Revista Española
de Educación Médica

UNIVERSIDAD DE MURCIA

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

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

Italo Campos Montenegro [1]*, Maximiliano Martínez Mena [2]

[1] Kinesiology Program, Faculty of Life Sciences, Viña del Mar University, Viña del Mar, Chile, italo.campos@uvm.cl, https://orcid.org/0009-0007-0325-3344

[2] Kinesiology Program, Faculty of Life Sciences, Viña del Mar University, Viña del Mar, Chile; maximiliano.martinez@uvm.cl; https://orcid.org/0009-0006-6181-0324

* Correspondence: italo.campos@uvm.cl

Received: 7/5/26; Accepted: 4/6/26; Published: 6/6/26

Summary.
Aim: To analyze the association between self-reported conditions of use of a GPT-based virtual tutor and the subjective learning experience in Kinesiology students, considering academic motivation as the primary outcome. **Methods:** This was an analytical, cross-sectional observational study of a 2025 cohort of Kinesiology students exposed to a GPT (Global Pedagogical Tutor) implemented as a supplementary learning support. A structured survey was administered to measure seven subjective dimensions of experience: motivation, self-efficacy, facilitation of learning, reduction of academic anxiety/stress, perceived quality, interaction/usability, and acceptance/perceived value. Three self-reported usage conditions were assessed: intensity of tutor use, self-reported AI literacy (using a single item), and academic checking behaviors. Associations were analyzed using bivariate Pearson correlations. The primary outcome was academic motivation; secondary outcomes were adjusted for multiplicity using the Benjamini-Hochberg FDR. **Results:** Twenty-three students were included. All three self-reported usage conditions were positively associated with academic motivation: intensity of use ($r = 0.703$; 95% CI 0.409 to 0.865; $p < 0.001$), single-item self-reported AI literacy ($r = 0.767$; 95% CI 0.519 to 0.896; $p < 0.001$), and academic checking behaviors ($r = 0.773$; 95% CI 0.530 to 0.899; $p < 0.001$). In secondary outcomes, the associations were positive and remained statistically significant after FDR adjustment. The final course grade was described only as a contextual reference, without inferences of academic effectiveness. **Conclusions:** In this small cohort, self-reported use of a GPT-based virtual tutor was 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 records of interaction.

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

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.

1. Introduction

Generative artificial intelligence (AAI), and in particular large language models (LLMs), has been rapidly incorporated into higher education as a tool capable of providing on-demand, customizable, and continuously available academic support. In this scenario, platforms such as ChatGPT have begun to operate de facto as virtual tutors to answer questions, organize content, provide feedback on assignments, and support independent study. Recent literature suggests that these tools can offer benefits in terms of perceived learning, motivation, and some performance indicators, although with significant heterogeneity across contexts, instructional designs, and usage patterns (1-3).

In parallel, the incorporation of LLMs in higher education has been discussed in diverse international contexts, including experiences and studies developed in North America, Africa, Europe, and Latin America. This expansion has intensified concerns about misuse, substitution of cognitive effort, academic integrity, and dependence on automatically generated responses. These tensions are especially relevant when generative tools are integrated into evaluative tasks or activities where the learning objective requires elaboration, reasoning, and independent judgment, rather than simply access to information. Therefore, the current debate no longer focuses solely on whether these tools are available or accepted by students, but rather on how they are used and under what conditions that use is educationally valuable (4-6). From this perspective, variables such as the ability to formulate useful prompts, verify information, and guide interaction toward plausible academic goals may be more relevant than the raw frequency of use. This distinction is particularly important in higher health education, where learning requires not only access to information but also critical selection, applied reasoning, and responsible use of knowledge. Along these lines, some recent studies and reviews have suggested that the effectiveness of conversational tutors depends largely on the pedagogical design and the type of interaction they promote, rather than on their mere availability or a nonspecific increase in their use (3, 7-8).

Despite the growing body of literature on generative AI in education, evidence applied in real-world health training contexts that simultaneously examines students' subjective experiences and the tutor's specific usage conditions remains limited. Many available studies describe acceptance, perceived usefulness, or overall effects of chatbot use, but pay less attention to whether the perceived value of the resource depends more on the quantity of use or on the quality or competence with which that use is carried out. This gap is relevant because a pedagogically responsible implementation requires not only more interaction with the tool, but also a more precise understanding of which usage features are associated with more favorable learning experiences (1-2, 6).

In this context, the present study analyzes the implementation of a personalized virtual tutor based on GPT (Global Positioning Technique) for Kinesiology students at the University of Viña del Mar during 2025, in the Semiology in Traumatology and Imaging course. The objective was to evaluate the association between three self-reported usage conditions—intensity of use, AI literacy, and academic verification behaviors—and different dimensions of the subjective experience associated with the tutor, considering academic motivation as the primary outcome. It was hypothesized that higher levels of AI literacy and verification behaviors would be associated with a more favorable experience, while intensity of use would play a less consistent role. Thus, the study seeks to provide applied evidence to guide the design and implementation of tutors based on language models in university health training contexts.

2. Methods

2.1 Study design

A cross-sectional, analytical observational study was conducted using a survey administered to undergraduate students who were exposed to a personalized virtual tutor based on GPT during 2025. The analytical objective was to examine the association between self-reported tutor usage and various subjective dimensions of the learning experience. The target population consisted of the 2025 cohort of students enrolled in the Semiology in Traumatology and Imaging course within the Kinesiology program at Universidad Viña del Mar (Chile). The tutor was made available to all enrolled students during the academic semester as a supplementary and voluntary resource. Eligibility criteria included: (i) enrollment in the course during 2025; (ii) providing written informed consent; and (iii) completing the survey. During the data collection period, a total of 31 eligible students were identified. Twenty-three students responded to the survey, representing a response rate of 74%. The unit of analysis was the student. For the correlational analyses, the 23 available records were considered, without imputing missing data. The final course grade was analyzed solely as a contextual reference and for exploratory purposes.

2.2 Virtual tutor and implementation

A personalized virtual tutor was implemented in ChatGPT via a Plus account (model 4), configured with relevant course materials, including teaching planning documents. The resource was presented as support for independent work and did not replace regular instruction. Prior to its implementation, a 30-minute general training session was conducted on its capabilities, explicitly stating that it should be used as a learning support tool, not as a substitute for assessments or academic work.

2.3 Instrument and procedure

A structured online survey using 5-point Likert scales was administered to measure seven subjective dimensions of the tutor experience: motivation, self-efficacy, facilitation of learning, reduction of academic anxiety/stress, perceived tutor quality, interaction/usability, and acceptance/perceived value. The acceptance/perceived value dimension included items related to the tutor's perceived usefulness for learning, intention to continue using the tutor, perceived improvement in the learning experience, and perceived benefits beyond the course. The instrument was developed specifically for this study, based on recurring conceptual dimensions found in the literature on subjective experience of using conversational tutors and learning support. Before its administration, the items were reviewed by experts (n = 3) to assess clarity, relevance, and coherence with the proposed constructs, resulting in adjustments to wording and content. Given its specific nature for this implementation, the instrument was not subjected to prior formal psychometric validation, so its results should be interpreted as self-reported measures of use and experience in this cohort.

2.4 Variables and operationalization

Three main self-reported predictors were defined:

- **Intensity of tutor use:** a composite index derived from three ordinal indicators: frequency of use during the semester, use at key times, and typical session duration. Each indicator was transformed into a standardized score (z-score), and the simple average of the three standardized scores was then calculated to obtain a synthetic index of intensity of use. Higher values indicated greater relative intensity of tutor use.
- **AI Literacy:** Single Likert item from 1 to 5 that assessed self-perception of ability to formulate clear and specific prompts for academic purposes.
- **Academic verification behaviors:** average of three items aimed at verifying information, detecting errors or hallucinations, and explaining sources or processes when appropriate.

Academic motivation was defined a priori as the primary outcome. Secondary outcomes included self-efficacy, facilitation of learning, reduction of academic anxiety/stress, perceived quality, interaction/usability, and acceptance/perceived value. Each dimension was calculated as the average of the corresponding items, with higher scores indicating a more favorable experience.

2.5 Statistical analysis

Descriptive analyses of tutor usage variables and scores per dimension were performed using means and standard deviations. The internal consistency of each dimension was estimated using Cronbach's alpha, interpreted as an indicator of internal coherence of the items in this cohort and not as evidence of formal psychometric validation of the instrument. Considering the small sample size, the exploratory nature of the study, and the self-reported nature of the variables, the associations between tutor usage conditions and subjective experience dimensions were analyzed using Pearson bivariate correlations. Since the subjective dimensions were calculated as averages of Likert-type items, they were treated as approximate continuous variables for the correlational analysis. This decision aimed to maintain the study's associative objective through a more parsimonious analytical strategy proportional to the cohort size, avoiding potentially unstable multivariate estimates.

Academic motivation remained the primary outcome. Secondary outcomes included self-efficacy, facilitation of learning, reduction of academic anxiety/stress, perceived quality, interaction/usability, and acceptance/perceived value. Multiplicity was controlled for secondary outcomes using the Benjamini-Hochberg FDR procedure, applied to the family of contrasts corresponding to the associations between the three tutor usage conditions and the six secondary dimensions of subjective experience.

The final course grade was examined solely as an exploratory analysis, without drawing inferences about the academic effectiveness of the online tutor. Analyses were performed using Python. The use of language models was limited to supporting code generation, without automating analytical decisions, results selection, or final interpretation. Since the study included the entire available cohort for a given course, a convenience sample was used, and no formal sample size calculation was performed. Due to the voluntary nature of both participation and the use of the tutor, self-selection, self-reporting, and social desirability bias were considered potential sources of bias. To partially mitigate these risks, students were informed that their decision to participate or not would have no academic consequences, and the analytical data was pseudo-anonymized before analysis.

2.6 Ethical considerations

The study was conducted in accordance with the ethical principles applicable to research involving human subjects and in compliance with the Declaration of Helsinki, as required by this journal's editorial policy for original works with human participation. The research was approved by the Scientific Ethics Committee of the University of Viña del Mar (CEC-UVM), Minutes No. 55-25, August 12, 2025. Participation was voluntary and based on prior informed consent. Students were informed that their decision to participate or not would have no academic consequences. For eligibility verification and linking to academic records, nominal identification was requested at the initial data collection stage; subsequently, the database was pseudonymized, removing direct identifiers and restricting access to nominative information to the research team exclusively for verification and linking purposes. The survey was administered using an Outlook form, and the results were accessible only to the principal investigator. The data were used exclusively for research purposes, treated with confidentiality safeguards, and reported in aggregate form.

3. Results

3.1 Participants and analytical basis

During the study period, 31 students were eligible. Of these, 23 responded to the survey, representing a response rate of 74%. All 23 available records were used for the correlational analyses. Missing data were not imputed. The analysis of the final course grade was conducted using the 23 available records and was interpreted solely for exploratory purposes.

3.2 Use of the virtual tutor and general experience profile

As shown in Table 1, the indicators of tutor use were in the intermediate to high ranges: the overall frequency of use was 3.22 ± 1.04, use at key moments 3.35 ± 1.15, and the typical session duration 3.04 ± 1.43, on ordinal scales from 1 to 5. In addition to the indicators of intensity of use, self-reported AI literacy had a mean of 3.91 ± 0.79, while academic verification behaviors had a mean of 3.88 ± 0.78. Both indicators were expressed on scales from 1 to 5, where higher values indicated greater self-perception of competence or a higher frequency of verification behaviors.

The subjective dimensions showed average scores close to the high end of the scale. Means ranged from 3.76 to 4.01 points (Table 2). The highest mean was observed in facilitation of learning (4.01 ± 0.95), followed by interaction/usability (4.00 ± 0.97), self-efficacy (3.98 ± 0.99), motivation (3.90 ± 1.01), perceived quality of the tutor (3.90 ± 0.94), reduction of academic anxiety/stress (3.78 ± 1.03), and acceptance/perceived value (3.76 ± 0.97). In this sample, the dimensions showed high internal consistency estimated by Cronbach's alpha (0.92 to 0.97), which supports the internal coherence of the items in this cohort, although this does not constitute complete psychometric validation of the instrument.

Table 1. Usage descriptors and main predictors.

Indicator	n	Average	OF
Frequency of use of the virtual tutor during the semester	23	3.22	1.04
Frequency of use at key times (outside of business hours/before delivery)	23	3.35	1.15
Typical duration of usage sessions	23	3.04	1.43
Self-reported AI literacy	23	3.91	0.79
Academic verification behaviors	23	3.88	0.78

SD = standard deviation. The indicators are expressed on ordinal or Likert-type scales from 1 to 5; higher scores indicate greater frequency, duration, self-perception of competence, or frequency of checking behaviors.

Table 2. Scores (mean ± SD) and internal consistency (Cronbach's α) ($n = 23$).

Dimension	k	Average	OF	α
Motivation	4	3.90	1.01	0.97
Self-efficacy	4	3.98	0.99	0.96
Facilitating learning	3	4.01	0.95	0.94
Reduction of academic anxiety/stress	4	3.78	1.03	0.95
Perceived quality of the tutor	4	3.90	0.94	0.92
Interaction/usability	4	4.00	0.97	0.97
Acceptance/perceived value	4	3.76	0.97	0.93

SD = standard deviation; α = Cronbach's alpha.

3.3 Primary outcome: academic motivation

The three self-reported tutor use conditions were positively associated with academic motivation (Table 3). The highest correlation was observed for academic checking behaviors ($r = 0.773$; 95% CI 0.530 to 0.899; $p < 0.001$), followed by self-reported single-item AI literacy ($r = 0.767$; 95% CI 0.519 to 0.896; $p < 0.001$) and usage intensity index ($r = 0.703$; 95% CI 0.409 to 0.865; $p < 0.001$).

3.4 Secondary Results

Exploratory associations between self-reported tutor use conditions and secondary dimensions of subjective experience were positive (Table 4). After controlling for multiplicity using the Benjamini-Hochberg FDR, all evaluated secondary associations remained statistically significant ($q < 0.001$). Observed correlations ranged from $r = 0.642$ to $r = 0.873$. The strongest associations were observed between self-reported single-item AI literacy and acceptance/perceived value ($r = 0.873$; 95% CI 0.721 to 0.945), self-reported single-item AI literacy and reduced academic anxiety/stress ($r = 0.826$; 95% CI 0.627 to 0.924), and academic checking behaviors with acceptance/perceived value ($r = 0.836$; 95% CI 0.646 to 0.928). These results should be interpreted as exploratory associations between self-reported variables, without establishing causal directionality or independence between conditions of use.

Table 3. Bivariate correlations between tutor usage conditions and academic motivation.

Primary result	Associated variable	Pearson's r	95% CI	p
Motivation	Usage intensity index	0.703	[0.409; 0.865]	<0.001
Motivation	AI Literacy	0.767	[0.519; 0.896]	<0.001
Motivation	Academic verification behaviors	0.773	[0.530; 0.899]	<0.001

Table 4. Correlations between the tutor's usage conditions and the secondary dimensions of subjective experience.

Secondary outcome	Associated variable	Pearson's r	95% CI	nominal p	q FDR
Self-efficacy	Usage intensity index	0.716	[0.431; 0.871]	<0.001	<0.001
	AI Literacy	0.737	[0.467; 0.882]	<0.001	<0.001
	Verification behaviors	0.726	[0.449; 0.876]	<0.001	<0.001
Facilitating learning	Usage intensity index	0.724	[0.445; 0.875]	<0.001	<0.001
	AI Literacy	0.746	[0.481; 0.886]	<0.001	<0.001
	Verification behaviors	0.763	[0.511; 0.894]	<0.001	<0.001
Reduction of academic anxiety/stress	Usage intensity index	0.665	[0.348; 0.845]	<0.001	<0.001
	AI Literacy	0.826	[0.627; 0.924]	<0.001	<0.001
	Verification behaviors	0.781	[0.544; 0.903]	<0.001	<0.001
Perceived quality	Usage intensity index	0.686	[0.381; 0.856]	<0.001	<0.001
	AI Literacy	0.732	[0.459; 0.879]	<0.001	<0.001
	Verification behaviors	0.779	[0.540; 0.902]	<0.001	<0.001
Interaction/ usability	Usage intensity index	0.664	[0.347; 0.845]	<0.001	<0.001
	AI Literacy	0.677	[0.367; 0.852]	<0.001	<0.001
	Verification behaviors	0.785	[0.552; 0.905]	<0.001	<0.001
Acceptance/ perceived value	Usage intensity index	0.642	[0.313; 0.834]	<0.001	<0.001
	AI Literacy	0.873	[0.721; 0.945]	<0.001	<0.001
	Verification behaviors	0.836	[0.646; 0.928]	<0.001	<0.001

Note. FDR = false discovery rate; q = p-value adjusted using the Benjamini-Hochberg procedure. All variables correspond to self-reported measurements.

3.5 Academic performance: exploratory analysis

The final course grade in the analytical cohort showed a mean of 4.42 ± 0.68 and a median of 4.40 [4.10; 4.75], with a range between 3.20 and 5.70. Given the small sample size and the exploratory nature of the academic performance analysis, the final course grade is described only as a contextual reference for the cohort. These results do not allow for inferences to be drawn regarding the academic effectiveness of the online tutor or a robust association between usage conditions and academic performance.

4. Discussion

The main finding of this study was that self-reported conditions of virtual tutor use were positively associated with academic motivation and various subjective dimensions of experience. In particular, single-item self-reported AI literacy, academic verification behaviors, and intensity of use showed positive associations with motivation. This pattern suggests that, in this cohort, students who reported a more active, competent, and attentive relationship with the tutor also tended to rate their learning experience more favorably.

These results are consistent with recent literature that has described potential benefits of using ChatGPT and educational chatbots on variables such as satisfaction, motivation, perceived learning, and support for independent study (1-3). However, the present study offers a specific nuance: the observed associations are concentrated in subjective experience and self-reported conditions of use, and therefore should be interpreted as contextual exploratory evidence and not as a demonstration of educational effectiveness. This distinction is relevant because the available literature shows heterogeneous results depending on the context, instructional design, type of tutor, form of pedagogical integration, and outcomes evaluated.

The association between self-reported AI literacy and subjective experience is particularly relevant. Although in this study it was operationalized through a single item referring to the self-perceived ability to formulate clear and specific prompts for academic purposes, its relationship with motivation, perceived quality, and other dimensions of experience suggests that the way in which students understand and guide their interaction with the tool can influence their evaluation of the tutor. This result aligns with studies that have highlighted the need to accompany the incorporation of generative AI with explicit pedagogical criteria, strategies for intentional use, and frameworks for responsible implementation (7, 10-11).

Academic verification behaviors also showed positive associations with subjective experience. This finding should be interpreted with caution, as it is based on self-reported measures rather than direct observation of verification practices. Even so, it is consistent with the idea that the formative value of language-based tutors depends on the quality of the interaction they foster, the student's ability to review their responses, and the clear definition of the tool's role as a learning support tool. In higher health education, this is particularly critical because the use of AI requires maintaining academic judgment, faculty supervision, and professional responsibility regarding the information generated.

Intensity of use was also positively associated with motivation and secondary dimensions of subjective experience. This result should be interpreted with more caution than an objective measure of interaction, since intensity was self-reported and may reflect different phenomena: greater interest in the subject, a greater need for support, greater familiarity with digital tools, or a more favorable overall perception of the tutor. In this sense, the finding partially coincides with studies that have reported favorable effects of tutors or educational chatbots, but it also confirms the need to consider how, why, and under what conditions the interaction occurs, rather than interpreting frequency of use as a sufficient indicator of learning (2-3).

An important point is that the findings of this study are primarily situated in the realm of subjective experience rather than objective academic performance. In contrast to experimental studies that have observed improvements in learning, motivation, or engagement under controlled conditions, this work was conducted in a naturalistic, voluntary, and cross-sectional setting, with a small sample and without standardization of tutor exposure (3, 9). Therefore, the final course grade was considered only as a contextual reference and not as evidence of academic impact. This difference does not invalidate the potential pedagogical value of the tutor, but it does limit the scope of the results: the study reports on the perceived experience in a specific cohort, not on the causal effectiveness of the resource.

From a curricular perspective, the results suggest that the incorporation of GPT-based tutors should be accompanied by explicit guidelines on responsible academic use. AI literacy, prompt formulation, response verification, and understanding the tutor's limitations emerge as relevant dimensions for future implementations. This approach aligns with authors who have emphasized the need to integrate generative AI tools through clear pedagogical frameworks, preventing their technical availability from being mistaken for effective educational integration (1, 5, 11).

This study has significant limitations. First, the cross-sectional design prevents establishing temporality, directionality, or causality among the variables. Second, the sample was small and restricted to a single cohort of Kinesiology students, which limits the generalizability of the findings. Third, the central variables were self-reported, and therefore may be influenced by recall bias, social desirability bias, or an overall favorable perception of the tutor. Fourth, the instrument was constructed specifically for this implementation; although it showed high internal consistency and was reviewed by experts for clarity and relevance, it lacks formal psychometric validation. Fifth, AI literacy was operationalized through a single item, which limits the conceptual depth with which this construct can be interpreted. Furthermore, objective records of interaction with the tutor and baseline covariates such as prior performance, prior experience with AI, or initial academic self-efficacy were not available. Finally, the study did not include pre-registration, so its results should be understood as exploratory and hypothesis-generating.

Even with these limitations, the study provides applied evidence from a real-world higher education health context. It was conducted in a regular course, with a tutor trained using official materials, and with students exposed to a naturalistic approach. Its main contribution lies not in demonstrating the tutor's causal effectiveness, but in showing that subjective experience with GPT-based tools can be related to how students report using, understanding, and verifying them. This information can guide future research with larger samples, longitudinal or controlled designs, validated instruments, and objective records of interaction.

5. Conclusions

- In this small cohort of kinesiology students, self-reported conditions of use of a GPT-based virtual tutor were positively associated with a more favorable subjective experience. Single-item self-reported AI literacy, academic verification behaviors, and intensity of use showed positive associations with academic motivation and various dimensions of perceived experience.
- These findings should be interpreted as exploratory, contextual, and self-reported associations, without inferring causality or effectiveness on academic performance. Taken together, the results suggest that how students report using, understanding, and verifying information generated by an AI-based tutor may be relevant to their subjective evaluation of the resource. Future research should consider larger samples, longitudinal or controlled designs, validated instruments, objective interaction records, and baseline measures to more accurately assess the role of AI literacy and responsible academic use practices in higher health education.

Funding: There was no funding.
Declaration of conflict of interest: The authors declare that they have no conflict of interest.
Use of artificial intelligence in the research process: ChatGPT was used to configure a personalized virtual tutor, restricted to the corpus defined by the authors. Students accessed the tutor using free personal accounts. In the analytical phase, language models were used solely to support the generation of Python code for statistical calculations, without automating analytical decisions, result selection, or final interpretation. ChatGPT was also used to support the manuscript's editing process. The authors assume full responsibility for the final content.

Authors' contributions: ICM*: Study conception, methodological design, tutor implementation, data collection, analysis, drafting of the original manuscript, and final review. MMM: Support in methodological design, critical review of the manuscript, interpretation of results, and approval of the final version.

6. References

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 and artificial intelligence: obstacle or advantage for higher medical education? *Higher Medical Education,* **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 .