International Society of the Learning Sciences

AI-integrated virtual students for teacher training: Comparing simulation-based classroom dialogue with the real thing

Alex Barrett, Fengfeng Ke, Chih-Pu Dai, Luke West, Saptarshi Bhowmik, Nuodi Zhang
ab21b@fsu.edu, fke@fsu.edu, cd18m@fsu.edu, law19a@fsu.edu, sb17s@fsu.edu, nz21a@fsu.edu
Florida State University

Abstract: This paper reports on a study of AI-integrated virtual student agents designed for preservice teacher training. Eight preservice teachers engaged in a simulated science class where they conducted multiple lessons with the virtual students. Simulation lesson discourse was transcribed and compared to that of in-service teachers conducting lessons with real students. Results indicate that the virtual student authentically replicated declarative and interrogative patterns of discourse, but preservice teachers asked fewer questions than their in-service counterparts.

Introduction & theoretical background

Providing preservice teachers with authentic experience has been a focus for teacher educators. One way to accomplish this is through simulated teaching. Artificial intelligence (AI) can be used to render simulation-based virtual agents that reflect the cognitive-affective states of real students. However, AI-integrated virtual student agents need to be scrutinized for authenticity.

For agents to be perceived as authentic, cognitive-affective student models can be integrated into a language model (Dai & Ke, 2022) to generated discourse that is similar to that of real students. The difference between AI-generated discourse and that of human students remains an open question. This invites the use of discourse analysis as a tool for analyzing dialogue systems, which is a common tool for evaluating conversation agents to gain insight into the accuracy of the conversation agent's logic and the behavior of the human interlocutor (Hobert, 2019). A surface level inquiry into the function of statements made by the agents with preservice teachers on the one hand, and in-service teachers with real students on the other, can illuminate patterns of discourse that provide insight into both the agent's functionality and preservice teacher performance.

We therefore analyzed the classroom dialogue of preservice teachers and AI-integrated virtual students from a discourse perspective with the goal of comparing simulated and real classroom discourse, assessing both authenticity and efficacy for teacher training. The research question guiding this study is how do the mean frequencies and ratios of statement function (declarative, interrogative, imperative, and exclamatory) compare for teachers and students within and between actual and simulated classrooms?

Method

This split-plot study involved eight pre-service STEM teachers recruited from a teacher training program at a U.S. university. Participants underwent four-hour teaching practice sessions that included preparation, delivery, and reflection of teaching a STEM topic exemplified in Ambitious Science Teaching (Windschitl et al., 2018), a K-12 STEM teaching framework initiative. Lessons were delivered by the preservice teachers in *OpenSimulator*, a 3D virtual world, to AI-integrated virtual student agents. The agents were programmed with a generative, pre-trained transformer-based deep neural network model trained on authentic STEM classroom dialogue (see Bhowmik et al., 2022).

The text-based interactions between the preservice teachers and the AI-integrated virtual student agents resulted in a transcript of over 12,000 words for analysis across 12 separate teaching sessions. Additionally, 24,000 words were manually transcribed from actual teacher-student dialogue for comparison.

Data were analyzed by coding the transcripts by frequency of statement function based on the four typical function types: declarative, interrogative, imperative, and exclamatory. Frequencies were averaged per teaching session and ratios were calculated to allow comparisons between class contexts (actual and simulated). Select comparisons among these figures were made using an analysis of variance (ANOVA).

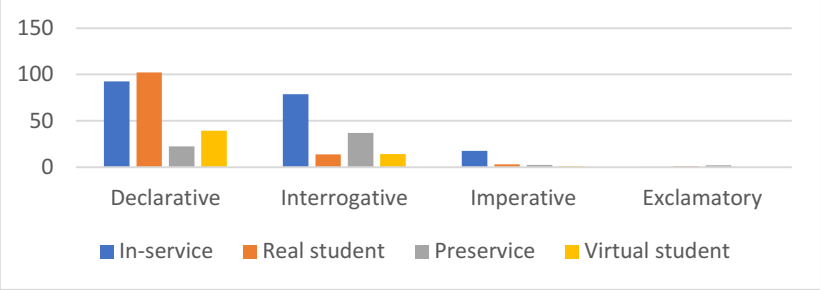
Results

Figure 1 depicts mean frequencies of statement type from the actual and simulated classroom sessions. In the simulated classroom, preservice teachers, on average, used significantly fewer declarative statements than the virtual student (F (1, 22) = 7.36, p = .013), but significantly more interrogatives (F (1, 22) = 8.22, p = .009). Imperative use was non-significant between preservice teachers and the virtual student (p = .181). For both

simulated and real contexts, exclamatory statements were too infrequent to warrant comparisons. In-service teachers in an actual classroom context, on average, also used significantly more interrogative statements (F (1, 16) = 14.24, p = .002). However, unlike in the simulated classroom, in-service teachers did not have a significant difference in declarative statement use from real students (F (1, 16) = .185, p = .673), and used imperatives significantly more (F (1, 16) = 14.13, p = .002).

Figure 1

Mean frequencies of statement types in dialogues from simulated and actual classrooms.



Note: Actual classroom sessions tended to be almost twice as long as simulated sessions and therefore between-group comparisons in this figure should be avoided.

Table 1 shows the teacher-to-student mean ratios for statement functions for dialogues that took place in both the simulated and actual classrooms, allowing for between-group comparisons. For declarative statements, the mean teacher-to-student ratios were non-significant between simulated and actual contexts (F (1, 19) = 1.24, p = .280), as were the imperative ratios (F (1, 8) = .10, p = .765). Mean interrogative ratios, on the other hand, were significantly different between contexts (F (1, 19) = 9.95, p = .005).

Table 1

Teacher/student mean ratios for statement types from simulated and actual classrooms.

	Declarative	Interrogative	Imperative	Exclamatory
In-service teacher / Real student	0.93	16.89	1.02	-
Pre-service teacher / Virtual student	0.71	4.04	1.36	-

Conclusion

To summarize, within both contexts, students used more declarative statements than teachers, and teachers used more interrogatives than students. Between contexts, teacher-to-student interrogative ratios were much larger in real classrooms, with in-service teachers using them almost 17 times more than students, compared to preservice teachers' four-fold use. These findings can aid in better coaching of preservice teachers and in further developing the AI-integrated virtual student agent.

References

Bhowmik, S., Barrett, A., Ke, F. Yuan, X., Southerland, S., Dai, C.-P., West, L. & Dai, Z. (2022). Simulating students: An AI chatbot for teacher training. In Chinn, C., Tan, E., Chan, C. & Kali, Y. (Eds.). *Proceedings of the 16th International Conference of the Learning Sciences – ICLS 2022*.

Dai, C.-P. & Ke, F. (2022). Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers & Education: Artificial Intelligence*, online. https://doi.org/10.1016/j.caeai.2022.100087

Hobert, S. (2019). How Are You, Chatbot? Evaluating Chatbots in Educational Settings - Results of a Literature Review. In Pinkwart, N. & Konert, J. (Eds.). *DeLFI 2019: Society for Informatics,* 259-270. https://doi.org/10.18420/delfi2019_289

Windschitl, M., Thompson, J. & Braaten, M. (2018). *Ambitious Science Teaching*. Harvard Education Press.

Acknowledgment

Research reported in this article is funded by the National Science Foundation, grant 1632965 and grant 2110777.