AI-Powered Chatbot Assessments in Online Anatomy & Physiology

Exploratory Mixed-Methods Pilot Study

Jaime Cantú, M.S., M.B.A.
Associate Professor of Biology
Austin Community College

Are Our Assessments Measuring What Students Actually Need to Do?

What Traditional Exams Do Well

- Factual recall
- Breadth of content coverage
- Foundational knowledge

What May Be Missing

- Applying reasoning to patient scenarios
- Constructing scientific explanations
- Communicating with patients
- Thinking beyond multiple choice

Purpose of the Pilot Study

Explored AI-powered chatbot assessments in an online A&P I course, comparing two sections to surface competencies traditional exams may miss.

Explanation

- Articulate concepts in their own words

Application

- Transfer knowledge to clinical scenarios

Communication

- Translate content into patient-friendly language

Fairness & Anxiety

- Perceived equity and assessment-related stress

Exploratory pilot — findings inform future research, not definitive conclusions.

Online Course Environment

Both sections were fully online, using the same Blackboard LMS infrastructure for a consistent learning environment.

Blackboard LMS

- Course modules and learning objectives
- Assignment instructions and assessment links
- Grades and communication
- Identical structures and deadlines across both sections

Respondus LockDown Browser + Webcam

- Restricted access to unauthorized sites and apps
- Webcam monitoring during all assessments
- Both formats timed and proctored

Study Design: Two Sections, One Difference

Held Constant

- Fully online BIOL 2401
- Same instructor
- Same objectives and pacing
- Same labs and grading
- Same integrity expectations

The One Difference

- **Traditional:** Proctored multiple-choice/short-answer exams
- **AI-Supported:** Proctored AI chatbot assessments

Enrollment

- 24 enrolled per section
- AI section final: 16 students
- Traditional section final: 19 students

What Did the AI Assessment Look Like?

Students led a structured dialogue with a virtual patient, demonstrating mastery through explanation.

The Conversation

- Explained concepts using scientific terminology
- Translated technical language for the patient

The Chatbot's Role

- Asked for clarification and elaboration
- Did not provide answers
- Students had to drive the dialogue

What Students Demonstrated

- Content knowledge and scientific reasoning
- Conceptual explanation
- Patient-centered communication
- Real-world application

Mirrors how healthcare professionals communicate daily: translating complex biology into language patients can act on.

Methodology

Study Design

- Exploratory convergent mixed-methods pilot
- Convenience sample from two course sections
- Voluntary participation, IRB-approved
- Instructor blinded to participant identities until after grading

Data Analysis

- Likert-scale results summarized descriptively
- Open-ended responses coded for themes
- Qualitative findings as primary evidence

Four Survey Instruments

Pre-Assessment n = 15	Post-Assessment n = 11
AI End-of-Course n = 8	Traditional End-of-Course n = 4

FINDING 1

Moving Beyond Memorization

- Verbal explanation requires genuine understanding
- Students organized and externalized their thinking
- Made student thinking visible to evaluators and themselves

> "It pushed me to actually understand the material."

> "I had to explain it in my own words."

> "It moved me past memorize and regurgitate."

FINDING 2

Communication and Applied Reasoning

AI-section students moved beyond knowledge recall toward demonstrating how they think.

> *"I was really able to display what I know."*

> *"It allowed me to talk through my thought processes."*

Applied Reasoning

Concepts linked to real patient scenarios

Language Translation

Technical terms made patient-friendly

Process Visibility

Reasoning made audible and assessable

FINDING 3

A Different Assessment Experience

Assessment design shapes how students feel and perform.

AI Section

- Felt conversational, less pressured
- Easier to explain knowledge fully
- Lower perceived stress

"It felt more like a conversation than a test."

Traditional Section

- Valued predictability and familiarity
- Leveraged test-taking strategies
- Process of elimination still useful

"I can use process of elimination if I don't know the question."

Neither format is superior. Each reveals different aspects of student knowledge. What matters is alignment between format and intended outcomes.

FINDING 4

The AI Assessment Was Not Automatically Easier

The chatbot format introduced its own cognitive and technical demands alongside content mastery.

New Format, New Challenges

- Chatbot could be repetitive or hard to redirect
- Steering conversations required its own skill set
- AI familiarity varied across students
- Technology comfort raises equity concerns

"The AI could be difficult to steer in the right direction."

What Students Need to Succeed

- Practice assessments before it counts
- Clear instructions and transparent rubrics
- AI-literacy scaffolding for productive interaction
- Technical support to reduce access barriers

Limitations

An exploratory pilot designed to generate hypotheses, not causal conclusions.

Sample & Design

- Small, single-course convenience sample
- No random assignment
- Possible student self-selection

Data & Measurement

- Unequal survey response counts
- Responses not matched over time
- Relies on self-reported perceptions

Analysis & Scope

- No long-term learning outcomes measured
- Single-researcher qualitative analysis
- Limited generalizability

⚠ These limitations define the boundaries for interpretation and set the agenda for future, more rigorous research.

Implications for the Future of Assessment

Expanding what counts as evidence of learning — not replacing traditional exams.

Traditional Exams Remain Valuable	AI May Fill a Gap	Balance Is the Goal
• Essential for foundational knowledge and recall • Efficiently assess breadth at scale	• Surfaces reasoning, explanation, and application • Competencies central to clinical practice	• Mixed assessment system • Measure both knowledge and application

Effective implementation requires rigor, transparency, equity, and integrity.

Closing Takeaway

Assessment design signals what we value — and shapes what students prepare to demonstrate.

Healthcare Students Must

- Grasp complex biology deeply
- Apply knowledge to real situations
- Explain reasoning clearly
- Communicate with patients accessibly

Our assessments should reflect these expectations.

The Path Forward

- AI-supported assessments mirror real healthcare demands
- A **complement** to existing tools, not a replacement
- Larger, rigorous studies still needed before broad adoption

This pilot opens an important conversation about what we ask students to demonstrate.

Thank You

Questions and reflections welcome.

Study Contact

Jaime Cantú, M.S., M.B.A.
Associate Professor of Biology
Austin Community College

Jaime.Cantu@austincc.edu

Key Takeaway

When we expand what counts as evidence of learning, we expand what students are prepared to do.