Revista Española
de Educación Médica

Preferencias declaradas por estilos de tutoría virtual en dos cohortes curriculares de Kinesiología: estudio transversal exploratorio.

Declared preferences for virtual tutoring styles in two curricular cohorts of Kinesiology students: an exploratory cross-sectional study.

Italo Campos Montenegro [1]

[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

* Correspondencia: italo.campos@uvm.cl

Recibido: 24/6/26; Aceptado: 28/7/26; Publicado: 30/7/26

Resumen.

Objetivo: Comparar la preferencia declarada por estilos de tutoría virtual entre dos cohortes curriculares de estudiantes de Kinesiología y describir diferencias en utilidad percibida, estrategias de estudio y orientación funcional de uso del tutor. **Métodos**: Se realizó un estudio cuantitativo, observacional y transversal en estudiantes de 6.º y 8.º semestre de Kinesiología de la Universidad Viña del Mar. Los estudiantes tuvieron acceso libre a dos tutores virtuales —uno estructurado-explicativo y otro reflexivo-socrático— sin asignación, exposición estandarizada ni registro objetivo de las interacciones. La variable principal fue la preferencia declarada por el tutor reflexivo-socrático. Los resultados secundarios se obtuvieron mediante un cuestionario elaborado ad hoc y se interpretaron como indicadores contextuales de autorreporte. La diferencia entre cohortes curriculares en la preferencia por el tutor reflexivo-socrático se estimó mediante prueba exacta de Fisher y regresión logística penalizada de Firth, ajustada por frecuencia de uso y experiencia previa con tutores virtuales. **Resultados**: Participaron 44 de 70 estudiantes invitados. La preferencia por el tutor reflexivo-socrático fue reportada por el 10% de estudiantes de 6.º semestre y por un 70,8% de los de 8.º semestre. La asociación fue significativa en el análisis no ajustado y se mantuvo tras el ajuste mediante regresión logística penalizada de Firth. La utilidad percibida fue alta en ambos grupos. En los resultados secundarios se observaron patrones descriptivos diferentes en reflexión, justificación y elaboración de resúmenes o esquemas, pero ninguna de las ocho comparaciones conservó significación estadística después del ajuste de Holm. **Conclusiones**: Se observaron diferencias marcadas en la preferencia declarada por estilos de tutoría virtual entre las dos cohortes curriculares estudiadas. Debido a que cada cohorte correspondía a una asignatura y a tutores contextualizados diferentes, los resultados no permiten atribuir estas diferencias exclusivamente al nivel formativo. Los patrones observados permiten formular hipótesis sobre la relación entre configuración tutorial, contexto curricular, características de la tarea y grado de autonomía esperado, que deberán evaluarse mediante diseños con exposición equivalente y resultados educativos objetivos.

Palabras clave: tutor virtual; inteligencia artificial; ChatGPT; tutoría socrática; razonamiento clínico; educación en salud; educación superior.

Abstract

Objective: To compare the reported preference for virtual tutoring styles between two curricular cohorts of Kinesiology students and to describe differences in perceived usefulness, study strategies, and functional orientation of tutor use. **Methods:** A quantitative, observational, cross-sectional study was conducted with 6th and 8th semester Kinesiology students at Universidad Viña del Mar.

Students had open access to two virtual tutors—one structured-explanatory and the other reflective-Socratic—without assignment, standardized exposure, or objective recording of interactions. The primary variable was the reported preference for the reflective-Socratic tutor. Secondary outcomes were obtained through an ad hoc questionnaire and interpreted as contextual indicators of self-reporting. The difference between curricular cohorts in preference for the reflective-Socratic tutor was estimated using Fisher's exact test and Firth's penalized logistic regression, adjusted for frequency of use and prior experience with virtual tutors. **Results:** 44 of 70 invited students participated. Preference for the reflective-Socratic tutor was reported by 10% of sixth-semester students and by 70.8% of eighth-semester students. The association was significant in the unadjusted analysis and remained so after adjustment using Firth's penalized logistic regression. Perceived usefulness was high in both groups. Secondary outcomes showed different descriptive patterns in reflection, justification, and the creation of summaries or outlines, but none of the eight comparisons retained statistical significance after Holm's adjustment. **Conclusions**: Marked differences were observed in the reported preference for online tutoring styles between the two curricular cohorts studied. Because each cohort corresponded to a different subject and different contextualized tutors, the results do not allow these differences to be attributed solely to the educational level. The observed patterns allow us to formulate hypotheses about the relationship between tutoring configuration, curricular context, task characteristics, and expected degree of autonomy, which should be evaluated using designs with equivalent exposure and objective educational outcomes.

Keywords: virtual tutor; artificial intelligence; ChatGPT; Socratic tutoring; clinical reasoning; health professions education; higher education.

1. Introducción

La integración de tutores virtuales conversacionales en educación superior ha crecido de manera sostenida, consolidándose como una estrategia emergente para apoyar el aprendizaje en contextos diversos. Estos sistemas —desde chatbots educativos hasta agentes basados en modelos de lenguaje de gran escala, como ChatGPT— se han incorporado en funciones explicativas, asistenciales, evaluativas y reflexivas. La discusión actual ha comenzado a desplazarse desde la disponibilidad tecnológica hacia la calidad de la mediación pedagógica: cómo interactúan estos tutores con los estudiantes, qué tipo de participación cognitiva promueven y en qué condiciones su diseño resulta más adecuado para distintos perfiles formativos (1, 2). Una dimensión crítica en esta discusión es el estilo pedagógico con que el tutor virtual estructura la interacción. En términos generales, pueden distinguirse orientaciones estructurado-explicativas, centradas en organizar contenidos, entregar explicaciones guiadas, secuenciar información y apoyar la síntesis del material, y orientaciones reflexivo-socráticas, centradas en preguntas guía, contraste de supuestos, justificación de respuestas y estimulación del razonamiento autónomo. Esta distinción no representa una oposición absoluta entre entregar respuestas y formular preguntas, sino una diferencia en el énfasis pedagógico predominante. Mientras algunos diseños reducen la fricción inicial del aprendizaje mediante estructura, claridad y guía, otros buscan aumentar la participación activa del estudiante mediante reflexión, argumentación y metacognición (3, 4).

La evidencia reciente muestra que los tutores conversacionales pueden incidir en múltiples dimensiones del aprendizaje, incluyendo rendimiento, motivación, carga cognitiva, compromiso, aceptación tecnológica y razonamiento crítico. Sin embargo, su efectividad depende de la forma en que se integran pedagógicamente y del modo en que sus interacciones apoyan procesos de autorregulación, monitoreo metacognitivo y retroalimentación. Desde esta perspectiva, el valor del tutor virtual no reside únicamente en responder con rapidez o disponibilidad permanente, sino en organizar, desafiar y acompañar el proceso de aprendizaje de manera coherente con los objetivos formativos (5, 6). Esta precisión es especialmente relevante porque el uso frecuente de un tutor virtual no equivale necesariamente a un uso pedagógicamente funcional. Cuando el sistema entrega productos finales sin promover elaboración, verificación o justificación, puede facilitar el desempeño

inmediato y, al mismo tiempo, limitar la comprensión conceptual profunda. Estudios recientes han mostrado que ciertos apoyos de IA pueden mejorar el rendimiento durante la tarea o reducir la frustración sin producir ganancias equivalentes de aprendizaje, y que el uso de IA generativa sin resguardos pedagógicos puede afectar negativamente el aprendizaje cuando desplaza el esfuerzo cognitivo del estudiante. En esta línea, la sobredependencia de sistemas conversacionales de IA ha sido identificada como un riesgo para habilidades cognitivas como pensamiento crítico, juicio autónomo y toma de decisiones (7-9).

En este marco, la percepción estudiantil aporta señales relevantes para el diseño educativo, aunque requiere una interpretación cuidadosa. La satisfacción, la utilidad percibida o la preferencia declarada por un estilo de tutoría no demuestran por sí mismas efectividad educativa ni aprendizaje profundo. Un tutor puede ser valorado por su claridad, rapidez o disponibilidad, y aun así promover interacciones de baja exigencia cognitiva. Del mismo modo, un tutor reflexivo puede estimular razonamiento y autonomía, pero aumentar la carga cognitiva si no se ajusta al nivel formativo, al tipo de tarea o al grado de dominio previo del estudiante. Por ello, estudiar las preferencias estudiantiles resulta relevante cuando se interpretan como indicadores de adecuación pedagógica percibida y no como evidencia directa de superioridad de un estilo sobre otro (2, 4). Esta delimitación también ha sido aplicada en estudiantes de Kinesiología expuestos a un tutor basado en GPT, donde las asociaciones entre condiciones autorreportadas de uso y experiencia subjetiva se interpretaron como exploratorias y contextuales, sin inferir causalidad ni efectividad sobre el rendimiento académico (10).

Esta discusión adquiere especial relevancia en carreras de la salud, donde la formación profesional exige una progresión desde la comprensión y organización de contenidos hacia la integración de información, la toma de decisiones y la justificación de acciones clínicas o terapéuticas. En este tipo de trayectorias, la etapa formativa puede modificar la función pedagógica esperada del tutor virtual. Estudiantes en etapas intermedias podrían requerir mayor estructura, síntesis y guía para organizar el conocimiento, mientras que estudiantes avanzados podrían valorar con mayor intensidad la reflexión, el contraste de alternativas y la justificación de decisiones. A pesar del crecimiento de la literatura sobre inteligencia artificial educativa, la evidencia sobre intervenciones con IA en profesiones de la salud continúa siendo limitada, especialmente cuando se consideran resultados educativos complejos, trayectorias formativas y diseños pedagógicos ajustados al desarrollo progresivo del razonamiento profesional (11). En esta misma línea, se ha advertido que la IA generativa puede apoyar el pensamiento crítico en educación médica, pero también tensionar la autonomía cognitiva cuando su uso desplaza procesos de análisis, verificación y juicio propio del estudiante (12).

En este contexto, el presente estudio transversal compara la preferencia declarada por estilos de tutoría virtual entre dos cohortes curriculares de estudiantes de Kinesiología e integra indicadores de utilidad percibida, estrategias de estudio y orientación funcional de uso. Las cohortes correspondían a 6.º y 8.º semestre y utilizaron tutores con una arquitectura pedagógica común, aunque contextualizados a los contenidos y demandas de sus respectivas asignaturas. Por ello, el propósito del estudio es describir diferencias situadas entre ambos contextos curriculares y explorar la hipótesis de que las necesidades de mediación tutorial pueden variar a lo largo de la formación, sin atribuir dichas diferencias exclusivamente al nivel formativo.

2. Métodos

2.1 Diseño del estudio

Se realizó un estudio cuantitativo, transversal, observacional y de alcance asociacional, basado en una encuesta autoadministrada aplicada a estudiantes de pregrado de Kinesiología expuestos al uso de tutores virtuales como apoyo académico. El estudio se desarrolló en un contexto naturalista de

implementación educativa, sin asignación aleatoria ni control experimental de la exposición. El factor principal de comparación fue la cohorte curricular, definida por la pertenencia a 6.º u 8.º semestre y a la asignatura correspondiente. Debido a que cada cohorte utilizó tutores configurados con contenidos, actividades y demandas curriculares propias, el semestre, la asignatura y la implementación específica de los tutores se encontraron completamente superpuestos. En consecuencia, los resultados se interpretan como diferencias asociadas a dos contextos curriculares y no como una estimación independiente del efecto del nivel formativo. El estudio fue diseñado para describir preferencias, percepciones y patrones declarados de uso, sin estimar efectos sobre aprendizaje, rendimiento académico, razonamiento clínico, autonomía cognitiva ni transferencia educativa.

2.2 Contexto, población y muestra

La población accesible estuvo conformada por 70 estudiantes regulares de la carrera de Kinesiología de la Universidad Viña del Mar, pertenecientes a dos cohortes curriculares: una de 6.º semestre y otra de 8.º semestre. Cada cohorte correspondía a una asignatura diferente en la que se incorporaron tutores virtuales como recurso de apoyo al aprendizaje autónomo. Se utilizó un muestreo no probabilístico, por invitación censal a todos los estudiantes elegibles. Los criterios de inclusión fueron: pertenecer a 6.º u 8.º semestre de la carrera de Kinesiología, haber tenido acceso a los tutores virtuales dispuestos para la asignatura correspondiente, aceptar el consentimiento informado digital y responder la encuesta con completitud suficiente para el análisis. Se excluyeron registros con más de 30 % de ítems sin respuesta o con inconsistencias evidentes no recuperables. La muestra final estuvo compuesta por 44 estudiantes, correspondientes al 62,9 % de la población invitada. Dado el carácter exploratorio y contextual del estudio, no se realizó un cálculo formal de tamaño muestral orientado a estimar efectos poblacionales. La muestra corresponde a los estudiantes elegibles que aceptaron participar y completaron la encuesta con suficiencia para el análisis. En consecuencia, los resultados deben interpretarse como evidencia asociacional situada en una cohorte específica, sin pretensión de representatividad estadística de otros programas, instituciones o carreras de la salud.

2.3 Descripción de los tutores virtuales

En cada cohorte curricular se dispusieron dos tutores virtuales conversacionales creados mediante la función "Crear GPT" de ChatGPT Plus. Los tutores fueron configurados de acuerdo con los resultados de aprendizaje, contenidos, actividades y bibliografía oficial de cada asignatura. Cada curso contó con un tutor estructurado-explicativo y un tutor reflexivo-socrático, ambos disponibles para uso voluntario y libre. Los estudiantes podían acceder a cualquiera de los dos tutores y decidir autónomamente cuál utilizar, con qué frecuencia y para qué propósito académico. No se asignó una secuencia de uso, un tiempo mínimo de exposición ni una cantidad equivalente de interacciones con cada tutor. Tampoco se registraron el número de sesiones, la duración de las interacciones, el tutor utilizado en cada ocasión ni la proporción de uso correspondiente a cada estilo. El acceso libre permitió integrar los tutores en condiciones naturalistas de uso académico, pero no aseguró que todos los participantes utilizaran o compararan ambos estilos en condiciones equivalentes. En consecuencia, la preferencia declarada puede reflejar una combinación de familiaridad diferencial, autoselección, intensidad de uso, necesidades académicas particulares y afinidad previa con determinado tipo de interacción. Por esta razón, no se interpreta como una medida experimental de preferencia ni como evidencia de superioridad pedagógica de un tutor sobre otro.

Aunque ambas cohortes curriculares contaron con la misma arquitectura pedagógica general —un tutor estructurado-explicativo y un tutor reflexivo-socrático—, los tutores no correspondieron a una intervención idéntica. Cada par fue configurado con los resultados de aprendizaje, contenidos, actividades y bibliografía propios de la asignatura correspondiente. Por tanto, la comparación debe interpretarse como una comparación situada entre dos cohortes expuestas a tutores equivalentes en su orientación pedagógica general, pero contextualizados a asignaturas y demandas curriculares diferentes. Esta superposición impide aislar el efecto independiente del nivel formativo respecto de la

asignatura cursada, la complejidad de las tareas, el corpus curricular y las características específicas de implementación de cada tutor.

En términos operativos, el tutor estructurado-explicativo fue configurado para priorizar la organización progresiva de los contenidos, la entrega de explicaciones directas y comprensibles, la descomposición de problemas o tareas en pasos y el apoyo para elaborar síntesis, esquemas y estrategias de estudio. Su orientación pedagógica buscaba reducir la dificultad inicial de acceso a los contenidos mediante estructura explícita, secuenciación y guía académica. El tutor reflexivo-socrático fue configurado para conducir la interacción mediante preguntas guía, contrapreguntas, contraste de supuestos y solicitud de justificación de las respuestas. Su orientación pedagógica buscaba que el estudiante explicitara su razonamiento, examinara alternativas y revisara sus propias decisiones antes de avanzar en la interacción. También incorporaba preguntas destinadas a favorecer reflexión metacognitiva sobre el proceso seguido para responder.

Para favorecer la reproducibilidad, las instrucciones nucleares utilizadas en la configuración de ambos estilos tutoriales se presentan en el Material suplementario 1. El documento incluye el propósito pedagógico, las reglas de interacción, el tipo de apoyo esperado, la forma de estructurar las respuestas y las restricciones comunes. Las instrucciones corresponden a las configuraciones efectivamente utilizadas durante la implementación y se presentan diferenciando la arquitectura pedagógica compartida del corpus curricular específico de cada asignatura. Ambos tutores fueron configurados con restricciones comunes orientadas a mantener su uso dentro del marco académico de la asignatura. Se estableció que debían responder a partir de los resultados de aprendizaje, contenidos, actividades y bibliografía incorporados en su configuración. Asimismo, no debían reemplazar al docente, emitir diagnósticos médicos, calificar trabajos, resolver evaluaciones formales, entregar apoyo psicológico ni responder consultas ajenas al ámbito curricular definido. No se recolectaron ni analizaron las conversaciones mantenidas entre estudiantes y tutores virtuales.

2.4 Instrumento y variables

La recolección de datos se realizó mediante una encuesta estructurada autoadministrada en Microsoft Forms. El instrumento fue elaborado ad hoc por el investigador a partir de la pregunta de investigación, los objetivos del estudio y las dimensiones consideradas relevantes para caracterizar la experiencia de uso de los tutores virtuales en las asignaturas participantes. El cuestionario fue diseñado para esta implementación educativa específica y no como una escala psicométrica destinada a medir constructos generalizables. El instrumento se organizó en cinco dominios: datos generales; experiencia previa y frecuencia de uso de tutores virtuales; preferencia declarada por estilo de tutoría; utilidad percibida, apoyo percibido al rendimiento y estrategias de estudio asistidas por tutor; y orientación funcional de uso. Esta última dimensión incluyó ítems relacionados con explicación estructurada, guía hacia la respuesta correcta, preguntas reflexivas, justificación de decisiones o respuestas, y elaboración de resúmenes o esquemas a partir de las respuestas del tutor.

La encuesta incluyó preguntas cerradas de selección, ítems tipo Likert de cinco puntos —desde 1 = totalmente en desacuerdo hasta 5 = totalmente de acuerdo— y dos preguntas abiertas destinadas a recoger los aspectos más valorados del tutor virtual y propuestas de mejora. La variable principal se obtuvo mediante una pregunta cerrada en la que los participantes declararon preferir el tutor estructurado-explicativo o el tutor reflexivo-socrático. En consecuencia, esta variable representa una elección declarada puntual y contextual, y no una escala psicométrica ni la medición de un constructo latente. La variable principal de comparación fue la cohorte curricular, codificada como 6.º u 8.º semestre. Como covariables se consideraron la frecuencia de uso y la experiencia previa con tutores virtuales. La frecuencia de uso se registró mediante una pregunta ordinal de autorreporte con las categorías “Nunca”, “Ocasional”, “Semanal”, “Varias veces por semana” y “Diariamente”. Esta variable representó la frecuencia general de utilización de tutores virtuales durante el periodo de implementación y no permitió identificar qué tutor fue utilizado, cuántas veces se accedió a cada uno,

cuánto duraron las interacciones ni si ambos estilos fueron efectivamente comparados. La categoría "Nunca" identificó a estudiantes que, pese a disponer de acceso, declararon no haber utilizado los tutores durante el periodo estudiado. En consecuencia, la frecuencia de uso se consideró una covariable contextual de autorreporte y no una medida objetiva ni tutor-específica de exposición. La preferencia declarada debe interpretarse dentro de estas condiciones de exposición no controlada. Su asociación con la cohorte curricular no permite determinar si las diferencias observadas derivaron de una comparación equivalente entre estilos tutoriales.

Para los resultados secundarios se construyeron tres puntajes compuestos exploratorios a partir de agrupaciones predefinidas de ítems: utilidad percibida, estrategias de estudio asistidas por tutor y apoyo percibido al rendimiento. Los indicadores relativos a la orientación funcional del tutor fueron analizados individualmente. La consistencia interna de los tres puntajes compuestos se estimó mediante alfa de Cronbach en la muestra estudiada. Estos coeficientes se utilizaron únicamente para describir el grado de relación interna entre los ítems y no como evidencia de validez de contenido, constructo o criterio. El instrumento no fue sometido a validación psicométrica formal ni a análisis factorial. Por tanto, sus resultados deben interpretarse como indicadores contextuales y exploratorios de preferencia, percepción y uso declarado. La redacción completa de los ítems, sus formatos de respuesta y la composición de los puntajes compuestos exploratorios se presentan en el Material suplementario 2.

2.5 Procedimiento de recolección de datos

La implementación educativa de los tutores virtuales se realizó como parte del apoyo académico regular de las asignaturas involucradas. La recolección de datos con fines de investigación se efectuó posterior a la aprobación del Comité Ético Científico de la Universidad Viña del Mar. Durante enero de 2026 se envió a los estudiantes una invitación institucional con el enlace a la encuesta en Microsoft Forms. El uso de los tutores fue voluntario y no formó parte de una actividad obligatoria, secuencia experimental o dosificación pedagógica preestablecida. No se solicitó a los estudiantes utilizar ambos tutores antes de responder la encuesta. Antes de responder, los participantes accedieron a un consentimiento informado digital en el que se explicaban los objetivos del estudio, el carácter voluntario de la participación, la confidencialidad de los datos y la ausencia de consecuencias académicas asociadas a participar o no participar. Solo quienes aceptaron el consentimiento informado pudieron continuar con la encuesta. Los datos fueron exportados desde Microsoft Forms a una base anonimizada para su análisis.

2.6 Gestión y resguardo de datos

Los datos fueron tratados de forma anónima o seudonimizada para el análisis. No se recolectaron identificadores directos ni registros de conversaciones con los tutores virtuales. Las respuestas abiertas fueron revisadas para detectar y eliminar potenciales elementos identificatorios antes del análisis cualitativo. La base de datos fue almacenada con acceso restringido al investigador responsable y los resultados se reportaron de forma agregada.

2.7 Análisis de datos

Se realizó un análisis descriptivo de la muestra mediante frecuencias, porcentajes, medias y desviaciones estándar, según la naturaleza de las variables. La asociación entre cohorte curricular —8.º frente a 6.º semestre— y preferencia declarada por el tutor reflexivo-socrático se examinó mediante la prueba exacta de Fisher. Se reportaron la odds ratio y su intervalo de confianza al 95 %. Adicionalmente, se ajustó un modelo de regresión logística penalizada de Firth, utilizando la preferencia por el tutor reflexivo-socrático como variable dependiente. El modelo incorporó como predictores la cohorte curricular, la frecuencia general de uso autorreportada y la experiencia previa con tutores virtuales. Se utilizó regresión penalizada debido al tamaño de la muestra y a la presencia de celdas pequeñas en la tabla de contingencia principal, condiciones en las que la regresión logística

convencional puede producir estimaciones inestables. El modelo se mantuvo deliberadamente parsimonioso e incluyó únicamente las covariables predefinidas consideradas relevantes para interpretar la asociación principal. Los intervalos de confianza se estimaron mediante perfil de verosimilitud penalizada. La asignatura y la configuración específica de los tutores no fueron incorporadas como covariables, debido a que coincidían completamente con la cohorte curricular y sus efectos no podían estimarse de manera independiente. Por esta razón, el estimador de cohorte se interpretó como una comparación contextual entre los dos grupos estudiados y no como el efecto aislado del nivel formativo. La frecuencia de uso correspondió a una medida general de autorreporte y no a una medida específica de exposición a cada tutor. En consecuencia, su incorporación al modelo no permitió controlar la familiaridad diferencial, la intensidad de las interacciones ni la comparación efectiva entre ambos estilos tutoriales. Los estimadores ajustados se interpretaron considerando el carácter exploratorio del estudio, el tamaño muestral y la amplitud de los intervalos de confianza.

Como análisis de sensibilidad, se repitió el modelo excluyendo a los dos participantes que declararon no haber utilizado los tutores virtuales durante el periodo de implementación. El objetivo fue determinar si la diferencia principal entre cohortes dependía de la inclusión de estos casos. Este análisis no fue diseñado para corregir la ausencia de exposición equivalente, ya que tampoco se dispuso de registros objetivos o específicos de uso para los participantes que declararon haber utilizado los tutores. Para los tres puntajes compuestos exploratorios —utilidad percibida, estrategias de estudio asistidas por tutor y apoyo percibido al rendimiento— se estimó la consistencia interna mediante alfa de Cronbach. El análisis se complementó con correlaciones ítem-total corregidas y con el cálculo del alfa si se eliminaba cada ítem. Para el puntaje compuesto por dos reactivos se examinó también la correlación entre ambos ítems. Estos análisis se utilizaron para describir el comportamiento interno de los puntajes en la muestra estudiada y no como evidencia de validez, unidimensionalidad o consolidación psicométrica del instrumento. No se realizaron análisis factoriales, dado que el estudio no fue diseñado como un proceso de validación instrumental. Los tres puntajes compuestos y los cinco indicadores funcionales individuales se compararon entre cohortes mediante la prueba U de Mann–Whitney bilateral. Para cada comparación se reportó Cliff's delta como tamaño de efecto no paramétrico. Los valores positivos indicaron puntuaciones mayores en la cohorte de 8.º semestre y los valores negativos, puntuaciones mayores en la cohorte de 6.º semestre. Las ocho comparaciones secundarias se consideraron una única familia analítica y se corrigieron mediante el procedimiento secuencial de Holm para controlar el error familiar derivado de la multiplicidad. Se reportaron los valores de p sin ajustar y ajustados. El resultado principal de preferencia por estilo tutorial no fue incluido en esta corrección, debido a que correspondió a la variable primaria y fue analizado mediante una hipótesis diferenciada. Los resultados secundarios se interpretaron considerando conjuntamente los valores ajustados, la dirección y magnitud de Cliff's delta y el carácter exploratorio del estudio.

También se examinó la asociación entre frecuencia general de uso autorreportada y utilidad percibida mediante el coeficiente rho de Spearman. Debido al diseño transversal, esta relación fue interpretada sin establecer dirección causal.

Las respuestas abiertas fueron examinadas mediante análisis de contenido temático con una aproximación deductivo-inductiva. El procedimiento incluyó la lectura completa de las respuestas, la construcción de categorías temáticas acotadas, la codificación de los contenidos y la síntesis de los temas principales. Los resultados cualitativos se utilizaron para contextualizar los hallazgos cuantitativos y no como evidencia independiente de resultados educativos. Todos los contrastes fueron bilaterales y se utilizó un nivel de significación de 0,05. Los análisis estadísticos se realizaron en Python, versión 3.13. El código empleado para el procesamiento y análisis de los datos, junto con una versión anonimizada de la matriz analítica, se encuentra disponible previa solicitud razonable al autor correspondiente, de acuerdo con los compromisos éticos y de confidencialidad establecidos para el estudio.

2.8 Consideraciones éticas

El estudio fue aprobado por el Comité Ético Científico para la Investigación en Seres Humanos, Animales y Todo Otro Ser Vivo de la Universidad Viña del Mar, mediante Acta de Evaluación Ética N.º 09-26, emitida el 9 de enero de 2026, correspondiente al proyecto "Preferencias y efectos del uso de tutores virtuales académicos y socráticos en estudiantes de Kinesiología de la Universidad Viña del Mar". La resolución estableció que los procedimientos propuestos no presentaban observaciones éticas. Todos los participantes otorgaron consentimiento informado digital antes de responder la encuesta. La participación fue voluntaria, anónima y sin consecuencias académicas. Se resguardó la confidencialidad de los datos y los resultados se presentan únicamente de forma agregada.

3. Resultados

3.1 Participantes y patrón de uso

Respondieron 44 estudiantes de un total de 70 invitados, lo que corresponde a una tasa de respuesta global de 62,9 %. La muestra estuvo compuesta por 20 estudiantes de 6.º semestre y 24 estudiantes de 8.º semestre. La edad promedio fue de 24,8 años (DE = 4,0; n = 43), considerando como dato perdido un registro implausible ingresado como "2022". La distribución por género fue equilibrada en la muestra total, con 22 mujeres y 22 hombres. La experiencia previa con tutores virtuales fue frecuente en ambos grupos: 17/20 estudiantes de 6.º semestre y 18/24 estudiantes de 8.º semestre reportaron haber utilizado previamente este tipo de herramienta. La frecuencia general de uso autorreportada mostró un patrón heterogéneo. En la cohorte de 6.º semestre predominó el uso ocasional y varias veces por semana, sin estudiantes que reportaran uso diario. En la cohorte de 8.º semestre predominó el uso ocasional; dos estudiantes declararon no haber utilizado los tutores y tres reportaron uso diario. Esta variable no permitió identificar cuál de los dos tutores fue utilizado ni la intensidad de exposición específica a cada estilo. Los dos casos clasificados como "Nunca" se mantuvieron en la descripción de la muestra y en el análisis principal, dado que pertenecían a la cohorte elegible, tuvieron acceso al recurso y respondieron la pregunta de preferencia. Su influencia fue examinada mediante un análisis de sensibilidad. La edad se reporta con n=43 debido a un registro implausible ingresado como "2022", tratado como dato perdido solo para esta variable. La frecuencia corresponde al uso general autorreportado de tutores virtuales y no distingue entre el tutor estructurado-explicativo y el tutor reflexivo-socrático.

Tabla 1. Características de la muestra y patrón de uso.

Característica	6.º semestre (n=20)	8.º semestre (n=24)	Total (n=44)
Edad, media (DE)	24,1 (4,2) [n=19]	25,5 (3,8) [n=24]	24,8 (4,0) [n=43]
Mujeres, n (%)	7 (35,0)	15 (62,5)	22 (50,0)
Hombres, n (%)	13 (65,0)	9 (37,5)	22 (50,0)
Experiencia previa con tutores virtuales, n (%)	17 (85,0)	18 (75,0)	35 (79,5)
Frecuencia general de uso autorreportada, n (%)	—	—	—
Nunca	0 (0,0)	2 (8,3)	2 (4,5)
Ocasional	8 (40,0)	9 (37,5)	17 (38,6)
Semanal	5 (25,0)	5 (20,8)	10 (22,7)
Varias veces/semana	7 (35,0)	5 (20,8)	12 (27,3)
Diariamente	0 (0,0)	3 (12,5)	3 (6,8)

3.2 Preferencia por estilo de tutor virtual

Se observó una diferencia marcada en la preferencia declarada por el tutor reflexivo-socrático entre las dos cohortes curriculares. En la cohorte de 6.º semestre, 2 de 20 estudiantes (10,0 %) prefirieron este estilo de tutoría, mientras que en la cohorte de 8.º semestre la preferencia fue reportada por 17 de 24 estudiantes (70,8 %). La diferencia absoluta entre grupos fue de 60,8 puntos porcentuales y la razón de proporciones fue 7,08. La prueba exacta de Fisher mostró una asociación estadísticamente significativa entre cohorte curricular y preferencia por el tutor reflexivo-socrático ($p < 0,001$), con un OR de 21,86 (IC95 %: 3,97–120,31).

En el modelo de regresión logística penalizada de Firth, ajustado por frecuencia de uso y experiencia previa con tutores virtuales, la diferencia entre cohortes se mantuvo. Los estudiantes de la cohorte de 8.º semestre presentaron mayores odds de preferir el tutor reflexivo-socrático que los estudiantes de la cohorte de 6.º semestre (aOR = 14,79; IC95 %: 3,64–85,67; $p < 0,001$). La frecuencia de uso (aOR = 1,14; IC95 %: 0,57–2,37; $p = 0,716$) y la experiencia previa con tutores virtuales (aOR = 0,61; IC95 %: 0,09–4,03; $p = 0,609$) no mostraron asociaciones estadísticamente significativas con la preferencia declarada. El estimador de cohorte representa la diferencia conjunta entre los contextos estudiados y no permite separar el nivel formativo de la asignatura y de la configuración específica de los tutores.

El análisis de sensibilidad, realizado tras excluir a los dos participantes que declararon no haber utilizado los tutores virtuales, mostró un patrón consistente con el análisis principal (aOR = 12,87; IC95 %: 3,11–75,30; $p < 0,001$). Este resultado indica que la diferencia observada entre cohortes no dependió de la inclusión de esos dos casos. Sin embargo, el análisis no permite corregir la exposición diferencial entre tutores, ya que no se dispuso de registros objetivos ni tutor-específicos de utilización entre quienes declararon haberlos usado.

Tabla 2. Preferencia declarada por estilo de tutoría virtual y modelos de asociación.

Variable	6.º semestre (n = 20)	8.º semestre (n = 24)	OR/aOR Firth (IC95 %)	p
Preferencia por tutor estructurado-explicativo, n (%)	18 (90,0)	7 (29,2)	—	—
Preferencia por tutor reflexivo-socrático, n (%)	2 (10,0)	17 (70,8)	OR = 21,86 (3,97–120,31)	<0,001
Diferencia absoluta en preferencia por tutor reflexivo-socrático	—	+60,8 puntos porcentuales	—	—
Razón de proporciones para preferencia por tutor reflexivo-socrático	—	7,08	—	—
Modelo ajustado Firth: cohorte de 8.º vs. cohorte de 6.º semestre	—	—	aOR = 14,79 (3,64–85,67)	<0,001
Frecuencia de uso, por categoría	—	—	aOR = 1,14 (0,57–2,37)	0,716
Experiencia previa con tutores virtuales, sí vs. no	—	—	aOR = 0,61 (0,09–4,03)	0,609
Análisis de sensibilidad sin participantes que declararon "Nunca"	—	—	aOR = 12,87 (3,11–75,30)	<0,001

OR = odds ratio no ajustado; aOR = odds ratio ajustado; IC95 % = intervalo de confianza al 95 %. El OR no ajustado corresponde a la asociación entre cohorte curricular y preferencia por tutor reflexivo-socrático estimada a partir de la tabla de contingencia. El modelo ajustado se estimó mediante regresión logística penalizada de Firth e incluyó cohorte curricular, frecuencia general de uso autorreportada y experiencia previa con tutores virtuales

Los intervalos de confianza del modelo penalizado corresponden a perfil de verosimilitud penalizada. El análisis de sensibilidad excluyó a los dos participantes que declararon no haber utilizado el tutor virtual durante el periodo de implementación. Debido a la superposición completa entre cohorte, asignatura y configuración de los tutores, el estimador no representa el efecto independiente del nivel formativo. El ajuste por frecuencia general de uso y el análisis de sensibilidad no controlan la exposición específica a cada tutor.

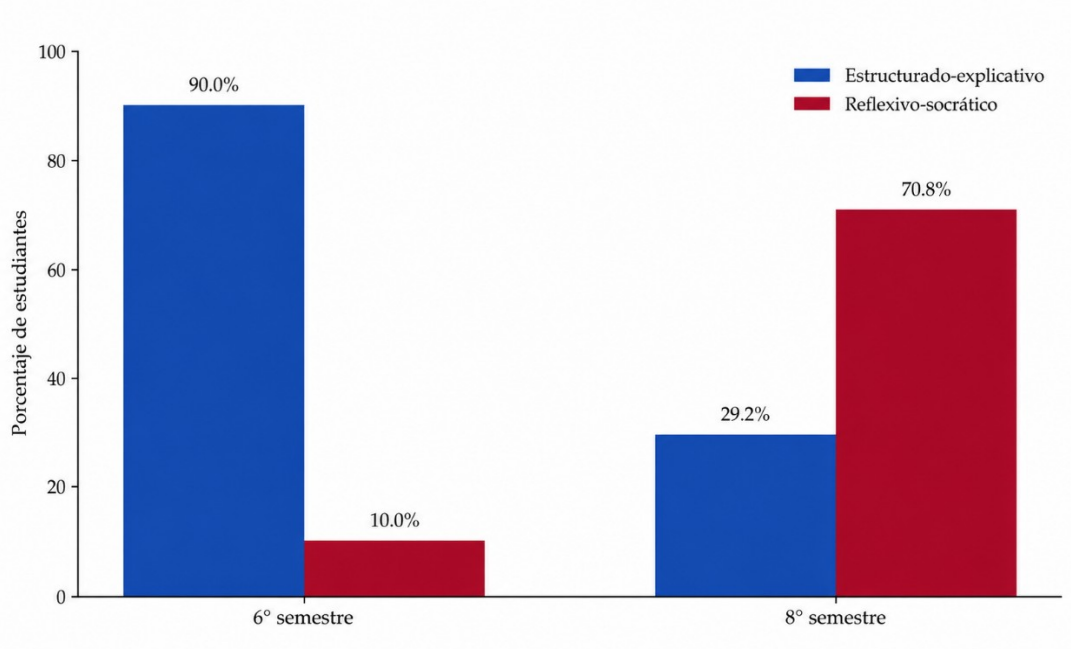


Figura 1. Preferencia declarada por estilo de tutoría virtual según cohorte curricular.

3.3 Utilidad percibida, estrategias de estudio e indicadores funcionales

Los tres puntajes compuestos exploratorios presentaron valores elevados en ambas cohortes. La utilidad percibida estuvo compuesta por seis ítems y mostró una consistencia interna de $\alpha = 0{,}790$. Las correlaciones ítem-total corregidas variaron entre 0,401 y 0,694, mientras que los coeficientes alfa obtenidos al eliminar cada ítem oscilaron entre 0,720 y 0,792. La exclusión de un reactivo no produjo una mejora relevante de la consistencia interna del puntaje. El puntaje de estrategias de estudio, también compuesto por seis ítems, presentó una consistencia interna más modesta, con $\alpha = 0{,}683$. Las correlaciones ítem-total corregidas variaron entre 0,282 y 0,658, y los coeficientes alfa obtenidos al eliminar cada ítem oscilaron entre 0,562 y 0,690. Ningún reactivo individual explicó por sí solo la menor consistencia interna observada. Este comportamiento sugiere que la menor homogeneidad puede estar relacionada con la diversidad de estrategias incluidas en el puntaje, entre ellas reformulación de conceptos, elaboración de resúmenes, análisis de perspectivas, reflexión sobre las respuestas, comparación de razonamientos y revisión de errores. Por esta razón, se interpretó como una agrupación exploratoria contextual y no como una escala unidimensional consolidada.

El apoyo percibido al rendimiento estuvo compuesto por dos ítems y presentó una consistencia interna de $\alpha = 0{,}867$. La correlación entre ambos reactivos fue $r = 0{,}770$. Debido a su reducido número de ítems, este resultado se interpretó como un puntaje compuesto breve y no como una escala psicométrica consolidada. En conjunto, estos análisis describen el comportamiento interno de los puntajes en la muestra estudiada, pero no constituyen evidencia de validez o unidimensionalidad del instrumento. Los puntajes compuestos no mostraron diferencias estadísticamente significativas entre cohortes. La utilidad percibida presentó medias de 4,48 (DE = 0,50) en 6.º semestre y 4,41 (DE = 0,67) en 8.º semestre (p sin ajustar = 0,923; p ajustado por Holm = 1,000; $\Delta = -0{,}02$). Las estrategias de estudio

asistidas por tutor presentaron medias de 4,43 (DE = 0,56) y 4,19 (DE = 0,56), respectivamente (p sin ajustar = 0,179; p ajustado = 0,716; Δ = −0,24). El apoyo percibido al rendimiento también fue comparable entre cohortes, con medias de 4,25 (DE = 0,72) y 4,17 (DE = 0,88) (p sin ajustar = 0,912; p ajustado = 1,000; Δ = −0,02). En los indicadores funcionales individuales, ambas cohortes presentaron puntuaciones altas en explicaciones detalladas y estructuradas: 4,70 (DE = 0,57) en 6.º semestre y 4,79 (DE = 0,51) en 8.º semestre (p sin ajustar = 0,524; p ajustado = 1,000; Δ = 0,08). La guía directa hacia la respuesta correcta presentó medias de 4,40 (DE = 0,88) y 4,75 (DE = 0,61), respectivamente (p sin ajustar = 0,100; p ajustado = 0,501; Δ = 0,23). La cohorte de 8.º semestre presentó puntuaciones superiores en los ítems referidos a preguntas que promueven reflexión —4,15 (DE = 0,99) frente a 4,67 (DE = 0,70)— y justificación de decisiones o respuestas —4,45 (DE = 0,94) frente a 4,83 (DE = 0,48)—. Sin embargo, ninguna de estas diferencias conservó significación estadística después del ajuste de Holm: preguntas reflexivas, p sin ajustar = 0,046, p ajustado = 0,324 y Δ = 0,30; justificación de decisiones o respuestas, p sin ajustar = 0,047, p ajustado = 0,324 y Δ = 0,27.

La diferencia de mayor magnitud entre los resultados secundarios correspondió a la elaboración de resúmenes o esquemas. La media fue de 4,65 (DE = 0,59) en 6.º semestre y 3,75 (DE = 1,29) en 8.º semestre, con un tamaño de efecto de Δ = −0,44. Aunque el valor sin ajustar fue p = 0,007, la diferencia no alcanzó el umbral convencional de significación después del ajuste de Holm (p ajustado = 0,057). En consecuencia, ninguna de las ocho comparaciones secundarias entre cohortes conservó significación estadística tras el control por multiplicidad. Estos resultados deben interpretarse como patrones exploratorios de autorreporte y no como evidencia confirmatoria de diferencias funcionales entre cohortes. La frecuencia general de uso autorreportada se asoció positivamente con la utilidad percibida del tutor virtual (rho de Spearman = 0,399; p = 0,007). Los participantes que declararon una mayor frecuencia de uso también informaron una mayor utilidad percibida. Debido al diseño transversal, esta asociación no permite establecer si una mayor valoración del tutor favoreció su utilización, si el uso más frecuente aumentó la percepción de utilidad o si ambas variables estuvieron influidas por factores personales o académicos no medidos.

Tabla 3. Puntajes compuestos exploratorios e indicadores funcionales autorreportados.

Resultado	6.º semestre, media (DE)	8.º semestre, media (DE)	p sin ajustar	p ajustado por Holm	Cliff's Δ
Utilidad percibida, α = 0,79	4,48 (0,50)	4,41 (0,67)	0,923	1,000	−0,02
Estrategias de estudio, α = 0,68	4,43 (0,56)	4,19 (0,56)	0,179	0,716	−0,24
Apoyo percibido al rendimiento, α = 0,87	4,25 (0,72)	4,17 (0,88)	0,912	1,000	−0,02
Explicaciones detalladas y estructuradas	4,70 (0,57)	4,79 (0,51)	0,524	1,000	0,08
Guía directa hacia la respuesta correcta	4,40 (0,88)	4,75 (0,61)	0,100	0,501	0,23
Preguntas que promueven reflexión	4,15 (0,99)	4,67 (0,70)	0,046	0,324	0,30
Justificación de decisiones/respuestas	4,45 (0,94)	4,83 (0,48)	0,047	0,324	0,27
Elaboración de resúmenes/esquemas	4,65 (0,59)	3,75 (1,29)	0,007	0,057	−0,44

Las comparaciones entre cohortes se realizaron mediante la prueba U de Mann–Whitney bilateral. Las ocho comparaciones secundarias se corrigieron mediante el procedimiento secuencial de Holm. Cliff's delta (Δ) se reportó como tamaño de efecto no paramétrico; los valores positivos indican puntuaciones mayores en la cohorte de 8.º semestre y los negativos, puntuaciones mayores en la cohorte de 6.º semestre. La utilidad percibida y las estrategias de estudio estuvieron compuestas por seis ítems cada una, mientras que el apoyo percibido al

rendimiento estuvo compuesto por dos ítems. Los coeficientes alfa describen la consistencia interna observada en esta muestra y no constituyen evidencia de validez, unidimensionalidad o consolidación psicométrica de los puntajes. Para el puntaje de apoyo percibido al rendimiento, la correlación entre los dos reactivos fue r = 0,770. Ninguna comparación secundaria conservó significación estadística después del ajuste por multiplicidad.

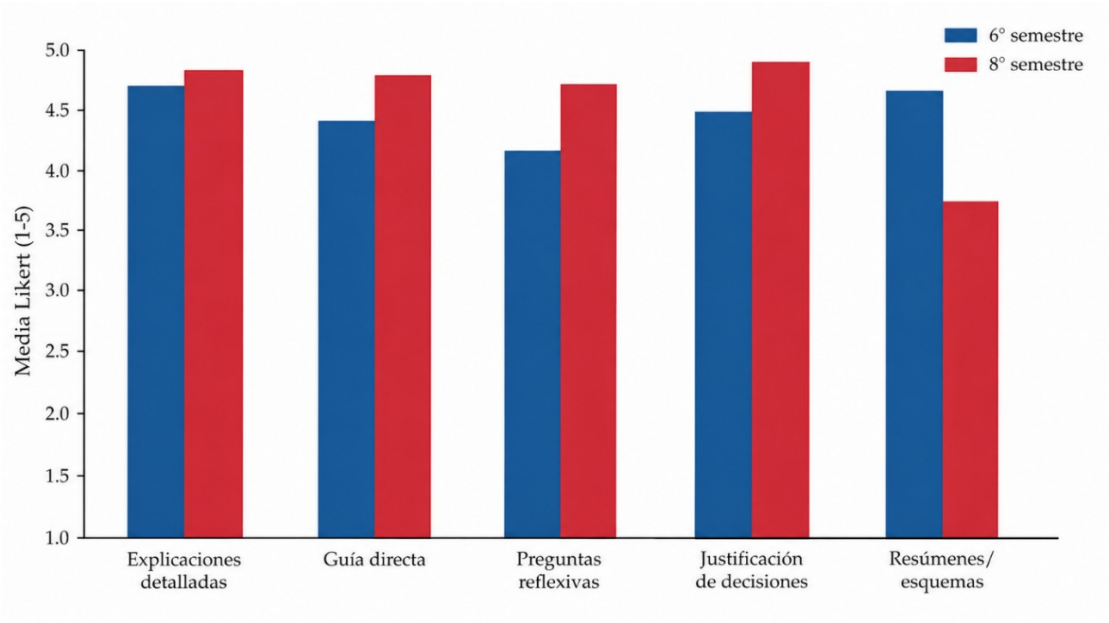


Figura 2. Perfil descriptivo de los indicadores funcionales autorreportados.

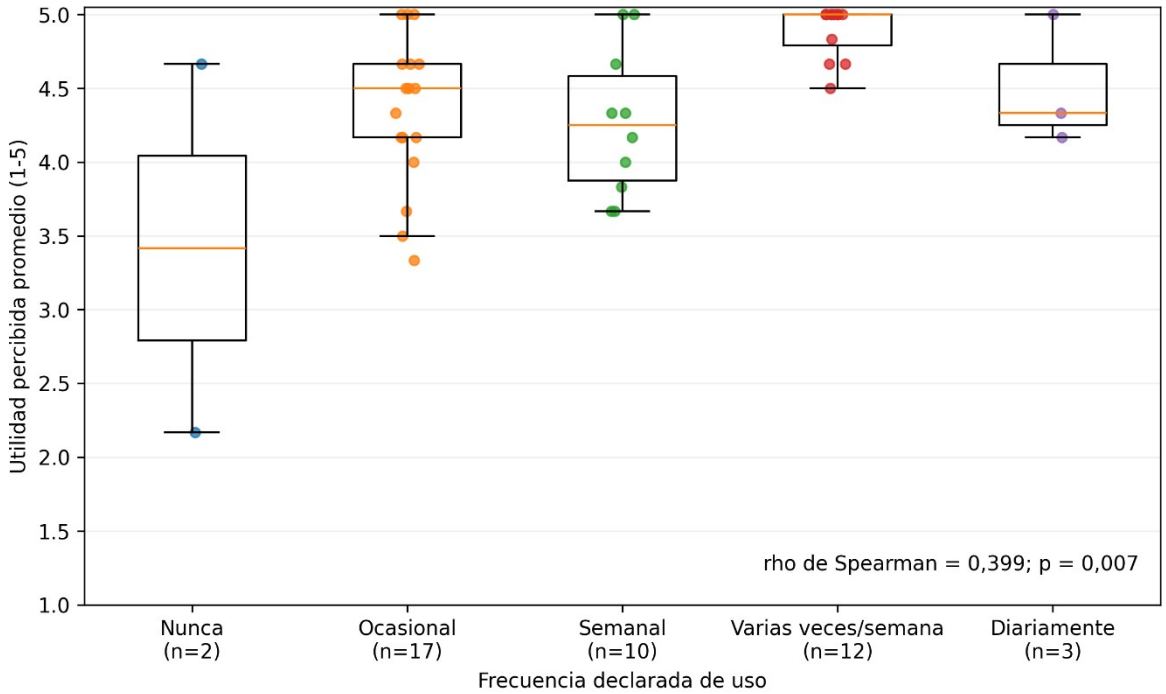


Figura 3. Distribución de la utilidad percibida según frecuencia general de uso autorreportada de tutores virtuales.

Tabla 4. Síntesis temática de respuestas abiertas.

Tema	Interpretación	Frecuencia	Ejemplos breves
Claridad, estructura y explicación sencilla	El tutor es valorado como apoyo para ordenar y comprender contenidos.	Alta	"explicaciones sencillas y detalladas"; "orden y estructura"; "explica detalladamente".
Resolución de dudas y precisión	Los estudiantes valoran respuestas rápidas, específicas y alineadas con contenidos de la asignatura.	Alta	"responder mis dudas"; "respuesta precisa"; "relacionado con la materia".
Reflexión, razonamiento y análisis de errores	Los participantes describen usos o valoraciones vinculados con cuestionamiento, revisión de errores y razonamiento, con mayor presencia en la cohorte de 8.º semestre.	Moderada-alta	"cuestionar mi aprendizaje"; "cuestionar mis errores"; "ayuda a razonar".
Apoyo para estudio y evaluación	El tutor se usa para preparación de certámenes, resúmenes, flashcards o pruebas tipo.	Moderada	"estudio pre prueba"; "pruebas tipo"; "resúmenes, flash cards".
Mejoras en personalización, formato y recursos visuales	Las mejoras solicitadas apuntan a interfaz, imágenes, app propia y banco de datos personalizado.	Moderada	"más personalizado"; "uso de imágenes"; "formato más amigable"; "app independiente".
Mayor ajuste curricular/bibliográfico	Solicitan más apego al syllabus, clases, bibliografía y criterios docentes.	Moderada	"apegado al syllabus"; "bibliografías certeras"; "más específico en kinesiología".

Las categorías corresponden a una síntesis temática de respuestas abiertas y se presentan para contextualizar los resultados cuantitativos. Las referencias a reflexión, razonamiento y aplicación clínica corresponden a expresiones de los participantes y no a resultados educativos evaluados objetivamente.

3.4 Respuestas abiertas

Las respuestas abiertas permitieron contextualizar las valoraciones declaradas en los ítems cuantitativos. Los aspectos más mencionados se agruparon en claridad, estructura, rapidez, resolución de dudas y apoyo para estudiar. En la cohorte de 8.º semestre aparecieron con mayor frecuencia expresiones vinculadas con reflexión, análisis del error, razonamiento y aplicación clínica, como la posibilidad de cuestionar el propio aprendizaje, revisar errores o aplicar conocimientos en situaciones clínicas. Estas expresiones representan percepciones y usos descritos por los participantes; no constituyen medidas directas de razonamiento clínico, reflexión efectiva ni transferencia del aprendizaje. Las mejoras sugeridas se concentraron en mayor personalización, incorporación de recursos visuales, generación de actividades tipo quiz o certamen, un formato más amigable y mayor correspondencia con el syllabus, la bibliografía y los contenidos de clase. En conjunto, ambas cohortes mantuvieron una alta valoración de la estructura y la explicación, junto con patrones descriptivos diferentes en algunas funciones tutoriales declaradas. La cohorte de 8.º semestre presentó puntuaciones superiores en preguntas reflexivas y justificación, mientras que la de 6.º semestre mostró una puntuación mayor en elaboración de resúmenes o esquemas. Sin embargo, ninguna de las ocho comparaciones secundarias conservó significación estadística después del ajuste

de Holm. Por tanto, estos resultados deben interpretarse como orientaciones funcionales percibidas y no como diferencias confirmatorias entre cohortes ni como evidencia de cambios objetivos en reflexión, autonomía o razonamiento clínico.

4. Discusión

El presente estudio comparó la preferencia declarada por estilos de tutoría virtual entre dos cohortes curriculares de estudiantes de Kinesiología e integró indicadores de utilidad percibida, estrategias de estudio y orientación funcional de uso. El principal hallazgo fue una diferencia marcada entre ambas cohortes: en 6.º semestre predominó la preferencia por el tutor estructurado-explicativo, mientras que en 8.º semestre la mayoría declaró preferir el tutor reflexivo-socrático. Esta diferencia se mantuvo tras ajustar por frecuencia general de uso y experiencia previa con tutores virtuales. Sin embargo, debido a que cada cohorte correspondía a una asignatura diferente y utilizó tutores contextualizados a contenidos y demandas curriculares propias, el hallazgo no permite identificar un efecto independiente del nivel formativo.

Los estudiantes de ambas cohortes mantuvieron una alta valoración de la estructura, la explicación detallada y la guía académica. La cohorte de 8.º semestre presentó puntuaciones superiores en los ítems relacionados con preguntas reflexivas y justificación, mientras que la de 6.º semestre mostró una puntuación mayor en elaboración de resúmenes o esquemas. No obstante, ninguna de las ocho comparaciones secundarias conservó significación estadística después del ajuste de Holm. La diferencia de mayor magnitud correspondió a la elaboración de resúmenes o esquemas ($\Delta = -0{,}44$; p ajustado = 0,057), por lo que debe considerarse una señal exploratoria y no un hallazgo confirmatorio. Estos patrones pueden estar relacionados con necesidades diferentes de mediación tutorial, pero también con la naturaleza de las asignaturas, la complejidad de las tareas, las demandas evaluativas y las características específicas de los tutores implementados. Estos resultados pueden situarse en relación con la literatura sobre agentes conversacionales educativos, que ha descrito su potencial para entregar apoyo inmediato, facilitar el acceso a información, personalizar interacciones y acompañar estrategias de estudio (1, 2). A la vez, la evidencia reciente indica que el valor educativo de estos sistemas depende de su diseño pedagógico y de su capacidad para favorecer procesos de autorregulación, monitoreo metacognitivo y retroalimentación significativa (5, 6). Desde esta perspectiva, los tutores virtuales requieren ser analizados de acuerdo con las funciones pedagógicas que buscan promover durante la interacción, además de su disponibilidad tecnológica o aceptación estudiantil.

La mayor preferencia declarada por el tutor reflexivo-socrático en la cohorte de 8.º semestre puede relacionarse conceptualmente con estudios sobre agentes conversacionales socráticos basados en modelos de lenguaje. En esas investigaciones, este tipo de mediación ha sido evaluado en relación con pensamiento reflexivo, razonamiento crítico y desempeño académico bajo condiciones específicas de diseño (3, 4). El presente estudio no evaluó esos resultados y, por tanto, la coincidencia se limita al interés declarado por funciones como las preguntas guía, el contraste de supuestos y la justificación de respuestas. Del mismo modo, la evidencia sobre retroalimentación generada por ChatGPT en tareas de razonamiento clínico aporta un marco pertinente para futuras evaluaciones, pero no permite inferir que las preferencias observadas en esta muestra se traduzcan en mejoras del razonamiento o del aprendizaje (13). En esta misma línea, Corral-Gudino y Marcos plantean el uso de tutores socráticos como una estrategia para preservar el papel activo del estudiante y reducir la delegación de tareas cognitivas en actividades mediadas por IA generativa (14). Esta propuesta es conceptualmente coherente con la orientación reflexivo-socrática utilizada en el presente estudio, aunque sus efectos sobre aprendizaje y autonomía no fueron evaluados en esta muestra.

En carreras de la salud, las funciones vinculadas con explicación, contraste y justificación resultan relevantes porque la formación profesional incorpora progresivamente tareas de integración de información y fundamentación de decisiones clínicas o terapéuticas. En este estudio, la cohorte de

8.º semestre manifestó mayor preferencia declarada por el tutor reflexivo-socrático y presentó puntuaciones descriptivamente superiores en los ítems relacionados con preguntas reflexivas y justificación. Sin embargo, estas diferencias secundarias no conservaron significación estadística después del ajuste por comparaciones múltiples. Por ello, pueden interpretarse como expectativas percibidas de mediación tutorial, pero no como evidencia confirmatoria de una orientación funcional diferenciada ni como demostración de mayor razonamiento clínico, reflexión o autonomía.

El aporte del estudio se ubica en la articulación entre contexto curricular, estilo pedagógico del tutor y educación superior en salud. Una proporción importante de la evidencia sobre inteligencia artificial educativa procede de contextos generales de educación superior, programación o tareas académicas con menor especificidad clínica. La formación en Kinesiología, en cambio, exige integrar conocimiento biomédico, razonamiento funcional, análisis contextual y fundamentación de decisiones vinculadas con problemas de salud. Considerando que la evidencia sobre IA en educación de profesiones de la salud todavía presenta limitaciones metodológicas y resultados heterogéneos (11), este trabajo aporta una descripción inicial de cómo dos cohortes curriculares pueden diferir en las funciones pedagógicas que declaran valorar en un tutor virtual. Una revisión de alcance publicada en esta revista mostró que la evidencia sobre IA en formación clínica se concentra principalmente en escenarios simulados o estructurados, presenta una alta heterogeneidad de intervenciones y resultados, y dispone de escaso seguimiento longitudinal (15). Esta situación refuerza la necesidad de interpretar los resultados del presente estudio dentro de las características específicas de las asignaturas y de la implementación analizada.

La asociación positiva entre frecuencia general de uso autorreportada y utilidad percibida debe interpretarse con cautela. El diseño transversal impide establecer la dirección de esta relación: quienes percibieron mayor utilidad pudieron utilizar más los tutores, mientras que una exposición más frecuente pudo aumentar la familiaridad y la valoración del recurso. Además, la frecuencia fue registrada de manera global y no permitió distinguir el tutor utilizado, la duración de las interacciones ni la exposición relativa a cada estilo. Por ello, esta variable no controla la familiaridad diferencial ni permite establecer que la preferencia declarada surgiera de una comparación equivalente entre ambos tutores. La utilidad percibida y la frecuencia de uso tampoco constituyen evidencia directa de aprendizaje profundo, razonamiento clínico o desempeño académico. Estudios recientes han mostrado que ciertos apoyos de IA pueden mejorar el desempeño durante la tarea o reducir la frustración sin producir ganancias equivalentes de aprendizaje, especialmente cuando el sistema entrega respuestas finales o desplaza procesos cognitivos centrales del estudiante (7, 8). Asimismo, la sobredependencia de sistemas conversacionales de IA ha sido descrita como un riesgo para habilidades como pensamiento crítico, juicio autónomo y toma de decisiones (9).

La integración de los resultados cuantitativos y de las respuestas abiertas permite contextualizar los patrones observados. Los puntajes compuestos exploratorios mostraron alta utilidad percibida, apoyo percibido al rendimiento y uso de estrategias de estudio asistidas por tutor en ambas cohortes. Los indicadores funcionales permitieron describir matices en las valoraciones declaradas, aunque ninguna de las diferencias secundarias conservó significación estadística después del ajuste por multiplicidad. Estos resultados corresponden a indicadores contextuales de autorreporte y no a mediciones validadas de constructos educativos. Asimismo, las menciones a reflexión, análisis del error, razonamiento y aplicación clínica en las respuestas abiertas describen la experiencia percibida por los participantes, pero no verifican que estos procesos hayan ocurrido ni que hayan mejorado como resultado de la interacción con los tutores. Las respuestas abiertas mostraron que los estudiantes valoraron la claridad, la estructura, la rapidez, la resolución de dudas y el apoyo para estudiar; en la cohorte de 8.º semestre aparecieron con mayor frecuencia referencias a reflexión, análisis del error, razonamiento y aplicación clínica. Esta información cualitativa contextualiza los patrones descriptivos, pero no confirma diferencias funcionales entre cohortes ni compensa la ausencia de resultados educativos objetivos.

Los patrones observados permiten plantear hipótesis para el diseño y evaluación futura de tutores virtuales. En contextos curriculares donde se valoren funciones de organización y síntesis, podría examinarse experimentalmente una mayor presencia de explicaciones estructuradas, ejemplos guiados y comprobaciones de comprensión. En tareas que requieran integración y justificación, podría evaluarse una mediación basada en preguntas guía, contraste de alternativas, identificación de supuestos y análisis de errores. Estas posibilidades se derivan de preferencias y usos declarados en dos cohortes específicas y no constituyen recomendaciones sobre la configuración más efectiva ni evidencia de una progresión tutorial óptima.

Otra línea de investigación corresponde a evaluar si las orientaciones explícitas sobre el uso académico de tutores virtuales modifican la forma en que los estudiantes interactúan con estas herramientas. Solicitar explicaciones, formular preguntas guía, contrastar respuestas y verificar información representan estrategias de interacción diferentes, cuyo valor educativo no fue examinado en este estudio. Su incorporación y evaluación pueden fundamentarse en la literatura sobre autorregulación, delegación cognitiva y uso responsable de IA. En carreras de la salud, esta pregunta adquiere interés por la relevancia profesional de verificar información y fundamentar las decisiones, aunque sus efectos deberán ser evaluados mediante diseños específicos.

La principal limitación interpretativa del estudio corresponde a la relación estrecha entre cohorte curricular, nivel formativo, asignatura y configuración específica de los tutores. Aunque ambas cohortes contaron con un tutor estructurado-explicativo y un tutor reflexivo-socrático, cada par fue construido a partir del corpus curricular propio de la asignatura correspondiente. En consecuencia, la asociación observada entre cohorte y preferencia por estilo de tutoría no puede atribuirse exclusivamente a la etapa formativa del estudiante. También puede estar influida por la complejidad de las tareas, las demandas evaluativas, el tipo de contenido, la familiaridad con la herramienta y las características específicas de implementación de cada tutor. El diseño transversal impide establecer causalidad o examinar cambios intraindividuales a lo largo de la trayectoria curricular.

Todas las variables analizadas correspondieron a autorreportes de preferencia, utilidad, estrategias de estudio y orientación funcional de uso. No se incorporaron medidas objetivas de aprendizaje, rendimiento académico, calidad de la justificación, razonamiento clínico, pensamiento reflexivo, autonomía cognitiva ni transferencia a tareas auténticas. En consecuencia, las referencias de los participantes a reflexión, análisis del error o aplicación clínica no permiten concluir que estos procesos hayan ocurrido, mejorado o sido causados por el tutor virtual. Tampoco puede establecerse la efectividad educativa ni la superioridad de uno de los estilos tutoriales.

El cuestionario fue elaborado ad hoc para esta implementación y no fue sometido a un proceso formal de validación de contenido, constructo o criterio. La variable principal correspondió a una pregunta cerrada de preferencia declarada, mientras que los resultados secundarios se construyeron mediante agrupaciones predefinidas de ítems. Los análisis de consistencia interna aportan información descriptiva sobre el comportamiento de estos puntajes en la muestra estudiada, pero no demuestran unidimensionalidad ni sustituyen una evaluación psicométrica completa. La utilidad percibida presentó $\alpha = 0{,}790$, con correlaciones ítem-total corregidas entre 0,401 y 0,694. El puntaje de estrategias de estudio mostró una consistencia interna más modesta, con $\alpha = 0{,}683$ y correlaciones ítem-total corregidas entre 0,282 y 0,658. La eliminación de cualquier reactivo produjo coeficientes entre 0,562 y 0,690, por lo que ningún ítem individual explicó la menor homogeneidad observada. Esta característica puede estar relacionada con la diversidad de estrategias incluidas y refuerza su interpretación como una agrupación exploratoria contextual. El apoyo percibido al rendimiento estuvo integrado por dos ítems, presentó $\alpha = 0{,}867$ y una correlación entre reactivos de $r = 0{,}770$, por lo que debe interpretarse como un puntaje compuesto breve. En consecuencia, estos resultados

expresan indicadores contextuales de percepción y uso declarado y no mediciones estandarizadas de constructos educativos consolidados.

Otra limitación corresponde al tamaño y las características de la muestra. Se trabajó con una muestra pequeña, no probabilística y proveniente de una sola carrera e institución, lo que restringe la generalización de los hallazgos. Aunque se observó una distribución diferente de género entre cohortes, esta variable no se incorporó al modelo principal para mantener la parsimonia del análisis en una muestra reducida. Su posible influencia sobre las preferencias declaradas debería evaluarse en estudios con mayor tamaño muestral.

El uso voluntario y naturalista constituye otra limitación relevante. Aunque los estudiantes tuvieron acceso a ambos tutores, no se exigió una comparación directa ni se estandarizaron el tiempo, la secuencia o la cantidad de interacciones. Tampoco se registraron trazas objetivas de uso que permitieran identificar el tutor utilizado, la duración de cada sesión o la exposición acumulada a cada estilo. La frecuencia autorreportada fue general y, por tanto, no permitió controlar la familiaridad diferencial ni la intensidad de uso específica. En estas condiciones, la preferencia declarada puede estar influida por autoselección, necesidades académicas particulares, motivación previa o exposición desigual.

Los análisis secundarios incluyeron ocho comparaciones entre cohortes, para las que se aplicó el procedimiento secuencial de Holm. Después del control por multiplicidad, ninguna diferencia conservó significación estadística. Aunque el ítem de elaboración de resúmenes o esquemas presentó la diferencia de mayor magnitud, su valor ajustado fue $p = 0,057$. Estos resultados requieren cautela debido al tamaño muestral, la naturaleza autorreportada de las variables y el carácter exploratorio del análisis.

Futuras investigaciones deberían incorporar diseños cruzados o experimentales en los que los mismos participantes utilicen ambos estilos tutoriales con contenidos equivalentes, secuencias contrabalanceadas y tiempos de exposición comparables. También sería necesario registrar, bajo consentimiento explícito y resguardos éticos, el número de sesiones, la duración de las interacciones, el tutor utilizado y los patrones de diálogo. Estas medidas permitirían distinguir una preferencia informada de la familiaridad diferencial y examinar la relación entre formas específicas de interacción y resultados objetivos de aprendizaje, razonamiento clínico simulado o calidad de la justificación. También sería pertinente desarrollar y evaluar instrumentos con evidencias formales de validez y consistencia para medir las dimensiones educativas de interés.

En síntesis, el estudio aporta evidencia situada sobre las preferencias declaradas y las funciones tutoriales valoradas en dos cohortes curriculares de Kinesiología. Ambas cohortes mostraron una alta valoración de la estructura, la guía y la explicación, junto con una diferencia marcada en la preferencia por el estilo tutorial. Los indicadores funcionales secundarios mostraron algunos patrones descriptivos, pero ninguna comparación conservó significación estadística después del ajuste por multiplicidad. Los hallazgos permiten formular preguntas e hipótesis sobre la configuración pedagógica de los tutores conversacionales, pero no sustentan recomendaciones sobre su efectividad ni demuestran efectos sobre aprendizaje, reflexión, autonomía profesional o razonamiento clínico.

5. Conclusiones

- En las dos cohortes curriculares estudiadas se observaron diferencias marcadas en la preferencia declarada por estilos de tutoría virtual. En la cohorte de 8.º semestre predominó la preferencia por el tutor reflexivo-socrático, mientras que en la de 6.º semestre predominó la preferencia por el tutor estructurado-explicativo.

- La estructura, la explicación detallada y la guía académica fueron valoradas en ambas cohortes. Se observaron patrones descriptivos diferentes en preguntas reflexivas, justificación y elaboración de resúmenes o esquemas, pero ninguna de las ocho comparaciones secundarias conservó significación estadística después del ajuste de Holm. Estos resultados describen preferencias y orientaciones funcionales autorreportadas; el estudio no evaluó resultados objetivos de aprendizaje, razonamiento clínico, reflexión, autonomía ni transferencia educativa.
- Dado que cada cohorte correspondía a una asignatura diferente y utilizó tutores contextualizados a demandas curriculares específicas, las diferencias observadas no pueden atribuirse exclusivamente al nivel formativo. Además, la exposición a cada tutor fue voluntaria, no estandarizada y registrada únicamente mediante una frecuencia general de autorreporte, por lo que no puede asegurarse que la preferencia declarada procediera de una comparación equivalente entre ambos estilos.
- Los patrones observados permiten formular la hipótesis de que la relación entre configuración tutorial, contexto curricular, características de la tarea y grado de autonomía esperado merece ser evaluada de manera específica. Futuros estudios cruzados, longitudinales o experimentales, con exposición equivalente y medidas objetivas de aprendizaje y desempeño, deberán determinar si alguna configuración produce resultados educativos diferentes.

Financiación: No hubo financiación.

Declaración de conflicto de interés: El autor declara no tener ningún conflicto de interés.

Uso de inteligencia artificial en el proceso de investigación: Se utilizó ChatGPT Plus para configurar tutores virtuales personalizados mediante la función "Crear GPT", restringidos a los resultados de aprendizaje, contenidos, actividades y bibliografía oficial de las asignaturas participantes. Los estudiantes tuvieron acceso libre a un tutor estructurado-explicativo y a un tutor reflexivo-socrático, sin que se recolectaran ni analizaran conversaciones mantenidas con dichas herramientas. En la fase analítica, se utilizaron Codex y ChatGPT únicamente como apoyo técnico para la generación y depuración de código en Python destinado al procesamiento de datos y cálculos estadísticos. No se ingresaron datos identificatorios de estudiantes en herramientas de inteligencia artificial. Las decisiones metodológicas, la selección de resultados, la interpretación estadística, la redacción final y las conclusiones fueron realizadas y verificadas por el autor, quien asume plena responsabilidad por el contenido del manuscrito.

Contribuciones del autor: I.C.M. realizó la conceptualización, la metodología, el diseño de los tutores virtuales, la elaboración del instrumento de recolección de datos, la gestión ética, la recolección de datos, el análisis formal, la interpretación de los resultados, la redacción del borrador original, la revisión y edición del manuscrito, la visualización de los resultados y la aprobación de la versión final.

Anexos: Material suplementario 1 (Arquitectura pedagógica e instrucciones nucleares de los tutores virtuales) y 2 (Instrumento de recolección de datos).

6. Referencias

1. Okonkwo CW, Ade-Ibijola A. Chatbots applications in education: a systematic review. *Comput Educ Artif Intell*. **2021**, 2, 100033. https://doi.org/10.1016/j.caeai.2021.100033.
2. Abdallah N, Katmah R, Khalaf K, Jelinek HF. Systematic review of ChatGPT in higher education: navigating impact on learning, wellbeing, and collaboration. *Soc Sci Humanit Open*. **2025**, 12, 101866. https://doi.org/10.1016/j.ssaho.2025.101866.
3. Xi L, Zhang Y, Wang Q. Investigating the effects of an LLM-based Socratic conversational agent on students' academic performance and reflective thinking in higher education. *Comput Educ*. **2026**, 241, 105494. https://doi.org/10.1016/j.compedu.2025.105494.
4. Fakour H, Imani M. Socratic wisdom in the age of AI: a comparative study of ChatGPT and human tutors in enhancing critical thinking skills. *Front Educ*. **2025**, 10, 1528603. https://doi.org/10.3389/feduc.2025.1528603.

5. Chang DH, Lin MP-C, Hajian S, Wang QQ. Educational design principles of using AI chatbot that supports self-regulated learning in education: goal setting, feedback, and personalization. *Sustainability*. **2023**, 15, 12921. https://doi.org/10.3390/su151712921.
6. Guan R, Raković M, Chen G, Gašević D. How educational chatbots support self-regulated learning? A systematic review of the literature. *Educ Inf Technol*. **2025**, 30, 4493–4518. https://doi.org/10.1007/s10639-024-12881-y.
7. Bassner P, Lenk-Ostendorf B, Beinstingel R, Wasner T, Krusche S. Less stress, better scores, same learning: the dissociation of performance and learning in AI-supported programming education. *Comput Educ Artif Intell*. **2026**, 10, 100537. https://doi.org/10.1016/j.caeai.2025.100537.
8. Bastani H, Bastani O, Sungu A, Ge H, Kabakcı Ö, Mariman R. Generative AI without guardrails can harm learning: evidence from high school mathematics. *Proc Natl Acad Sci U S A*. **2025**, 122, e2422633122. https://doi.org/10.1073/pnas.2422633122.
9. Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn Environ*. **2024**, 11, 28. https://doi.org/10.1186/s40561-024-00316-7.
10. Campos Montenegro I, Martínez Mena M. Alfabetización en IA y experiencia subjetiva con un tutor virtual basado en GPT en estudiantes de Kinesiología: estudio transversal. *Rev Esp Edu Med*. **2026**, 7(4). https://doi.org/10.6018/edumed.715291.
11. Feigerlova E, Hani H, Hothersall-Davies E. A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. *BMC Med Educ*. **2025**, 25, 129. https://doi.org/10.1186/s12909-025-06719-5.
12. Izquierdo-Condoy JS, Arias-Intriago M, Tello-De-la-Torre A, Busch F, Ortiz-Prado E. Generative artificial intelligence in medical education: enhancing critical thinking or undermining cognitive autonomy? *J Med Internet Res*. **2025**, 27, e76340. https://doi.org/10.2196/76340.
13. Çiçek FE, Ülker M, Özer M, Kıyak YS. ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial. *Postgrad Med J*. **2025**, 101, 458–463. https://doi.org/10.1093/postmj/qgae170.
14. Corral-Gudino L, Marcos M. La escritura en educación médica en la era de la inteligencia artificial generativa: recomendaciones para fomentar la creatividad y el pensamiento crítico en un entorno de aprendizaje asistido por IA generativa. *Rev Esp Edu Med*. **2026**, 7(3). https://doi.org/10.6018/edumed.708901.
15. Bonilla Mejia JJ, Cortés Fuenzalida TD, Polanco Aliaga DH, Martínez Carrillo C, Herrera Alcaíno AA. Inteligencia artificial en la formación clínica práctica de estudiantes de medicina de pregrado: una revisión de alcance de aplicaciones, resultados y brechas. *Rev Esp Edu Med*. **2026**, 7(3). https://doi.org/10.6018/edumed.710071.



© 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/)