



AI-Powered Chatbot Assessments in Online Anatomy & Physiology: A Mixed-Methods Study

Jaime Cantú

Austin Community College, United States

Abstract

As online health sciences education expands, concerns continue regarding whether traditional exam-based assessments adequately capture applied understanding, communication skills, and clinical reasoning. This mixed-methods study examines the use of AI-powered chatbot assessments as an alternative to traditional high-stakes exams in a fully online Anatomy and Physiology I course at a higher education institution. Two course sections with identical learning objectives and instructional design differed only in assessment format: one used traditional proctored exams, while the other implemented structured, proctored AI chatbot conversations simulating virtual patient interactions. Quantitative survey data collected at multiple time points examined student confidence, perceived learning, fairness, and assessment-related anxiety, while qualitative open-ended responses explored student experiences with explanation, application, and engagement. Results indicate that students in both formats reported increased confidence over time; however, students in the AI-based assessment section more frequently described deeper conceptual understanding, improved ability to explain anatomical and physiological concepts, and reduced test anxiety. Qualitative findings highlight explanation and dialogue as key mechanisms supporting learning, with students characterizing chatbot assessments as more reflective of real-world healthcare communication. These findings align with constructivist learning theory and cognitive load theory, suggesting that AI-supported conversational assessments may better capture applied competency than traditional exams. This study contributes empirical evidence to ongoing discussions about ethical, equitable, and pedagogically thorough integration of AI into online assessments.

Keywords: artificial intelligence, assessment, anatomy and physiology, online learning, mixed methods

Introduction

Assessment practices in higher education, particularly in foundational and gateway health science courses, have continued to rely heavily on traditional measures of student learning even as instructional delivery models and student demographics have shifted significantly [11,13,14]. Traditional assessments, such as multiple-choice and essay-based exams, have been criticized for prioritizing memorization over application and for inadequately measuring clinical reasoning and communication skills [12,15]. These limitations are especially noticeable in Anatomy and Physiology (A&P), a foundational course for many health science pathways where student success is connected to later academic and clinical preparation [3]. In A&P, students must not only understand complex biological systems but also explain and apply that knowledge in patient-centered contexts.

Recent advances in generative artificial intelligence (AI) offer new opportunities to rethink assessment design. AI-powered chatbot assessments allow students to engage in structured dialogue, justify reasoning, and translate technical language into accessible explanations. Prior research has shown that AI-driven learning environments can enhance engagement, personalize instruction, and improve knowledge retention [8,20]. In medical education specifically, AI-supported tools have been identified as having the potential to support clinical reasoning, diagnostic decision-making, intelligent tutoring, and feedback in physician training [10]. However, empirical research examining AI as a direct assessment mechanism, rather than as tutoring or practice, remains limited, particularly in introductory A&P courses. This article summarizes a mixed-methods study designed to address this gap by examining how AI-powered chatbot assessments function as alternatives to traditional exams in a fully online A&P I course.



Course Context and Assessment Design

The study was conducted across two fully online BIOL 2401 sections taught by the same instructor at a Austin Community College in Austin, Texas, USA. Both sections followed identical learning objectives, pacing, laboratory experiences, grading thresholds, and academic integrity measures. Students completed weekly laboratory assignments, chapter-based quizzes, and extensive student-created video presentations requiring explanation, synthesis, and analogy-based reasoning.

The primary difference between sections was the lecture assessment format. In the traditional section, students completed proctored, time-restricted midterm and final exams. In the AI section, these exams were replaced with structured, proctored, time-restricted AI chatbot conversations. During these conversations, students led a dialogue with a virtual patient, explained anatomical and physiological concepts using correct terminology, and then reframed their explanations in patient-friendly language. The chatbot prompted clarification and elaboration but did not provide answers, aligning assessment with applied competency rather than recall.

This design reflects prior research emphasizing the value of low-stakes, explanation-centered assessment and adaptive feedback in reducing extraneous cognitive load and supporting deeper learning [6,17,18].

Methodology

A convergent mixed-methods design was used, integrating quantitative survey data with qualitative open-ended responses. Participants were drawn from a convenience sample of students enrolled in two existing sections of Anatomy and Physiology. Students were not randomly assigned to the AI-supported or traditional assessment section; rather, they registered for sections through the institution's normal enrollment process. The AI-supported section was identified in the course schedule as using AI as a main assessment tool, while the traditional section included the same general online course expectations but did not include that AI assessment statement.

The course began with 24 enrolled students. By the end of the semester, enrollment was 16 students in the AI-supported section and 19 students in the traditional section. Although formal demographic data were not collected because demographics were not the primary focus of the study, the participant group appeared to be mostly female, which is consistent with the student population commonly represented in health science prerequisite courses.

Four surveys were administered:

- Pre-assessment survey (combined sections; $n = 15$)
- End-of-course survey, AI section ($n = 8$)
- End-of-course survey, traditional section ($n = 4$)
- Post-assessment survey (combined sections; $n = 11$)

Quantitative items used Likert-scale measures to assess confidence, anxiety, perceived fairness, and alignment between assessments and learning. Qualitative items invited students to reflect on what they found easy or difficult and how assessments influenced their understanding.

Because the course was intensive in pace and workload, student participation varied across the semester. Some students completed surveys at the beginning and end of the course but did not consistently participate in the middle, while others engaged with only selected data collection points. As a result, survey response counts differed across instruments. Rather than treating participation as uniform across the full course, the analysis considered each survey as a separate data point while also looking across responses for broader patterns in students' experiences with assessment.

Results

Appendix Tables 1–3 provide a structured overview of the four survey instruments used in this study, including each instrument's timing, participant group, sample size, analytic purpose, selected qualitative excerpts, and descriptive quantitative findings. Because the surveys were administered at different points in the course and were not identical across all instruments, the tables are used to organize the findings rather than to suggest direct statistical comparison across groups. Table 1 outlines the purpose of each survey, Table 2 presents representative de-identified student responses, and Table 3 summarizes selected Likert-scale means related to confidence, explanation, application, anxiety or stress,



and assessment fairness. The following sections discuss the quantitative and qualitative findings separately while drawing on these appendix tables as supporting evidence.

Quantitative Findings

Pre-assessment data indicated that students entered BIOL 2401 with moderate confidence in their ability to succeed, but lower confidence in applying and explaining A&P concepts, alongside moderate-to-high assessment anxiety. This baseline aligns with prior research demonstrating that traditional assessment structures often amplify anxiety in health science students [2]. Pre-assessment results also showed most students agreed that assessments should help them apply learning to real-life situations and that they learn better when required to explain concepts in their own words. These baseline responses suggest that students were already receptive to applied, explanation-centered assessment, while also entering the course with concerns about assessment pressure and unfamiliar formats.

Across end-of-course and post-assessment surveys, students reported increased confidence overall, regardless of assessment format. Post-assessment Likert means clustered in the Agree to Strongly Agree range for items related to application ($M \approx 4.18$ – 4.36 on a 5-point scale), fairness ($M \approx 4.27$), and perceived learning value. Students in the AI section more frequently endorsed items related to explanation and application, suggesting that assessment format influenced how confidence developed rather than whether confidence increased.

Given small sample sizes, particularly in the traditional end-of-course survey, results are best interpreted as descriptive patterns rather than inferential comparisons. Nonetheless, the consistency of favorable ratings across instruments supports the feasibility of applied, AI-supported assessment in online A&P contexts.

Qualitative Findings

Qualitative responses provided insight into why students perceived AI-based assessments as beneficial. Pre-assessment open-ended responses showed that several students distinguished memorization from meaningful understanding. One student explained that previous science exams often required them to “memorize and then regurgitate the information,” making it difficult to remember material later. Another student noted that writing information in their own words helped them learn more effectively. These responses provide important baseline context for interpreting the end-of-course findings, as students entered the course already aware that traditional study and testing practices did not always support long-term understanding.

Students in the AI section frequently emphasized explanation as a learning mechanism: “It pushed me to actually understand the material because I had to explain it in my own words.” This aligns with constructivist learning theory [16,19] and with research demonstrating that explanation and articulation support deeper conceptual understanding [9].

Students also reported reduced anxiety compared to traditional exams: “It felt more like a conversation than a test.” This finding echoes prior research showing that lower-stakes, formative assessment structures can mitigate test-related stress [5,15].

Students in the traditional section valued predictability but acknowledged limitations: “I studied to pass the test, but I don’t know if I could explain everything later.” Across both sections, students demonstrated increased metacognitive awareness, distinguishing memorization from understanding; an outcome emphasized in competency-based medical education literature [10,12].

Discussion

When taking the results together, the quantitative and qualitative findings suggest that assessment format shapes learning orientation. Both traditional and AI-based assessments supported confidence growth; however, AI-powered chatbot assessments uniquely emphasized explanation, reasoning, and communication, competencies central to healthcare education.

These findings align with prior research demonstrating that AI-driven and adaptive assessment systems can reduce extraneous cognitive load, increase engagement, and support applied learning



[7,18,20]. They also reinforce concerns raised in the literature regarding equity and access: several students noted that learning how to interact with the AI itself was part of the challenge, underscoring the need for AI literacy scaffolding [1,4].

Limitations

This study is limited by small sample sizes, reliance on self-reported data, and implementation within a single course and institution. As such, findings should be interpreted as exploratory. However, the convergence of quantitative trends and qualitative student voice strengthens confidence in the overall conclusions and aligns with broader findings in AI and medical education research.

Conclusion and Implications

This article provides preliminary evidence that AI-powered chatbot assessments may serve as a viable and pedagogically meaningful alternative to traditional exams in fully online Anatomy and Physiology courses. In this study, students perceived AI-based assessments as supporting deeper understanding, clearer explanation, and reduced assessment anxiety while still maintaining rigor and alignment with course objectives. Although the findings are exploratory, they suggest that AI-supported assessment can move students beyond simply identifying the correct answer and toward explaining why an answer makes sense, how body systems connect, and how scientific knowledge can be communicated in patient-centered ways.

One of the most important implications of this work is that assessment format matters. Students in both sections demonstrated growth in confidence, but the AI-supported assessment appeared to shape confidence differently by requiring students to verbalize their reasoning, respond to follow-up prompts, and explain concepts in both technical and accessible language. This matters in A&P because many students enrolled in these courses are preparing for healthcare-related fields where success depends not only on content knowledge but also on communication, clinical reasoning, and the ability to translate complex information for patients. In this sense, AI-powered chatbot assessments may better reflect the kinds of applied thinking students will need beyond the classroom.

At the same time, this study does not suggest that AI-based assessments should fully replace traditional exams. Traditional exams still provide value in measuring foundational knowledge, content recall, and readiness under structured conditions. However, they often fail to capture whether students can explain, apply, and communicate what they know. AI-based assessments may be most useful as a complementary assessment model that expands what counts as evidence of learning. When carefully designed, these assessments can help instructors evaluate dimensions of student understanding that are often difficult to measure through multiple-choice or short-answer exams alone.

The findings also point to the importance of intentional course design. AI assessment cannot simply be added to a course without preparation. Students need clear expectations, practice opportunities, rubrics, and guidance on how to interact with AI in ways that support learning rather than replace it. In this study, the AI chatbot did not provide answers; instead, it prompted students to clarify, elaborate, and explain their reasoning. This design choice is important because it positions AI as an assessment environment rather than a shortcut. For AI-supported assessments to be equitable and effective, instructors must also consider students' varying levels of comfort with technology and provide scaffolding for AI literacy.

For online A&P courses specifically, AI-powered chatbot assessments may help address a persistent challenge: how to create meaningful, rigorous, and authentic assessments in an asynchronous environment. Fully online courses often rely heavily on proctored exams to protect academic integrity, but those exams may not always reflect the depth of student learning. A structured chatbot assessment offers another possibility. It allows students to demonstrate knowledge through explanation while still working within time limits, proctoring expectations, and instructor-designed prompts. This creates a more applied assessment experience without removing academic accountability.

Future research should build on this exploratory study through larger sample sizes, more diverse institutional contexts, and longitudinal designs that examine whether AI-supported assessment influences student performance in later coursework, clinical programs, or professional communication tasks. Additional research should also compare student performance across different assessment types, including traditional exams, oral exams, written case studies, and AI chatbot conversations. Performance-



based outcome measures would strengthen the evidence base and help determine whether students who complete AI-supported assessments demonstrate stronger long-term retention, application, or communication skills.

Overall, this study suggests that AI-powered chatbot assessments have the potential to reframe assessment in Anatomy and Physiology from a process focused primarily on recall to one that values reasoning, explanation, and communication. If students are being prepared for careers where they must think critically, communicate clearly, and apply biological knowledge to real human situations, then assessment practices should reflect those goals. AI-supported assessment, when designed with rigor, transparency, and equity in mind, offers one promising pathway toward that broader vision of learning.

Appendix.

Table 1. Survey Instruments and Analytic Purpose

Survey instrument	Timing	Group	Number of Participants (n)	Main Purpose
Pre-assessment survey	Beginning of course	Combined sections	15	Establish baseline student confidence, anxiety, and expectations before major assessments.
Post-assessment survey	Post-assessment	Combined sections	11	Capture overall reflection on how assessments supported understanding, application, confidence, engagement, and stress.
End-of-course survey: AI section	End of course	AI-supported assessment section	8	Capture student perceptions after completing structured AI chatbot assessments.
End-of-course survey: Traditional section	End of course	Traditional assessment section	4	Capture student perceptions after completing conventional exams.

Table 2. Qualitative Findings by Survey Instrument

Survey instrument	Group	Number of Participants (n)	De-identified quote(s)
Pre-assessment survey	Combined sections	15	<p>“All I’m trying to do when it comes to exams is memorize and then regurgitate the information to get a good grade.”</p> <p>“Explaining concepts in my own words and teaching others has helped me understand the material the most.”</p>



Post-assessment survey	Combined sections	11	<p>"I feel as though I am not just trying to regurgitate memorized information."</p> <p>"I was able to talk more conceptually about the topic instead of just answering."</p> <p>"My motivation has been higher and my stress has lowered."</p>
End-of-course survey: AI section	AI-supported assessment	8	<p>"These exams were very helpful because they allowed me to talk through my thought processes of the material learned."</p> <p>"I was really able to display what I know, which is not something you typically get to do in a multiple-choice standard exam."</p> <p>"The AI could be difficult to steer in the right direction or avoid being redundant."</p>
End-of-course survey: Traditional section	Traditional assessment	4	<p>"I can use process of elimination to get an answer if I don't necessarily know the question."</p> <p>"The short essays have always been a little challenging; sometimes it seems like I can't remember every single thing on the prompt."</p> <p>"The presentation part... helped me recall the information well during the exams."</p>

Table 3. Quantitative Findings by Survey Instrument

Survey instrument	n	Confidence/ understanding	Explanation/ communication	Application	Anxiety/ stress	Assessment fairness
Pre-assessment survey	15	A&P confidence M = 3.87	Explain scientific ideas M = 4.07	Apply learning to real-life situations M = 4.20	Usually anxious before exams M = 4.33	Prepared for assessment format M = 3.73
Post-assessment survey	11	Helped understand A&P M = 4.27	Confident explaining concepts M = 4.09; Express clearly M = 4.36	Apply rather than memorize M = 4.18	Less anxious as course progressed M = 4.00	Fair M = 4.27
End-of-course survey: AI section	8	A&P confidence M = 4.38	Explain scientific ideas M = 4.25	Real-life application M = 4.38	Usually anxious before exams M = 3.50	Comfort with new formats M = 4.38; Prepared M = 4.00
End-of-course survey: Traditional section	4	A&P confidence M = 4.25	Explain scientific ideas M = 4.00	Real-life application M = 4.50	Usually anxious before exams M = 4.25	Prepared M = 4.25



REFERENCES

- [1] Aldossary A. S., Aljindi A. A. and Alamri J. M., "The Role of Generative AI in Education: Perceptions of Saudi Students", *Contemporary Educational Technology*, 2024, 16(4), <https://doi.org/10.30935/cedtech/15496>.
- [2] Brodersen L. and Lorenz R., "Perceived Stress, Physiological Stress Reactivity, and Exit Exam Performance in a Prelicensure Bachelor of Science Nursing Program", *International Journal of Nursing Education Scholarship*, 2020, 17(1), Article 20190121, <https://doi.org/10.1515/ijnes-2019-0121>.
- [3] Brown S. J., White S. and Power N., "Introductory Anatomy and Physiology in an Undergraduate Nursing Curriculum", *Advances in Physiology Education*, 2017, 41(1), 56–61, <https://doi.org/10.1152/advan.00112.2016>.
- [4] Chapagai S. D. and Adhikari B., "Exploring the Role of Artificial Intelligence in Education", *International Research Journal of MMC*, 2024, 5(5), 99–108, <https://doi.org/10.3126/irjmmc.v5i5.73633>.
- [5] Chauhan T., Hasan S., Vardhan G., et al., "Impact of Assessment Formats on Academic Performance and Stress", *Journal of Medical Evidence*, 2024, 5(4), 315–319.
- [6] Darejeh A., Moghadam T. S., Delaramifar M. and Mashayekh S., "AI-Powered Adaptive E-Learning Systems", *ICeLeT Proceedings*, 2024, 1–6.
- [7] Eltahir M. E. and Babiker F. M. E., "The Influence of Artificial Intelligence Tools on Student Performance in e-Learning Environments: Case Study", *Electronic Journal of e-Learning*, 2024, 22(9), 91–110, <https://doi.org/10.34190/ejel.22.9.3639>.
- [8] Gervacio A. P., "Generative AI and Motivated Engagement in Science Learning", *Environment and Social Psychology*, 2024, 9(11).
- [9] Linden K., Pemberton L. A. and Webster L., "Evaluating the Bones of Adaptive Learning: Do the Initial Promises Really Increase Student Engagement and Flexible Learning within First Year Anatomy Subjects?", 5th International Conference on Higher Education Advances (HEAd'19), Valencia, Editorial Universitat Politècnica de València, 2019, 331–339, <https://doi.org/10.4995/HEAd19.2019.9346>.
- [10] Masters K., "Artificial Intelligence in Medical Education", *Medical Teacher*, 2019, 41(9), 976–980, <https://doi.org/10.1080/0142159X.2019.1595557>.
- [11] McPherson M. S. and Bacow L. S., "Online Higher Education: Beyond the Hype Cycle", *Journal of Economic Perspectives*, 2015, 29(4), 135–154.
- [12] Miller G. E., "The Assessment of Clinical Skills/Competence/Performance", *Academic Medicine*, 1990, 65(9), S63–S67.
- [13] National Academies of Sciences, Engineering, and Medicine, "Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students' Diverse Pathways", Washington, DC, The National Academies Press, 2016, <https://doi.org/10.17226/21739>.
- [14] National Center for Education Statistics, "Undergraduate Enrollment", U.S. Department of Education, Institute of Education Sciences, 2023.
- [15] Perrella A., Koenig J., Kwon H., et al., "On Being Examined", *Advances in Physiology Education*, 2015, 39(4), 320–326.
- [16] Piaget J., "The Psychology of Intelligence", London, Routledge & Kegan Paul, 1950.
- [17] Rahiman H. U. and Kodikal R., "Revolutionizing Education: Artificial Intelligence Empowered Learning in Higher Education", *Cogent Education*, 2024, 11(1), Article 2293431, <https://doi.org/10.1080/2331186X.2023.2293431>.
- [18] Sweller J., "Cognitive Load During Problem-Solving", *Cognitive Science*, 1988, 12(2), 257–285.
- [19] Vygotsky L. S., "Mind in Society: The Development of Higher Psychological Processes", Cambridge, MA, Harvard University Press, 1978.
- [20] Zhai X., Chu X., Chai C. S., et al., "A Review of AI in Education from 2010 to 2020", *Complexity*, 2021, Article 8812542, <https://doi.org/10.1155/2021/8812542>.