Using Generative AI to Simulate Patient History-Taking in a Problem-Based Learning Tutorial: A Mixed-Methods Study

Allison Mool, Jacob Schmid, Thomas Johnston, William Thomas, Emma Fenner, Kevin Lu, Raya Gandhi, Adam Western, Brendan Seabold, Kodi Smith, Zachary Patterson, Hannah Feldt, Daniel Vollmer, Roshan Nallaveettil, Anthony Fanelli, Logan Schmillen, Shelley Tischkau, PhD, Anna T. Cianciolo, PhD, Pinckney Benedict, MFA, Richard Selinfreund, PhD

Allison Mool is a 2nd year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0006-5235-4646. (CRediT taxonomy roles: investigation, methodology, writing original draft)

Jacob Schmid is a 2nd year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0000-8930-3718. (CRediT taxonomy roles: conceptualization, investigation, methodology)

Thomas Johnston is a 4th year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0007-5319-9307. (CRediT taxonomy roles: methodology, writing original draft)

William Thomas is an undergraduate student in anthropology, Southern Illinois University, Carbondale, Illinois; ORCID: 0009-0000-8372-4183. (CRediT taxonomy roles: formal analysis, investigation, methodology)

Emma Fenner is a 4th year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0000-0002-9066-3507. (CRediT taxonomy roles: conceptualization)

Kevin Lu is a 2nd year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0000-0002-8459-5480. (CRediT taxonomy roles: investigation, resources)

Raya Gandhi is a 2nd year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0009-1601-3838. (CRediT taxonomy roles: formal analysis, investigation, resources)

Adam Western is a 3rd year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0009-5953-967X. (CRediT taxonomy roles: conceptualization, resources)

Brendan Seabold is a 3rd year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0008-4494-9863. (CRediT taxonomy roles: resources)
Kodi Smith is a 1st year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0004-5921-9004. (CRediT taxonomy roles: formal analysis, methodology, software, writing original draft)

Zachary Patterson is a 1st year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0008-7566-1088. (CRediT taxonomy roles: formal analysis, methodology, software, writing original draft)

Hannah Feldt is a 1st year medical student, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0009-0006-6189-4767. (CRediT taxonomy roles: formal analysis, methodology, software, writing original draft)

Daniel Vollmer is an undergraduate student in computer science, Southern Illinois University, Carbondale, Illinois; ORCID: 0009-0005-8874-4817. (CRediT taxonomy roles: software)

Roshan Nallaveettill is an undergraduate student in computer science, Southern Illinois University, Carbondale, Illinois; ORCID: 0009-0008-7033-5814. (CRediT taxonomy roles: software)

Anthony Fanelli, is an undergraduate student in computer science, Southern Illinois University, Carbondale, Illinois; ORCID: 0009-0007-6817-9827 (CRediT taxonomy roles: software)

Logan Schmillen is an undergraduate student in computer science, Southern Illinois University, Carbondale, Illinois; ORCID: 0009-0007-7287-1867 (CRediT taxonomy roles: software)

Shelley Tischkau is professor and Chair, Department of Pharmacology, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0000-0002-8016-6398. (CRediT taxonomy roles: conceptualization, investigation)

Anna T. Cianciolo is professor, Department of Medical Education, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0000-0001-5948-9304. (CRediT taxonomy roles: conceptualization, formal analysis, writing original draft)

Pinckney Benedict is professor, Department of English, and Director, Center for Virtual Expression (CVEX), Southern Illinois University, Carbondale, Illinois; ORCID: 0009-0001-5942-6968. (CRediT taxonomy roles: conceptualization, methodology, resources, visualization)

Richard Selinfreund is associate professor and Head of Pathology Group, Department of Medical Microbiology, Immunology and Cell Biology, Southern Illinois University School of Medicine, Springfield, Illinois; ORCID: 0000-0002-3778-5134. (CRediT taxonomy roles: funding acquisition, methodology, project administration, resources, supervision)
Correspondence should be addressed to Richard Selinfreund, Department of Medical Microbiology, Immunology and Cell Biology, Southern Illinois University School of Medicine, PO Box 19626, Springfield, IL 62794-9626; telephone: (217) 545-4580; email: rselinfreund39@siumed.edu.
Abstract

**Background:** Medical educators who implement problem-based learning (PBL) strive to balance realism and feasibility when simulating patient cases, aiming to stimulate collaborative group discussion, engage students’ clinical reasoning, motivate self-directed learning, and promote the development of actionable scientific understanding. Recent advances in generative artificial intelligence (AI) offer exciting new potential for patient simulation in PBL. **Method:** This study used a between-groups, mixed-methods approach to (1) form a comprehensive picture of Year 2 medical student interactions with a generative AI-simulated patient in a PBL tutorial, as compared to interactions with multimedia patient case materials; and (2) triangulate on the impact these interactions had on learning. Two groups of students (N = 13) gathered patient history information from a generative AI-enabled, 3D-animated avatar (AI condition). Two other student groups (N = 13) gathered patient history information from a multimedia database using keyword searching (Electronic PBL Module [ePBLM] condition). We used descriptive observation to explore student interactions with both forms of the simulated patient, and we quantitatively compared students’ perceptions of their learning experience and recall of patient history information across conditions. **Results:** Students in the AI condition rated their present, AI-augmented PBL learning experience—particularly its clinical accuracy and teamwork aspects—significantly higher than they rated their previous PBL learning experiences using ePBLMs. Recall of patient history information did not differ between conditions. Descriptive observation indicated that the AI avatar presented case content accurately, with an appropriate amount of information provided in response to students’ questions. Students were highly engaged as a group in taking a history from the avatar. However, although students used language suggestive of anthropomorphizing of the AI (e.g., gender pronouns), they appeared to orient to it as an augmented “question bank” for gathering patient history information, using a questioning strategy akin to querying an ePBLM. **Discussion:** Optimizing AI implementation to stimulate clinical reasoning and patient communication skills in PBL could include (1) starting early, perhaps in Year 1, before an alternative interactional framework can take hold; (2) orienting students to the AI to help them understand its capabilities; and (3) encouraging “play” with or “discovery learning” of the AI’s capabilities.
Introduction

In medical problem-based learning (PBL), patient cases provide the organizing structure of the curriculum, serving as triggers for learning and providing a platform to exercise thinking representative of physicians’ workplace cognition (Wee, Kek, & Sim, 2001).

Patient case design therefore is a key concern for PBL educators. Quality patient cases stimulate collaborative group discussion, engage clinical reasoning, motivate self-directed learning, and promote the development of actionable scientific understanding.

- Characteristics of a quality patient case – realistic, ideally based on a real patient, set in a medical context, sufficiently challenging for students to stimulate high-level cognitive engagement, designed based on clear learning objectives (Wee, Kek, & Sim, 2001; Azer, Peterson, Guerrero, & Edgren, 2012)
- Characteristics of quality interactions with patient case materials (questioning oneself and others, self-explanation and elaboration of ideas, testing one’s understanding, and self-identification of knowledge and comprehension gaps – Cianciolo & Regehr, 2019)

The diversity of approaches to simulating patients in PBL ranges from paper cases to interactive courseware to live patient actors, each with their own tradeoffs.

- Advantages of high-fidelity simulation (engaging and provides opportunity to practice clinical skills – Barrows & Tamblyn, 1980; facilitates patient-centered learning – MacLeod, 2011)
- Disadvantages of high-fidelity simulation (lack of control and potential to become too complex – Barrows & Tamblyn, 1980; cost in time, personnel, and funding)

Recent advances in generative artificial intelligence (AI) offer exciting new potential for patient simulation in PBL.

- Capabilities of generative AI for simulating people realistically – use of these capabilities in gaming and for other non-medical purposes
- Ways that generative AI provides the advantages of high-fidelity simulation without the disadvantages (guardrails)
- More realistic patients may stimulate renewed interest in PBL tutorials at a time when engagement in our own curriculum and elsewhere is lacking (Dawood, Rea, Decker, Kelley, & Cianciolo, 2021; Walling, Ista, Bonaminio, Paolo, Fontes, Davis, & Berardo, 2017; Kilgour, Grundy, & Monrouxe, 2016).

Generative AI-simulated patients offer higher fidelity patient simulation, with potentially greater control, than what is currently feasible in most medical PBL curricula. However, AI development can be time-consuming and expensive, and the technology is still evolving. Research is needed to explore the impact of using generative AI to simulate patients in PBL tutorials to understand how (and when) to best implement this new technology.
Methods

Overview

This study used a between-groups, mixed-methods approach (Schifferdecker & Reed, 2009) to (1) form a comprehensive picture of medical student interactions with a generative AI-simulated patient in a PBL tutorial, as compared to interactions with multimedia patient case materials; and (2) triangulate on the impact these interactions had on learning. Two groups of students gathered patient history information from a generative AI-enabled, 3D-animated avatar (AI condition). Two other student groups gathered patient history information from a multimedia database using keyword searching (Electronic PBL Module [ePBLM] condition). We used descriptive observation to explore student interactions with both forms of the simulated patient, and we quantitatively compared students’ perceptions of their learning experience and recall of patient history information across conditions.

Context/Setting

This study took place in March 2024 at a small, public, community-based medical school in the Midwestern U.S. (class size = 72), which was the first institution to implement its PBL curriculum using ePBLMs (Ryan & Koschmann, 1994). The 4-year curriculum follows a 2+2 design in which the first 2 years focus on preclinical, basic science instruction in a classroom setting with periodic clinical exposure, and the second 2 years focus on clinical clerkships and electives. The preclinical curriculum comprises spirally organized organ-system-based units; Year 1 instruction emphasizes anatomy, physiology, and histology, and Year 2 emphasizes pathophysiology and pharmacology. The curriculum is delivered using a hybrid PBL format (Lim, 2012) in which participation in PBL tutorials is required, and attendance at correlated lectures is recommended, but not mandatory.

PBL tutorials in our preclinical curriculum comprise small groups of 6-8 students and a facilitator who meet multiple times per week (3 times/week in Year 1, 2 times/week in Year 2) to exercise clinical reasoning with real-world patient cases, discuss understanding of the case content, identify basic science knowledge gaps (i.e., learning issues), and develop shared knowledge of the patient’s condition and its treatment (Koschmann, Glen, & Conlee, 1997; Cianciolo, Kidd, & Murray, 2016). Procedurally, students work collaboratively to “open” the case, which involves gathering patient information from history-taking, physical examination, laboratory studies, and imaging to understand the patient’s medical concerns, form a differential diagnosis, and identify learning issues to address with self-directed learning. Following self-directed learning, students “close” the case by discussing their self-directed learning findings, the implications of their new knowledge for understanding the patient case, and the patient’s care outcomes. The students comprising PBL tutorial groups change each unit, for a total of 94 unique groups assigned across both years of the preclinical curriculum.
Participants

Participants comprised 4 groups of second-year medical students (N = 26, 37% of the Class of 2026. Among the 14 participants who provided demographics, 79% self-identified as women and 20% self-identified as belonging to a marginalized racial/ethnic group. Students’ participation occurred during the first 2 days of the final unit in the preclinical curriculum (Endocrinology, Reproduction, and Gastroenterology [ERG]), approximately 3 months prior to taking Step 1 of the USMLE Board Examination, and after nearly 2 years’ experience in our PBL (ePBLM) curriculum. Participants were a convenience sample recruited via word of mouth by fellow second-year medical students serving as co-investigators in this study. There were no inclusion/exclusion criteria. Participants were allowed to choose their PBL tutorial group with knowledge of who else was in their group. Groups were then assigned to the AI or ePBLM condition depending on investigator availability and preference.

This study was deemed non-human subjects research by the Springfield Committee on Research Involving Human Subjects (6 February 2024). The facilitator for all 4 groups, a faculty member from the investigator team—who is locally considered a master PBL tutor with 16 years of tutoring experience—described the study to participants and received their verbal consent to participate with audio-recording. Participants were offered pizza and drinks during their tutorial, but no other incentives were provided.

Materials

Patient Case – Randy Rhodes. The patient case used in this study is a 54-year-old man—Randy Rhodes—who presents to the family medicine clinic with a rash on his arm. His diagnosis is Diabetes Mellitus Type 2. The case is designed to teach pathology, pharmacology, and other topics associated with this diagnosis (e.g., obesity and related metabolic syndromes) as one of 20 cases comprising the ERG PBL curriculum. Participants interacted with Randy Rhodes to gather patient history information in one of 2 ways, ePBLM or generative AI:

ePBLM. ePBLMs have been used to simulate patients in our PBL curriculum for over 30 years. Developed locally to replace the spiral-bound notebooks of comprehensive case information provided to students to role-play the patient, ePBLMs allow students to gather case information serially by using keyword searches to simulate history-taking, physical examination, testing/imaging, and procedures (Ryan & Koschmann, 1994). The ePBLM system returns answers to students’ search queries as realistically as possible using multimedia. For instance, it provides actual laboratory test or imaging results where possible, and it returns canned dialogue snippets in response to patient history queries (e.g., “I’ve had this rash for 3 months.”) as well as offering information about patient progression throughout their treatment. Figure 1 presents a screen capture of the ePBLM interface for the Randy Rhodes case.

Generative AI. Students gather patient history information from the AI-enabled patient via voice-to-voice interaction with a 3D-animated avatar situated in a clinical exam room.
environment created using Unreal Game Engine 5.0 and Metahuman (Figure 2). The content of the avatar’s responses to students’ history-taking questions—generated using the convai platform—is produced via prompts to a general large language model that have been engineered using the existing Randy Rhodes ePBLM database. The scope of the avatar’s responses is shaped by “guardrails” placed by the development team to optimize the case’s educational value. For instance, one guardrail prevents Randy Rhodes from revealing his diagnosis directly to students. Situated within a game engine, the AI also is capable of responding with different degrees of friendliness depending on how students interact with it. For example, the AI is trained to respond irritably to history questions laden with medical jargon and, if guardrails had not been implemented to prevent it, could storm out of the exam room.\(^1\) Interactions with the AI in this context may take forms other than questions; the avatar can respond to empathic statements and even tell jokes.

**Learner Perceptions Survey.** This paper-based 8-item survey assessed participants’ perceptions of PBL tutorials using either ePBLM or AI. Participants used a scale of 1 to 5 to rate how positively they view several aspects of tutorial quality and impact (impact on improving clinical knowledge, clinical simulation accuracy, impact on clinical knowledge retention, enjoyability, stressfulness, allowance for teamwork, interestingness, impact on confidence) when either format is used. Both forms of the survey (ePBLM and AI) are presented in the supplemental appendix. A composite score for each form was created by summing the ratings across all 8 items (Item 5 was reverse coded). The possible range of scores was 8 to 40, representing how positively participants view their overall PBL tutorial experience using each format.

**Patient History Information Quiz.** This paper-based quiz (supplemental appendix) consisted of 11 short-answer questions to assess participants’ knowledge of Randy Rhodes’s demographic information, medical history, and symptoms immediately following the PBL tutorial and 2 weeks later. Quizzes were graded by 2 student co-investigators masked to the condition (ePBLM or AI) and timing (immediate or delayed) when the quiz was completed. According to a faculty-verified key, each short answer could receive 0 points, ½ point, or 1 point, and the possible range of total scores was 0 to 11, representing overall recall of patient history information available in the Randy Rhodes case. Both co-investigators independently graded the first 5 quizzes to establish consistency with the rating criteria, after which 1 co-investigator graded the remaining 52 quizzes.

**Descriptive Observation Notes.** Five co-investigators conducted descriptive observation (Spradley, 1980) of the audio-recorded tutorials, documenting their notes using a template created locally by a well-established medical education researcher (supplemental appendix). The template prompted observers to (1) document a reflexivity reflection; (2) specify (with time stamp) and describe noteworthy occurrences observed in the audio-recording; and (3) report in bracketed fashion their reactions to these noteworthy occurrences. The lead observer was a

\(^1\)Randy’s developers removed the door from the examination room such that storming out is no longer possible. Instead, the avatar will disengage from a problematic interaction and steer the conversation back toward the rash on his arm, which is his pressing concern and motivation for seeing a doctor.
behavioral scientist with experience conducting observational research of PBL tutorials (Cianciolo, Kidd, & Murray, 2016). The other observers were first and second-year medical student co-investigators enrolled in our preclinical curriculum.

Procedure

All 4 PBL tutorials took place in a small laboratory room arranged to mimic a typical PBL tutorial classroom (Figure 3). The room was equipped with tables, chairs, a whiteboard, and the tools needed to interact with the patient in each group: a keyboard and mouse to search the ePBLM database and a push-to-talk microphone to communicate verbally with the AI-enabled patient. The room also contained a computer monitor for presenting the simulated patient visually.

After consent and before beginning their PBL tutorial, participants filled out the Learner Perceptions Survey – ePLBM. That is, participants rated their perceptions of ePBLM tutorials generally, based on their nearly 2 years’ experience in our PBL curriculum. Survey administration was proctored by 1 of 4 student co-investigators who attended the tutorials for this purpose as well as to answer participant questions, operate the push-to-talk microphone (AI groups), and manage the tutorial’s other technology, such as audio-recording. Audio-recording was accomplished via videoconferencing software (WebEx).

With minimal direction from the faculty facilitator, all 4 groups conducted a case opening, which ranged in duration from 36 to 75 minutes and consisted of history taking. After completing the case opening and before leaving the laboratory room, participants retook the Learner Perceptions Survey, this time rating their perceptions of PBL tutorials using the type of simulated patient they had just interacted with (ePBLM or AI). Participants also completed the Patient History Information Quiz.

Two weeks after the tutorials, participants individually retook the Patient History Information Quiz, proctored by a student co-investigator, in various campus locations of convenience for each participant. Student co-investigators graded the quizzes after all quizzes had been completed. Descriptive observation of the audio-recordings began after all 4 tutorials had occurred.

Data Analysis

Quantitative analysis. Quantitative analyses were conducted using linear mixed models to control for within-group variance while examining the effect of condition (ePBLM vs. AI) on participants’ perceptions and recall of patient history information. P-values of .05 or smaller were deemed statistically significant. All analyses were performed using R version 4.3.2 (R Core Team 2023).
Qualitative analysis. Observation notes were analyzed inductively to characterize the interactions among students and the simulated patient in each group. For each of the 4 groups, the lead analyst compiled the independent observers’ notes into one set. These compiled notes were arranged chronologically to create a detailed, multi-perspective characterization of noteworthy events and associated bracketed reflections that occurred during each tutorial. To highlight the different perspectives at a glance, the lead analyst represented each observers’ notes using a different text color. Next, all qualitative analysts independently reviewed each set of compiled field notes to identify regularities in the interactions among people and case materials. Flagged as regularities were recurring interactional behaviors such as consistencies in patient questioning strategies or commonalities among occurrences of pauses or silence in group interactions. Analysis also flagged interactional behaviors potentially reflective of how students oriented to the patient case or each other (e.g., referring to the patient with personal pronouns or seeking input from peers in asking history-taking questions). Finally, all qualitative analysts collaboratively reviewed regularities and noteworthy interactional behaviors within and across participant groups to develop patterns in PBL behavior associated with type of simulated patient.

Results

Learner Perceptions Survey Data

Linear mixed modeling of the Learner Perceptions Survey data revealed a significant effect of time (pre- vs. post-tutorial) for the AI condition only (Table 1). Specifically, participants’ post-tutorial ratings of their overall PBL learning experience were, on average, 6 points higher than their pre-tutorial ratings if their tutorial involved an AI-enabled patient. In other words, students in the AI condition rated their present, AI-augmented PBL learning experience significantly higher overall than they rated their previous PBL learning experiences using ePBLMs. Students in the ePBLM condition perceived their present PBL learning experience to be approximately the same as their previous PBL learning experiences using the same technology.

[INSERT TABLE 1 HERE.]

The item-level data in Table 2 indicate that students’ post-tutorial perceptions of their learning experience in the AI condition were higher than their pre-tutorial perceptions almost across the board. Particularly noteworthy is the difference in perceptions of clinical simulation accuracy; on average, students who gathered patient history information from an AI-enabled patient rated the clinical accuracy of their present PBL tutorial 1.6 points higher (on a 5-point scale) than the clinical accuracy of their previous PBL tutorials using ePBLMs.

[INSERT TABLE 2 HERE.]
**Patient History Quiz Data**

Linear mixed modeling of the Patient History Quiz data revealed a significant effect of time only (Table 3). Specifically, students’ recall of Randy Rhodes’s patient history information decreased significantly from immediate to delayed assessment, across both conditions (ePBLM and AI). Immediate and delayed recall of patient history information slightly favored the AI condition (10.1 vs. 9.4 immediate recall for the AI and ePBLM conditions, respectively; 8.6 vs. 7.9 delayed recall for the AI and ePBLM conditions, respectively), however this difference was nonsignificant, and the rate of forgetting did not differ between the 2 conditions. Of note, immediate recall performance was near ceiling (maximum quiz score = 11 points) in both conditions, and delayed recall was also very high.

[INSERT TABLE 3 HERE.]

**Observation Data**

Initial impressions from the observation of one group in the AI condition:

- The AI presented case content accurately, with an appropriate amount of information provided in response to students’ questions. The students appeared to find the avatar’s appearance, voice quality, and behavior plausible – with no language to suggest an aversive reaction to the patient representation (i.e., being in the uncanny valley).

- Although students used language suggestive of anthropomorphizing of the AI (e.g., gender pronouns, inferences about the AI’s emotional state), they appeared to orient to it as an augmented “question bank” for gathering patient history information. The format and style of their questions was consistent with how patient history information is gathered from the ePBLM (i.e., singularly focused on diagnosis) versus a more naturalistic conversation with the AI (i.e., diagnostically oriented, yet inclusive of rapport-building). Potential explanations for this include:
  - As rising third-years, the students were very experienced with using ePBLMs, and likely brought this framework for conducting PBL to the study.
    - This frame may have been reinforced by using a patient case that is a required part of the curriculum versus an optional, experimental case.
    - Further, the faculty facilitator did not offer participants a divergent perspective for how the AI might be used.
  - Voice communication with the AI was mediated by a student co-investigator, who operated the push-to-talk microphone at students’ request. The additional layer between the students and the AI may have created a communication bottleneck that reduced the likelihood of naturalistic interactions.
- Students were very engaged, as a group, in history taking with the AI-enabled Randy Rhodes.
Students seemed to enjoy the novelty of the AI avatar, commenting on its appearance and reactions to what they said. The whole group seemed engaged in troubleshooting the AI’s unexpected behaviors (i.e., irritability).

To communicate effectively with the AI, students had to frame their questions naturally instead of in terms of the keywords they were used to using. The whole group seemed engaged in troubleshooting what questions to ask.

Discussion

Although using a generative AI-enabled patient simulation in a PBL tutorial had limited impact on recall of patient history information—which was near ceiling in both the ePBLM and AI conditions—students in the AI condition reported that the tutorial was more enjoyable and facilitative of teamwork than did students in the ePBLM condition. Descriptive observation of participants’ audio-recorded interactions in the PBL tutorials reinforced students’ self-reports and help to characterize what was enjoyable and engaging about taking a history from an artificially intelligent patient. Observation also provides insight into how to maximize the impact of generative AI-enabled patient simulations for clinical learning in the PBL context.

Results in the context of the PBL literature WRT patient simulation.

Implications of results for optimizing AI implementation in PBL.

- Start implementation of AI early, perhaps in the first year, before an alternative framework can take hold.
- Provide students with an orientation to the AI agent to help them understand its capabilities (perhaps during the Year 1 orientation to PBL).
- Encourage “play” with or “discovery learning” of the AI’s capabilities.

Study limitations

- One case at one medical school (but more than one-third of the class)
- Study involved students very experienced with PBL and in the final unit of preclinical instruction preceding the Step 1 board examination
- Students were not randomly allocated to groups (although groups were randomly allocated to condition, and there did not appear to be systematic differences across condition in pre-existing social connections)
- Lack of video data to conduct more comprehensive observations
- Lack of a human-simulated patient to offer a comparison case (e.g., it is impossible to know whether students would orient to a human role-player as an augmented question bank)
Future applications of AI in PBL and elsewhere in the medical curriculum

- More diverse and varied patient interactions than traditional standardized patients while saving on time, scheduling, and training of standardized patients
- Potential for simulating telehealth visits and conducting standardized patient examinations
- Communication skills assessment (e.g., use of accessible language) through the adaptability of the AI to provide slightly different responses based on the phrasing of users' questions
- Scalability of AI across different stages of medical education, from preclinical students learning to communicate and diagnose to clinical students learning treatment and management to residents and fellows learning to navigate complex or uncommon patient interactions

Acknowledgements: We would like to thank Don Torry, PhD, for his guidance and continued focus related to research strategy; Debra L. Klamen, MD, MHPE, for her continued support for this project; and Sean McGinity BS, BA for his support to the quantitative data analysis. Finally, we would like to acknowledge the medical students from the Class of 2026 who participated in this study.

Funding/Support: This project received funding from 3 internal sources at Southern Illinois University School of Medicine (SIU-SOM): Dean’s Start-up Funds; SIU-SOM Concept Development Award; and an SIU-SOM Foundation Grant for Diabetes Research.

Previous presentations: Earlier versions of this work were presented at the 2024 Central Group on Educational Affairs Spring Meeting, Milwaukee, Wisconsin, April 3-5, 2024 and at the SIU-SOM 14th Annual Teaching and Learning Symposium, Springfield, IL, April 11, 2024.

Ethical approval: This project was deemed non-human subjects research by the Springfield Committee on Research Involving Human Subjects (6 February 2024).

References


Figure 1. Screen Capture of the Randy Rhodes ePBLM

**Case:** Randy Rhodes  
**Age:** 54  
**Gender:** Male

<table>
<thead>
<tr>
<th>Questions</th>
<th>Exams</th>
<th>Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Questions</td>
<td>Search Exams</td>
<td>Search Tests</td>
</tr>
<tr>
<td>View all Exams</td>
<td></td>
<td>View all Tests</td>
</tr>
</tbody>
</table>

**Setting and time of day**

Family Medicine Clinic, 9:00 AM

**Patient presentation and initially available information**

Presents with a rash.
Figure 2. Screen Capture of Randy Rhodes, the Generative AI-enabled Patient
Figure 3. Photo of the Laboratory Room in Which the PBL Tutorials Took Place
Table 1. Linear Mixed Model of Learner Perceptions Survey Data

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates std. Error</th>
<th>CI</th>
<th>Statistic</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>23.77</td>
<td>1.13</td>
<td>16.94 – 30.59</td>
<td>21.02</td>
</tr>
<tr>
<td>Time: post</td>
<td>0.23</td>
<td>0.98</td>
<td>-1.80 – 2.26</td>
<td>0.23</td>
</tr>
<tr>
<td>Condition: AI</td>
<td>-1.38</td>
<td>1.60</td>
<td>-6.49 – 3.72</td>
<td>-0.87</td>
</tr>
<tr>
<td>timepost:conditionAI</td>
<td>5.77</td>
<td>1.39</td>
<td>2.90 – 8.64</td>
<td>4.14</td>
</tr>
</tbody>
</table>

Random Effects

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td>6.30</td>
</tr>
<tr>
<td>$\tau_{00}$ id</td>
<td>10.00</td>
</tr>
<tr>
<td>$\tau_{00}$ group</td>
<td>0.00</td>
</tr>
<tr>
<td>$\tau_{11}$ group:conditionAI</td>
<td>0.00</td>
</tr>
<tr>
<td>$\rho_{01}$ group</td>
<td>-0.98</td>
</tr>
<tr>
<td>ICC</td>
<td>0.61</td>
</tr>
<tr>
<td>N group</td>
<td>4</td>
</tr>
<tr>
<td>N id</td>
<td>26</td>
</tr>
</tbody>
</table>

Observations 52

Marginal $R^2$ / Conditional $R^2$ 0.241 / 0.707
<table>
<thead>
<tr>
<th>Survey Item</th>
<th>Pre-Tutorial (N = 26)</th>
<th>Post-Tutorial (N = 26)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AI (N = 13)</td>
<td>ePBLM (N = 13)</td>
<td>AI (N = 13)</td>
<td>ePBLM (N = 13)</td>
</tr>
<tr>
<td>Improves clinical knowledge</td>
<td>3.8 (0.6)</td>
<td>3.8 (0.9)</td>
<td>4.2 (0.6)</td>
<td>3.6 (0.7)</td>
</tr>
<tr>
<td>Simulates clinical experiences accurately</td>
<td>2.6 (1.0)</td>
<td>3.1 (1.0)</td>
<td>4.2 (0.6)</td>
<td>3.5 (0.9)</td>
</tr>
<tr>
<td>Helps clinical knowledge retention</td>
<td>3.6 (1.0)</td>
<td>3.6 (1.0)</td>
<td>4.3 (0.6)</td>
<td>3.5 (0.9)</td>
</tr>
<tr>
<td>Is enjoyable</td>
<td>3.5 (0.8)</td>
<td>3.7 (0.6)</td>
<td>4.7 (0.5)</td>
<td>3.8 (0.8)</td>
</tr>
<tr>
<td>Is stressful*</td>
<td>4.1 (0.5)</td>
<td>3.9 (0.5)</td>
<td>3.8 (0.8)</td>
<td>4.2 (0.6)</td>
</tr>
<tr>
<td>Allows for teamwork</td>
<td>3.8 (0.4)</td>
<td>4.0 (0.4)</td>
<td>4.5 (0.7)</td>
<td>4.0 (0.6)</td>
</tr>
<tr>
<td>Stimulates interest in learning the disease at hand</td>
<td>3.4 (0.9)</td>
<td>3.8 (0.7)</td>
<td>4.4 (0.5)</td>
<td>3.8 (0.8)</td>
</tr>
<tr>
<td>Improves confidence interacting with this chief complaint</td>
<td>3.5 (1.0)</td>
<td>3.8 (0.7)</td>
<td>4.2 (0.8)</td>
<td>3.5 (0.7)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>28.4 (0.4)</td>
<td>29.8 (0.3)</td>
<td>34.4 (0.3)</td>
<td>30.0 (0.3)</td>
</tr>
</tbody>
</table>

*This item is reverse-coded such that a higher score represents lower perceived stress.*
Table 3. Linear Mixed Model of Patient History Quiz Data

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>std. Error</th>
<th>CI</th>
<th>Statistic</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>9.40</td>
<td>0.32</td>
<td>7.55 – 11.26</td>
<td>29.32</td>
<td><strong>0.004</strong></td>
</tr>
<tr>
<td>Time: post</td>
<td>-1.46</td>
<td>0.29</td>
<td>-2.05 – -0.87</td>
<td>-5.07</td>
<td><strong>&lt;0.001</strong></td>
</tr>
<tr>
<td>Condition: AI</td>
<td>0.69</td>
<td>0.41</td>
<td>-1.09 – 2.48</td>
<td>1.71</td>
<td>0.233</td>
</tr>
</tbody>
</table>

**Random Effects**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>σ²</td>
<td>1.08</td>
</tr>
<tr>
<td>τ₀₀ id</td>
<td>0.50</td>
</tr>
<tr>
<td>τ₀₀ group</td>
<td>0.00</td>
</tr>
<tr>
<td>τ₁₁ group.conditionAI</td>
<td>0.00</td>
</tr>
<tr>
<td>ρ₀₁ group</td>
<td>-1.00</td>
</tr>
<tr>
<td>N group</td>
<td>4</td>
</tr>
<tr>
<td>N id</td>
<td>26</td>
</tr>
</tbody>
</table>

Observations 52
Marginal R² / Conditional R² 0.382 / NA
Supplemental Appendix

Learner Perceptions Survey (ePBLM Condition)

Pre-Tutorial

Based on your past experience with ePBLM problem-based learning sessions, please give your opinions on each of these statements:

1. I think that ePBLM problem-based learning sessions improve my clinical knowledge.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

2. I think that ePBLM problem-based learning sessions simulate clinical experiences accurately.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

3. I think that seeing patients in ePBLM problem-based learning sessions helps me retain knowledge about the clinical situations.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

4. I think that ePBLM problem-based learning sessions are enjoyable.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

5. I think that ePBLM problem-based learning sessions are stressful.
   1) Strongly disagree
2) Disagree
3) Neither agree nor disagree
4) Agree
5) Strongly agree

6. I think that ePBLM problem-based learning sessions allow for teamwork.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

7. I think that ePBLM problem-based learning sessions stimulate my interest in learning about the disease at hand.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

8. I usually feel more confident interacting with chief complaints after ePBLM tutor group sessions.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree
Post-Tutorial

1. I think that this ePBLM problem-based learning session improved my clinical knowledge.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

2. I think that this ePBLM problem-based learning session simulated clinical experiences accurately.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

3. I think that seeing the patient in this ePBLM problem-based learning session will help me retain knowledge about the clinical situation.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

4. I think that this ePBLM problem-based learning session was enjoyable.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

5. I think that this ePBLM problem-based learning session was stressful.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

6. I think that this ePBLM problem-based learning session allowed for teamwork.
1) Strongly disagree
2) Disagree
3) Neither agree nor disagree
4) Agree
5) Strongly agree

7. I think that this ePBLM problem-based learning session stimulated my interest in learning about the disease at hand.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

8. I feel more confident interacting with this complaint after this ePBLM tutor group session.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree
Learner Perceptions Survey (AI Condition)

Pre-Tutorial

Based on your past experience with ePBLM problem-based learning sessions, please give your opinions on each of these statements:

1. I think that ePBLM problem-based learning sessions improve my clinical knowledge.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

2. I think that ePBLM problem-based learning sessions simulate clinical experiences accurately.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

3. I think that seeing patients in ePBLM problem-based learning sessions helps me retain knowledge about the clinical situations.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

4. I think that ePBLM problem-based learning sessions are enjoyable.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

5. I think that ePBLM problem-based learning sessions are stressful.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
4) Agree
5) Strongly agree

6. I think that ePBLM problem-based learning sessions allow for teamwork.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

7. I think that ePBLM problem-based learning sessions stimulate my interest in learning about the disease at hand.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

8. I usually feel more confident interacting with chief complaints after ePBLM tutor group sessions.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree
Post-Tutorial

1. I think that this autonomous AI problem-based learning session improved my clinical knowledge.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

2. I think that this autonomous AI problem-based learning session simulated clinical experiencing accurately.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

3. I think that seeing the patient in this autonomous AI problem-based learning session will help me retain knowledge about the clinical situation.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

4. I think that this autonomous AI problem-based learning session was enjoyable.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

5. I think that this autonomous AI problem-based learning session was stressful.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree
6. I think that this autonomous AI problem-based learning session allowed for teamwork.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

7. I think that this autonomous AI problem-based learning session stimulated my interest in learning about the disease at hand.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree

8. I feel more confident interacting with this complaint after this autonomous AI tutor group session.
   1) Strongly disagree
   2) Disagree
   3) Neither agree nor disagree
   4) Agree
   5) Strongly agree
Patient History Information Quiz

1. What was Randy’s chief complaint?

2. Did anything seem to help relieve Randy’s chief complaint?

3. What was Randy’s diet like?

4. What was Randy’s occupation?

5. Did Randy have any notable past medical history?

6. Was there any remarkable family medical history?

7. What were some of Randy’s lifestyle habits?

8. Did Randy have any weight changes? If so, what were they?

9. What were Randy’s sleep habits like?

10. What were Randy’s demographics?

11. Did Randy have any genitourinary symptoms?
### SIU School of Medicine Naturalistic Field Observation Tool

**Descriptive Observation Notes Template**

<table>
<thead>
<tr>
<th>Date of recording:</th>
<th>Case Type (AI or ePBLM):</th>
<th>Group Size (number of students):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer (your name):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer reflexivity:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description of Location/Space:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name and role of the research assistant in this session:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Take detailed notes on everything you observe that you deem to be noteworthy.* These notes should provide (but may not be limited to) detailed descriptions of anything related to student interactions with the PBL case medium (AI or ePBLM), case content, each other, and the tutor. Can include direct quotes.

<table>
<thead>
<tr>
<th>Timestamp (Taken from recording)</th>
<th>Observation of Activity (Who, What, When, Where, How)</th>
<th>Reflections (Your Thoughts On What You Observe)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: 0:07:21</td>
<td>Student asks the AI a question, referring to the AI by name, reacts with apparent frustration to the AI’s indirect response to their question.</td>
<td>Interesting that the student treats the AI like a person, and has emotional reactions to what it says.</td>
</tr>
</tbody>
</table>