orientar la interacción hacia metas académicas plausibles podrían ser más relevantes que la frecuencia bruta de uso. Esta distinción es particularmente importante en educación superior en salud, donde el aprendizaje exige no solo acceso a información, sino también selección crítica, razonamiento aplicado y uso responsable del conocimiento. En esta línea, algunos estudios y revisiones recientes han sugerido que la efectividad de los tutores conversacionales depende en gran medida del diseño pedagógico y del tipo de interacción que promueven, más que de su mera disponibilidad o de un aumento inespecífico de su uso (3, 7-8).

Pese al crecimiento de la literatura sobre IA generativa en educación, sigue siendo limitada la evidencia aplicada en contextos reales de formación en salud que examine simultáneamente la experiencia subjetiva del estudiantado y las condiciones específicas de uso del tutor. Buena parte de los estudios disponibles describe aceptación, utilidad percibida o efectos globales del uso de chatbots, pero con menor atención a si el valor percibido del recurso depende más de la cantidad de uso o de la calidad o competencia con que ese uso se lleva a cabo. Esta brecha es relevante, porque una implementación pedagógicamente responsable no requiere solo más interacción con la herramienta, sino una comprensión más precisa de qué rasgos del uso se asocian con experiencias de aprendizaje más favorables (1-2, 6).

En este contexto, el presente estudio analiza la implementación de un tutor virtual personalizado basado en GPT en estudiantes de Kinesiología de la Universidad Viña del Mar durante 2025, en la asignatura de Semiología en Traumatología e Imagenología. El objetivo fue evaluar la asociación entre tres condiciones de uso autorreportadas —intensidad de uso, alfabetización en IA y conductas de verificación académica— y distintas dimensiones de la experiencia subjetiva asociada al tutor, considerando la motivación académica como resultado primario. Se planteó como hipótesis que mayores niveles de alfabetización en IA y de conductas de verificación se asociarían con una experiencia más favorable, mientras que la intensidad de uso desempeñaría un papel menos consistente. Con ello, el estudio busca aportar evidencia aplicada para orientar el diseño e implementación de tutores basados en modelos de lenguaje en contextos universitarios de formación en salud.

2. Métodos

2.1 Diseño del estudio

Se realizó un estudio observacional analítico de corte transversal, basado en una encuesta aplicada a estudiantes de pregrado expuestos durante 2025 a un tutor virtual personalizado basado en GPT. El objetivo analítico fue examinar la asociación entre condiciones autorreportadas de uso del tutor y distintas dimensiones subjetivas de la experiencia de aprendizaje. La población objetivo correspondió a la cohorte 2025 de estudiantes inscritos en la asignatura de Semiología en Traumatología e Imagenología, de la carrera de Kinesiología de la Universidad Viña del Mar (Chile). El tutor se puso a disposición de la totalidad de estudiantes matriculados/as durante el semestre académico como recurso complementario y de uso voluntario. Fueron elegibles quienes: (i) estuvieron matriculados/as en la asignatura durante 2025; (ii) aceptaron participar mediante consentimiento informado escrito; y (iii) respondieron la encuesta. Durante el periodo de recolección, el número total de estudiantes elegibles fue de 31. Respondieron la encuesta 23 estudiantes, lo que correspondió a una tasa de respuesta de 74%. La unidad de análisis fue el/la estudiante. Para los análisis correlacionales se consideraron los 23 registros disponibles, sin imputación de datos faltantes. La nota final del curso se analizó únicamente como referencia contextual y con carácter exploratorio.

2.2 Tutor virtual e implementación

Se implementó un tutor virtual personalizado GPT en ChatGPT a través de una cuenta plus (modelo 4o), configurado con materiales relevantes de la asignatura, incluyendo documentos de planificación docente. El recurso se presentó como apoyo al trabajo autónomo y no reemplazó la docencia regular. Antes del período de uso se realizó una capacitación general de 30 minutos sobre sus posibilidades de uso y se indicó explícitamente utilizarlo como apoyo formativo, no como sustituto de evaluaciones o productos académicos.

2.3 Instrumento y procedimiento

Se aplicó una encuesta estructurada online con escalas Likert de 5 puntos para medir siete dimensiones subjetivas de experiencia con el tutor: motivación, autoeficacia, facilitación del aprendizaje, reducción de ansiedad/estrés académico, calidad percibida del tutor, interacción/usabilidad y aceptación/valor percibido. La dimensión aceptación/valor percibido integró ítems referidos a utilidad percibida del tutor para el aprendizaje, intención de continuidad de uso, mejora percibida de la experiencia de aprendizaje y beneficios percibidos más allá de la asignatura. El instrumento fue de elaboración propia y se construyó ad hoc para este estudio a partir de dimensiones conceptuales recurrentes en la literatura sobre experiencia subjetiva de uso de tutores conversacionales y apoyo al aprendizaje. Antes de su aplicación, los ítems fueron sometidos a juicio de expertos ($n = 3$) para evaluar claridad, pertinencia y coherencia con los constructos propuestos, realizándose ajustes de redacción y contenido. Dado su carácter específico para esta implementación, el instrumento no fue sometido a una validación psicométrica formal previa, por lo que sus resultados deben interpretarse como medidas autorreportadas de uso y experiencia en esta cohorte.

2.4 Variables y operacionalización

Se definieron tres predictores principales autorreportados:

- **Intensidad de uso del tutor:** índice compuesto derivado de tres indicadores ordinales: frecuencia de uso durante el semestre, uso en momentos clave y duración típica de sesión. Cada indicador fue transformado a una puntuación estandarizada (z-score) y posteriormente se calculó el promedio simple de las tres puntuaciones estandarizadas para obtener un índice sintético de intensidad de uso. Valores más altos indicaron mayor intensidad relativa de uso del tutor.
- **Alfabetización en IA:** ítem único Likert de 1 a 5 que evaluó la autopercepción de capacidad para formular prompts claros y específicos con fines académicos.
- **Conductas de verificación académica:** promedio de tres ítems orientados a verificación de información, detección de errores o alucinaciones y explicitación de fuentes o proceso cuando correspondía.

La motivación académica se definió a priori como resultado primario. Los resultados secundarios fueron autoeficacia, facilitación del aprendizaje, reducción de ansiedad/estrés académico, calidad percibida, interacción/usabilidad y aceptación/valor percibido. Cada dimensión se calculó como el promedio de los ítems correspondientes, de modo que puntajes mayores indicaran una experiencia más favorable.

2.5 Análisis estadístico

Se realizaron análisis descriptivos de las variables de uso del tutor y de los puntajes por dimensión mediante medias y desviaciones estándar. La consistencia interna de cada dimensión se estimó mediante alfa de Cronbach, interpretado como un indicador de coherencia interna de los ítems en esta cohorte y no como evidencia de validación psicométrica formal del instrumento.

Considerando el tamaño muestral reducido, el carácter exploratorio del estudio y la naturaleza autorreportada de las variables, las asociaciones entre las condiciones de uso del tutor y las dimensiones subjetivas de experiencia se analizaron mediante correlaciones bivariadas de Pearson. Dado que las dimensiones subjetivas fueron calculadas como promedios de ítems tipo Likert, se trataron como variables continuas aproximadas para el análisis correlacional. Esta decisión buscó mantener el objetivo asociativo del estudio mediante una estrategia analítica más parsimoniosa y proporcional al tamaño de la cohorte, evitando estimaciones multivariadas potencialmente inestables.

La motivación académica se mantuvo como resultado primario. Los resultados secundarios fueron autoeficacia, facilitación del aprendizaje, reducción de ansiedad/estrés académico, calidad percibida, interacción/usabilidad y aceptación/valor percibido. Para los resultados secundarios se controló la multiplicidad mediante el procedimiento FDR de Benjamini-Hochberg, aplicado sobre la familia de contrastes correspondiente a las asociaciones entre las tres condiciones de uso del tutor y las seis dimensiones secundarias de experiencia subjetiva.

La nota final del curso se examinó únicamente como análisis exploratorio, sin establecer inferencias de efectividad académica del tutor virtual. Los análisis fueron realizados en Python. El uso de modelos de lenguaje se limitó al apoyo en generación de código, sin automatizar decisiones analíticas, selección de resultados ni interpretación final. Dado que se trabajó con la cohorte completa disponible de una asignatura, el estudio utilizó una muestra por conveniencia y no se realizó cálculo formal de tamaño muestral. Por el carácter voluntario tanto de la participación como del uso del tutor, se consideraron como posibles fuentes de sesgo la autoselección, el autorreporte y la deseabilidad social. Para mitigar parcialmente estos riesgos, se informó a los/las estudiantes que su decisión de participar o no participar no tendría consecuencias académicas, y la base analítica fue pseudoanonimizada antes del análisis.

2.6 Consideraciones éticas

El estudio fue realizado de acuerdo con los principios éticos aplicables a investigación en seres humanos y en conformidad con la Declaración de Helsinki, tal como exige la política editorial de esta revista para trabajos originales con participación humana. La investigación contó con aprobación del Comité Ético Científico de la Universidad Viña del Mar (CEC-UVM), Acta N.º 55-25, 12 de agosto de 2025. La participación fue voluntaria y mediada por consentimiento informado previo. Se informó a los/las estudiantes que su decisión de participar o no participar no tendría consecuencias académicas. Para fines de verificación de elegibilidad y vinculación con el registro académico, se solicitó identificación nominal en la etapa inicial de recolección; posteriormente, la base fue pseudoanonimizada, eliminando identificadores directos y restringiendo el acceso a la información nominativa al equipo investigador exclusivamente para fines de verificación y vinculación. La encuesta fue administrada mediante un formulario de Outlook, y los resultados fueron accesibles únicamente para el investigador principal. Los datos fueron utilizados exclusivamente con fines de investigación, tratados con resguardo de confidencialidad y reportados de forma agregada.

3. Resultados

3.1 Participantes y base analítica

Durante el periodo de estudio, 31 estudiantes fueron elegibles. De ellos, 23 respondieron la encuesta, lo que correspondió a una tasa de respuesta del 74%. Para los análisis correlacionales se consideraron los 23 registros disponibles. No se realizó imputación de datos faltantes. El análisis de la nota final del curso se efectuó con los 23 registros disponibles y fue interpretado únicamente con carácter exploratorio.

3.2 Uso del tutor virtual y perfil general de experiencia

Tal como se expresa en la tabla 1, los indicadores de uso del tutor se ubicaron en rangos intermedios a altos: la frecuencia general de uso fue $3,22 \pm 1,04$, el uso en momentos clave $3,35 \pm 1,15$ y la duración típica de las sesiones $3,04 \pm 1,43$, en escalas ordinales de 1 a 5. Además de los indicadores de intensidad de uso, la alfabetización en IA autorreportada presentó una media de $3,91 \pm 0,79$, mientras que las conductas de verificación académica presentaron una media de $3,88 \pm 0,78$. Ambos indicadores se expresaron en escalas de 1 a 5, donde valores mayores indicaron mayor autopercepción de competencia o mayor frecuencia de conductas de verificación.

Tabla 1. Descriptivos de uso y predictores principales.

Indicador	n	Media	DE
Frecuencia de uso del tutor virtual durante el semestre	23	3,22	1,04
Frecuencia de uso en momentos clave (fuera de horario/antes de entrega)	23	3,35	1,15
Duración típica de las sesiones de uso	23	3,04	1,43
Alfabetización en IA autorreportada	23	3,91	0,79
Conductas de verificación académica	23	3,88	0,78

DE = desviación estándar. Los indicadores se expresan en escalas ordinales o tipo Likert de 1 a 5; puntajes mayores indican mayor frecuencia, duración, autopercepción de competencia o frecuencia de conductas de verificación.

Las dimensiones subjetivas mostraron puntajes promedio cercanos al rango alto de la escala. Las medias se ubicaron entre 3,76 y 4,01 puntos (tabla 2). La mayor media se observó en facilitación del aprendizaje ($4,01 \pm 0,95$), seguida por interacción/usabilidad ($4,00 \pm 0,97$), autoeficacia ($3,98 \pm 0,99$), motivación ($3,90 \pm 1,01$), calidad percibida del tutor ($3,90 \pm 0,94$), reducción de ansiedad/estrés académico ($3,78 \pm 1,03$) y aceptación/valor percibido ($3,76 \pm 0,97$). En esta muestra, las dimensiones mostraron alta consistencia interna estimada mediante alfa de Cronbach (0,92 a 0,97), lo que respalda la coherencia interna de los ítems en esta cohorte, sin constituir por sí mismo una validación psicométrica completa del instrumento.

Tabla 2. Puntuaciones (media \pm DE) y consistencia interna (α de Cronbach) (n = 23).

Dimensión	k	Media	DE	α
Motivación	4	3,90	1,01	0,97
Autoeficacia	4	3,98	0,99	0,96
Facilitación del aprendizaje	3	4,01	0,95	0,94
Reducción de ansiedad/estrés académico	4	3,78	1,03	0,95
Calidad percibida del tutor	4	3,90	0,94	0,92
Interacción/usabilidad	4	4,00	0,97	0,97
Aceptación/valor percibido	4	3,76	0,97	0,93

DE = desviación estándar; α = alfa de Cronbach.

3.3 Resultado primario: motivación académica

Las tres condiciones autorreportadas de uso del tutor se asociaron positivamente con la motivación académica (tabla 3). La correlación más alta se observó para conductas de verificación académica ($r = 0,773$; IC95% 0,530 a 0,899; $p < 0,001$), seguida por alfabetización en IA autorreportada mediante ítem único ($r = 0,767$; IC95% 0,519 a 0,896; $p < 0,001$) e índice de intensidad de uso ($r = 0,703$; IC95% 0,409 a 0,865; $p < 0,001$).

3.4 Resultados secundarios

Las asociaciones exploratorias entre las condiciones autorreportadas de uso del tutor y las dimensiones secundarias de experiencia subjetiva fueron positivas (tabla 4). Tras el control por multiplicidad mediante FDR de Benjamini-Hochberg, todas las asociaciones secundarias evaluadas mantuvieron significación estadística ($q < 0,001$). Las correlaciones observadas se ubicaron entre $r = 0,642$ y $r = 0,873$. Las asociaciones de mayor magnitud se observaron entre alfabetización en IA autorreportada mediante ítem único y aceptación/valor percibido ($r = 0,873$; IC95% 0,721 a 0,945), alfabetización en IA autorreportada mediante ítem único y reducción de ansiedad/estrés académico ($r = 0,826$; IC95% 0,627 a 0,924), y conductas de verificación académica con aceptación/valor percibido ($r = 0,836$; IC95% 0,646 a 0,928). Estos resultados deben interpretarse como asociaciones exploratorias entre variables autorreportadas, sin establecer direccionalidad causal ni independencia entre condiciones de uso.

Tabla 3. Correlaciones bivariadas entre las condiciones de uso del tutor y la motivación académica.

Resultado primario	Variable asociada	r de Pearson	IC95%	p
Motivación	Índice de intensidad de uso	0,703	[0,409; 0,865]	<0,001
Motivación	Alfabetización en IA	0,767	[0,519; 0,896]	<0,001
Motivación	Conductas de verificación académica	0,773	[0,530; 0,899]	<0,001

Tabla 4. Correlaciones entre las condiciones de uso del tutor y las dimensiones secundarias de experiencia subjetiva.

Resultado secundario	Variable asociada	r de Pearson	IC95%	p nominal	q FDR
Autoeficacia	Índice de intensidad de uso	0,716	[0,431; 0,871]	<0,001	<0,001
	Alfabetización en IA	0,737	[0,467; 0,882]	<0,001	<0,001
	Conductas de verificación	0,726	[0,449; 0,876]	<0,001	<0,001
Facilitación del aprendizaje	Índice de intensidad de uso	0,724	[0,445; 0,875]	<0,001	<0,001
	Alfabetización en IA	0,746	[0,481; 0,886]	<0,001	<0,001
	Conductas de verificación	0,763	[0,511; 0,894]	<0,001	<0,001
Reducción de ansiedad/estrés académico	Índice de intensidad de uso	0,665	[0,348; 0,845]	<0,001	<0,001
	Alfabetización en IA	0,826	[0,627; 0,924]	<0,001	<0,001
	Conductas de verificación	0,781	[0,544; 0,903]	<0,001	<0,001
Calidad percibida	Índice de intensidad de uso	0,686	[0,381; 0,856]	<0,001	<0,001
	Alfabetización en IA	0,732	[0,459; 0,879]	<0,001	<0,001
	Conductas de verificación	0,779	[0,540; 0,902]	<0,001	<0,001
Interacción/usabilidad	Índice de intensidad de uso	0,664	[0,347; 0,845]	<0,001	<0,001
	Alfabetización en IA	0,677	[0,367; 0,852]	<0,001	<0,001
	Conductas de verificación	0,785	[0,552; 0,905]	<0,001	<0,001
Aceptación/valor percibido	Índice de intensidad de uso	0,642	[0,313; 0,834]	<0,001	<0,001
	Alfabetización en IA	0,873	[0,721; 0,945]	<0,001	<0,001
	Conductas de verificación	0,836	[0,646; 0,928]	<0,001	<0,001

Nota. FDR = false discovery rate; q = valor p ajustado mediante el procedimiento de Benjamini-Hochberg. Todas las variables corresponden a medidas autorreportadas.

3.5 Rendimiento académico: análisis exploratorio

La nota final del curso en la cohorte analítica mostró una media de $4,42 \pm 0,68$ y una mediana de 4,40 [4,10; 4,75], con un rango entre 3,20 y 5,70. Dado el tamaño muestral reducido y el carácter

exploratorio del análisis de rendimiento académico, la nota final del curso se describe únicamente como referencia contextual de la cohorte. Estos resultados no permiten establecer inferencias sobre efectividad académica del tutor virtual ni sobre asociación robusta entre condiciones de uso y rendimiento académico.

4. Discusión

El hallazgo principal de este estudio fue que las condiciones autorreportadas de uso del tutor virtual se asociaron positivamente con la motivación académica y con distintas dimensiones subjetivas de experiencia. En particular, la alfabetización en IA autorreportada mediante ítem único, las conductas de verificación académica y la intensidad de uso mostraron asociaciones positivas con la motivación. Este patrón sugiere que, en esta cohorte, los estudiantes que reportaron una relación más activa, competente y cuidadosa con el tutor también tendieron a valorar de forma más favorable su experiencia de aprendizaje.

Estos resultados son coherentes con la literatura reciente que ha descrito beneficios potenciales del uso de ChatGPT y chatbots educativos sobre variables como satisfacción, motivación, aprendizaje percibido y apoyo al estudio autónomo (1-3). Sin embargo, el presente estudio aporta un matiz específico: las asociaciones observadas se concentran en experiencia subjetiva y condiciones autorreportadas de uso, por lo que deben interpretarse como evidencia exploratoria contextual y no como demostración de efectividad educativa. Esta distinción es relevante, porque la literatura disponible muestra resultados heterogéneos según contexto, diseño instruccional, tipo de tutor, forma de integración pedagógica y desenlaces evaluados.

La asociación entre alfabetización en IA autorreportada y experiencia subjetiva resulta particularmente relevante. Aunque en este estudio fue operacionalizada mediante un ítem único referido a la capacidad autopercebida para formular prompts claros y específicos con fines académicos, su relación con motivación, calidad percibida y otras dimensiones de experiencia sugiere que la manera en que los estudiantes comprenden y orientan la interacción con la herramienta puede influir en su valoración del tutor. Este resultado dialoga con trabajos que han señalado la necesidad de acompañar la incorporación de IA generativa con criterios pedagógicos explícitos, estrategias de uso intencional y marcos de implementación responsable (7, 10-11).

Las conductas de verificación académica también mostraron asociaciones positivas con la experiencia subjetiva. Este hallazgo debe leerse con prudencia, porque corresponde a una medida autorreportada y no a una observación directa de prácticas de verificación. Aun así, resulta consistente con la idea de que el valor formativo de los tutores basados en modelos de lenguaje depende de la calidad de la interacción que promueven, de la capacidad del estudiante para revisar las respuestas y de la delimitación del rol de la herramienta como apoyo al aprendizaje. En educación superior en salud, esta condición es especialmente sensible, porque el uso de IA requiere mantener juicio académico, supervisión docente y responsabilidad profesional frente a la información generada.

La intensidad de uso también se asoció positivamente con la motivación y con las dimensiones secundarias de experiencia subjetiva. Este resultado debe interpretarse con más cautela que una medida objetiva de interacción, ya que la intensidad fue autorreportada y puede reflejar distintos fenómenos: mayor interés por la asignatura, mayor necesidad de apoyo, mayor familiaridad con herramientas digitales o una percepción global más favorable hacia el tutor. En este sentido, el hallazgo coincide parcialmente con estudios que han reportado efectos favorables de tutores o chatbots educativos, pero también confirma la necesidad de considerar cómo, para qué y bajo qué

condiciones se produce la interacción, en lugar de interpretar la frecuencia de uso como indicador suficiente de aprendizaje (2-3).

Un punto importante es que los hallazgos del presente estudio se ubican principalmente en el plano de la experiencia subjetiva y no del rendimiento académico objetivo. En contraste con estudios experimentales que han observado mejoras en aprendizaje, motivación o compromiso en condiciones controladas, el presente trabajo se desarrolló en una implementación naturalista, voluntaria y transversal, con una muestra pequeña y sin estandarización de la exposición al tutor (3, 9). Por ello, la nota final del curso fue considerada únicamente como referencia contextual y no como evidencia de impacto académico. Esta diferencia no invalida el valor pedagógico potencial del tutor, pero acota el alcance de los resultados: el estudio informa sobre experiencia percibida en una cohorte específica, no sobre efectividad causal del recurso.

Desde una perspectiva curricular, los resultados sugieren que la incorporación de tutores basados en GPT debería acompañarse de orientaciones explícitas sobre uso académico responsable. La alfabetización en IA, la formulación de prompts, la verificación de respuestas y la comprensión de los límites del tutor aparecen como dimensiones relevantes para futuras implementaciones. Esta orientación es concordante con autores que han destacado la necesidad de integrar herramientas de IA generativa mediante marcos pedagógicos claros, evitando que su disponibilidad técnica sea confundida con integración educativa efectiva (1, 5, 11).

Este estudio presenta limitaciones importantes. En primer lugar, el diseño transversal impide establecer temporalidad, direccionalidad o causalidad entre las variables. En segundo lugar, la muestra fue pequeña y restringida a una única cohorte de estudiantes de Kinesiología, lo que limita la generalización de los hallazgos. En tercer lugar, las variables centrales fueron autorreportadas, por lo que pueden estar influidas por sesgo de recuerdo, deseabilidad social o una percepción global favorable hacia el tutor. En cuarto lugar, el instrumento fue construido ad hoc para esta implementación; aunque mostró alta consistencia interna y fue revisado por expertos en términos de claridad y pertinencia, no cuenta con validación psicométrica formal. En quinto lugar, la alfabetización en IA fue operacionalizada mediante un ítem único, lo que limita la profundidad conceptual con que puede interpretarse este constructo. Además, no se dispuso de registros objetivos de interacción con el tutor ni de covariables basales como rendimiento previo, experiencia previa con IA o autoeficacia académica inicial. Finalmente, el estudio no contó con preregistro, por lo que sus resultados deben entenderse como exploratorios y generadores de hipótesis.

Aun con estas limitaciones, el estudio aporta evidencia aplicada desde un contexto real de educación superior en salud. Fue realizado en una asignatura regular, con un tutor configurado a partir de materiales oficiales y con estudiantes expuestos a una implementación naturalista. Su aporte principal no radica en demostrar efectividad causal del tutor, sino en mostrar que la experiencia subjetiva con herramientas basadas en GPT puede relacionarse con la manera en que los estudiantes declaran usarlas, comprenderlas y verificarlas. Esta información puede orientar futuras investigaciones con muestras mayores, diseños longitudinales o controlados, instrumentos validados y registros objetivos de interacción.

5. Conclusiones

- En esta cohorte pequeña de estudiantes de Kinesiología, las condiciones autorreportadas de uso de un tutor virtual basado en GPT se asociaron positivamente con una experiencia subjetiva más favorable. La alfabetización en IA autorreportada mediante ítem único, las conductas de verificación académica y la intensidad de uso mostraron asociaciones positivas con la motivación académica y con distintas dimensiones de experiencia percibida.

- Estos hallazgos deben interpretarse como asociaciones exploratorias, contextuales y basadas en autorreporte, sin inferir causalidad ni efectividad sobre el rendimiento académico. En conjunto, los resultados sugieren que la forma en que los estudiantes declaran usar, comprender y verificar la información generada por un tutor basado en IA puede ser relevante para su valoración subjetiva del recurso. Futuras investigaciones deberán considerar muestras mayores, diseños longitudinales o controlados, instrumentos validados, registros objetivos de interacción y medidas basales que permitan evaluar con mayor precisión el rol de la alfabetización en IA y de las prácticas de uso académico responsable en educación superior en salud.

Financiación: No hubo financiación.

Declaración de conflicto de interés: Los autores declaran no tener ningún conflicto de interés.

Uso de inteligencia artificial en el proceso de investigación: Se utilizó ChatGPT para la configuración de un tutor virtual personalizado, restringido al corpus definido por los autores. Los estudiantes accedieron al tutor mediante cuentas personales gratuitas. En la fase analítica, los modelos de lenguaje se utilizaron únicamente como apoyo para la generación de código en Python para cálculos estadísticos, sin automatizar decisiones analíticas, selección de resultados ni interpretación final. Además, ChatGPT se utilizó como apoyo en la revisión de redacción del manuscrito. Los autores asumen plena responsabilidad por el contenido final.

Contribuciones de los autores: ICM*: Concepción del estudio, diseño metodológico, implementación del tutor, recolección de datos, análisis, redacción del borrador original y revisión final. MMM: Apoyo en diseño metodológico, revisión crítica del manuscrito, interpretación de resultados y aprobación de la versión final.

6. Referencias

1. Michel-Villarreal R, Vilalta-Perdomo E, Salinas-Navarro DE, Thierry-Aguilera R, Gerardou FS. Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. *Education Sciences*, **2023**, 13, 856. <https://doi.org/10.3390/educsci13090856>.
2. Sáiz-Manzanares MC, Marticorena-Sánchez R, Martín-Antón LJ, González Díez I, Almeida LS. Perceived satisfaction of university students with the use of chatbots as a tool for self-regulated learning. *Heliyon*, **2023**, 9, e12843. <https://doi.org/10.1016/j.heliyon.2023.e12843>.
3. Deng R, Jiang M, Yu X, Lu Y, Liu S. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, **2025**, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>.
4. Zhai X. ChatGPT User Experience: Implications for Education. *SSRN Electronic Journal*, **2022**. <https://doi.org/10.2139/ssrn.4312418>.
5. Baidoo-Anu D, Owusu Ansah L. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, **2023**, 7, 52-62. <https://doi.org/10.61969/jai.1337500>.
6. Laun M, Wolff F. Chatbots in education: Hype or help? A meta-analysis. *Learning and Individual Differences*, **2025**, 119, 102646. <https://doi.org/10.1016/j.lindif.2025.102646>.
7. Hardt D, Nagler M, Rincke J. Tutoring in (online) higher education: Experimental evidence. *Economics of Education Review*, **2023**, 92, 102350. <https://doi.org/10.1016/j.econedurev.2022.102350>.
8. Vega Jiménez J, Borja Gómez EE, Ramírez Álvarez PJ. ChatGPT e inteligencia artificial: ¿obstáculo o ventaja para la educación médica superior? *Educación Médica Superior*, **2023**, 37(2). <https://ems.sld.cu/index.php/ems/article/view/3851>.
9. Kestin G, Miller K, Klales A, Milbourne T, Ponti G. AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*. **2025**, 15, 17458. <https://doi.org/10.1038/s41598-025-97652-6>.

10. Sánchez-Vera F. Subject-Specialized Chatbot in Higher Education as a Tutor for Autonomous Exam Preparation: Analysis of the Impact on Academic Performance and Students' Perception of Its Usefulness. *Education Sciences*, **2025**, 15, 26. <https://doi.org/10.3390/educsci15010026>.
11. Fageeh AI. The rise of chatbots in higher education: Exploring user profiles, motivations, and integration strategies. *Social Sciences & Humanities Open*, **2025**, 12, 101996. <https://doi.org/10.1016/j.ssaho.2025.101996>.

Copyright



© 2026 University of Murcia. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Spain License (CC BY-NC-ND) (<http://creativecommons.org/licenses/by/4.0/>)