Pattern analysis of ambitious science talk between preservice teachers and AI-powered student agents

Alex Barrett*
Florida State University
Tallahassee, USA
abarrett3@fsu.edu

Fengfeng Ke
Florida State University
Tallahassee, USA
fke@fsu.edu

Nuodi Zhang
Florida State University
Tallahasse, USA
nzhang4@fsu.edu

Chih-Pu Dai
University of Hawai'i at Manoa
Honolulu, USA
cdai@hawaii.edu

Saptarshi Bhowmik
Florida State University
Tallahassee, USA
sb17s@fsu.edu

Xin Yuan
Florida State University
Tallahassee, USA
xyuan@cs.fsu.edu

Abstract

New frontiers in simulation-based teacher training have been unveiled with the advancement of artificial intelligence (AI). Integrating AI into virtual student agents increases the accessibility and affordability of teacher training simulations, but little is known about how preservice teachers interact with AI-powered student agents. This study analyzed the discourse behavior of 15 preservice teachers who undertook simulation-based training with AI-powered student agents. Using a framework of ambitious science teaching, we conducted a pattern analysis of teacher and student talk moves, looking for evidence of academically productive discourse. Comparisons are made with patterns found in real classrooms with professionally trained science teachers. Results indicated that preservice teachers generated academically productive discourse with AI-powered students by using ambitious talk moves. The pattern analysis also revealed coachable moments where preservice teachers succumbed to cycles of unproductive discourse. This study highlights the utility of analyzing classroom discourse to understand human-AI communication in simulation-based teacher training.

CCS Concepts

• ; • **Computing methodologies** → Artificial intelligence; Distributed artificial intelligence; Intelligent agents; • **Human-centered computing** → Human computer interaction (HCI); HCI design and evaluation methods; User studies; • **Applied computing** → Education; Interactive learning environments;

Keywords

Artificial intelligence, Teacher training, Discourse pattern analysis, Simulation

ACM Reference Format:
Alex Barrett, Fengfeng Ke, Nuodi Zhang, Chih-Pu Dai, Saptarshi Bhowmik, and Xin Yuan. 2025. Pattern analysis of ambitious science talk between preservice teachers and AI-powered student agents. In *LAK25: The 15th International Learning Analytics and Knowledge Conference (LAK 2025), March 03–07, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3706468.3706570

*Corresponding author.

LAK 2025, March 03–07, 2025, Dublin, Ireland

ACM ISBN 979-8-4007-0701-8/25/03
https://doi.org/10.1145/3706468.3706570

1 Introduction

Simulation-based learning has evolved from role playing with peers to highly immersive mixed reality environments. Recently, artificial intelligence (AI) has been integrated into immersive educational simulations to power virtual agents, affording personalized and situated learning [7]. One notable application of AI-powered agents in simulation-based learning has been in science teacher preparation, where preservice or novice science teachers practice classroom instruction with AI-powered students [8]. Integrating AI into teacher training simulations is more cost effective and accessible than the conventional approach of using a professional voice actor to puppeteer the students [9]. Furthermore, having student avatars puppeteered by adults leaves these simulations vulnerable to inauthentic student-teacher interactions. One of the essential purposes of science teacher training simulations is for science teachers to practice facilitating academically productive talk [16], therefore it is important for student-teacher interactions in the simulations to be as authentic as possible.

With recent developments in generative AI, teachable student agents powered by large language models (LLMs) and fine-tuned using natural language datasets from real-world science classrooms can be used [4]. Studies have shown these agents can generate discourse that reflects the idiosyncrasies of real students, increasing the authenticity of pre-service teacher training simulations [27, 28]. However, it is not yet known whether these agents can support multi-turn discussions that lead to rigorous science explanations, a key goal of ambitious science teaching practice [25]. Initial studies analyzing the sentence-level discourse between preservice teachers and AI-powered student agents have suggested there are similarities in the sentence function frequencies of virtual students and real students [2, 3], but less is understood about whether AI-powered student agents in teacher training simulations can help preservice teachers practice academically productive communication through ambitious teaching practices, or which specific talk moves preservice teachers might employ with AI-powered students.

Preservice science teachers' discourse interactions with AI-powered student agents are the central feature of simulation-based teacher training and therefore require scrutiny to both ensure the ecological validity of the training and locate design-based elements that can be improved upon. Therefore, the purpose of this study is to explore the discourse patterns between preservice science teachers and AI-powered student agents in simulation-based training and to share insights into human-AI interaction.

Specifically, our study is governed by the following research questions:

1. What are the frequent patterns of academically productive talk moves between preservice teachers and AI-powered student agents?

2. How do these discourse patterns compare to those of real-world classrooms?

2 Understanding ambitious science talk

2.1 Dialogic teaching in science classrooms

Science teaching involves adaptively applying instructional principles to facilitate learning through a means of purposeful dialogic communication. The dialogue that manifests between teachers and students in science classroom should promote epistemological understanding, argumentation skills, and science knowledge [20], as students collaborate with teachers in meaning making. This has historically been a struggle to achieve, as evidence has shown that science classrooms tend toward a monologic dynamic where the teacher is seen as the sole source of knowledge and students are passive recipients of that knowledge [11], or when dialogic interactions are designed to conform students to the teachers' point of view [22].

Ambitious Science Teaching (AST) [20] is a framework that follows the traditions of sociocultural theory [24] and advocates for dialogic teaching practices that engage students with science concepts, elicit their ideas, support their ongoing thinking, and collectively draw evidence-based explanations. At the discourse level, AST can be deconstructed into identifiable teacher talk moves which facilitate AST goals, and differentiated by conservative talk moves which can interfere with these goals [1].

2.2 Codifying science classroom dialogue

Academically productive science discourse is supported by ambitious talk moves that include probing, pressing, distributing participation, revoicing, and countering [22, 25]. A probing talk move is used to elicit student observations, thinking, or experiences. Pressing is used when asking for more information or examples stemming from a student turn at talk. Distributing participation seeks student-to-student and student-teacher cross talk to enrich or diversify the discussion. Revoicing is a strategic repeating or rephrasing of a student utterance that serves to highlight an idea or clarify an expression. And last, countering encourages students to think deeper about an idea by presenting an alternative perspective.

Conservative talk moves are those which hinder academically productive discourse, such as asking display questions, evaluating correctness, and minilecturing [1, 12]. Display questions are those which elicit a specific or predetermined answer. Evaluating correctness is a talk move that identifies a student utterance as correct or incorrect. And minilecturing is characterized as supplying information, such as an explanation or answer.

Student contributions can also be codified to better understand AST dialogic practice. Thompson et al. [23] categorized student talk according to the level of explanatory rigor. At the low rigor end this scale starts with student silence, understood as no student talk when talk is expected. This is followed by students providing definition or facts, often in response to teacher display questions; and then descriptions or observations, which are accountings of student experience or observation. On the high-rigor end of the scale are both under and fully theorized science explanations. A science explanation can be characterized by depicting a phenomenon as part of a process or relationship and talking about what is occurring on an unobservable level. These explanations can be considered under or fully theorized based on whether students include models, laws, or science theories, or going beyond simple cause and effect relationships.

2.3 Patterns of dialogic science teaching

Investigations into the influence of novice teacher talk moves on student explanatory rigor have shown that ambitious talk moves lead to higher rigor [1, 12]. In a study by Grinath and Southerland [12] examining the talk moves of novice teachers in science classrooms, it was found that conservative talk moves were significantly positively associated with low-rigor student contributions, such as providing facts, whereas ambitious talk moves were significantly positively associated with observations and explanations. In a closer analysis, Barns et al. [1] also found that ambitious talk moves were more likely to be followed by higher levels of student explanatory rigor, particularly in dialogues where students themselves posed rigorous questions or where teachers pressed for explanations. These two studies show that novice teachers employ ambitious talk moves, and that these talk moves elicit more rigorous scientific contributions from students. However, it remains unclear what specific patterns of discourse are being used or if such talk moves would arise with AI-powered student agents.

The validity of classroom simulation systems with AI-powered student agents rely on the authenticity of the discourse produced by the agents. There are various protocols available for evaluating AI dialogue systems, including automated and human evaluation [10]. Human evaluation of classroom dialogic interaction captures subjective qualities such as appropriateness and relevancy, as rated by experienced teachers. Automated evaluation provides objective assessment criteria such as similarity to training data. Classroom dialogue is highly dynamic and subject to influences such as teacher expertise, student age, and class size [19], therefore, the validity of AI dialogue systems for classroom discourse can be supported by human evaluation, with further comparisons to a training corpus adding insight.

3 Method

This study uses a case study design to explore the dialogic behaviors of preservice teachers and AI-powered student agents in interaction. We also include the in situ classroom discourse of highly trained teachers and human students to use as a model comparison. This design allows us to address each of our research questions through a close examination of participants' discursive behaviors.

Figure 1: *EVETeach* environment with AI-powered students, showing the panels for dialogue history (bottom-left) and text input (top-right).

3.1 Simulation environment and design of the AI-powered student agents

This study used a classroom teaching simulator called *EVETeach*, which employs open-source 3D virtual world technology to render a middle school science classroom populated with six AI-powered student agents. Users control a teacher avatar and can interact with individual students or the whole class through a text-based dialogue box. *EVETeach* was designed and developed as part of an ongoing research project aimed at supporting preservice teacher learning [14]. Figure 1 depicts the virtual environment of *EVETeach*, showing the teacher avatar and student agents.

The student agents are controlled through a cloud-based program powered by a fine-tuned and customized generative pre-trained transformer (GPT) [4]. After evaluating multiple large language models (LLMs) we decided that GPT-2, an open-source large language model by OpenAI, was most suitable for our purposes due to its susceptibility to training data and low hallucination rates compared to larger models [15, 26]. The LLM was fine-tuned using data from 13.5 hours of recorded classroom teacher-student dialogue obtained from open-access videos. These recordings were manually transcribed and tagged to differentiate between teacher and student turns at talk as well as topic domain to guide the LLM during response generation. Using these data, we fine-tuned the LLM over 500 iterations with a temperature of 0.6, a setting we determined through iterative testing [4]. This program was deployed in a virtual machine through Google Cloud and integrated into the virtual world environment. This design allows users to query the student agents and get responses that are authentic to student discursive behavior. The performance of *EVETeach* has undergone extensive validation assessment through a process including both human and automated testing [3–5]. Figure 2 depicts a simplified architecture of the AI-powered student agents.

3.2 Participants and procedure

We used nonprobability sampling to recruit 15 preservice math and science teachers (all female) from a teacher education program at a university in the southern US. Participant informed consent was obtained in adherence to the ethical protocols of the supervising university. Each participant was scheduled for a 1-hour simulation session guided by a researcher facilitator. Participants were given remote access to the simulation program from their personal computers using Zoom videoconferencing. The facilitator provided a brief tutorial and then guided the participants over the session as they practiced engaging AI-students with a science topic and building consensus around it. Each session was recorded for further analysis.

3.3 Data, coding, and reliability

The text-based interactions between participating preservice teachers and the AI-powered student agents were obtained from each session, resulting in a transcript of 23,734 words across 2,431 turns at talk for analysis. For the purpose of comparison, a second dataset was obtained from the classroom recordings of two in-service teachers who were professionally trained in ambitious science teaching. These data were sampled from the *EVETeach* pretraining corpus and consisted of 31,845 words across 2,122 turns at talk. We refer to these datasets as the simulated classroom discourse and the in situ classroom discourse, respectively.

Two researchers did independent and systematic coding on the same 10% of both datasets using a coding framework adapted from Barnes et al. [1] (Table 1). In addition to the codes used by Barnes et

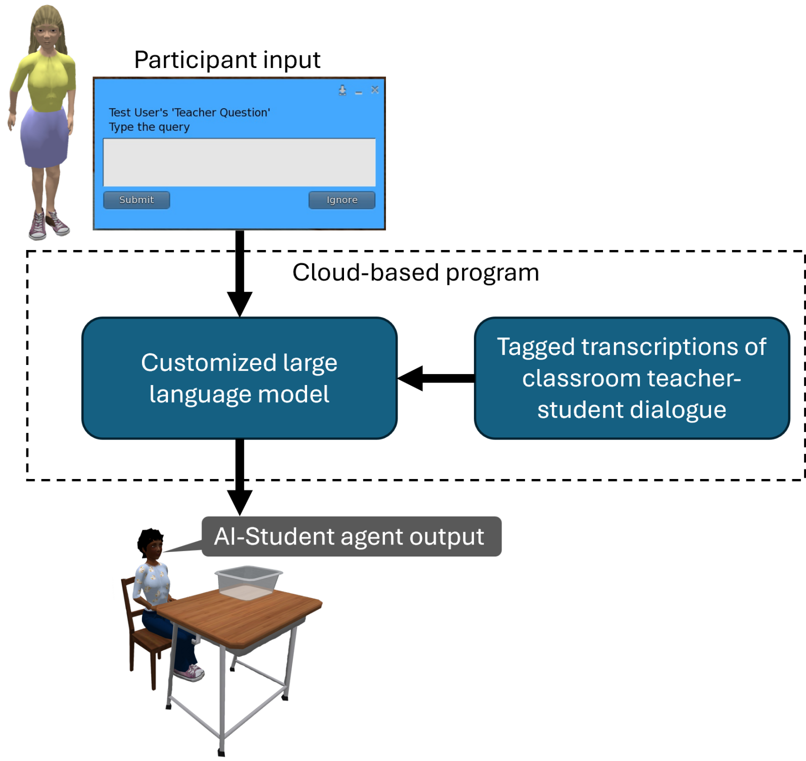

Figure 2: Architecture of the AI-powered student agent within the simulation environment.

al. [1] we added three additional codes to fully capture the teacher-student discourse. These codes are "counterclaim", an ambitious teacher talk move described by Teo [22] which involves presenting students with an alternative perspective; "pragmatic/management", which is both a teacher and student talk move that describes talk for communicating directions or socializing; and "question", a student talk move denoting an interrogative. After coding 10% of both datasets, researchers discussed any discrepancies until 100% agreement was reached and they felt comfortable coding the remaining data independently. Coding resulted in 2,638 identified talk moves in the simulated classroom discourse and 3,081 identified talk moves in the in situ classroom discourse.

3.4 Analysis

To answer the research questions, two analyses were performed. First, a generalized sequence pattern (GSP) algorithm was conducted in Python to analyze frequent talk move patterns [18]. Second, a Markov Chain transition matrix was generated using the markovchain package in R [21]. These analyses allowed us to discern the most common patterns of academically productive talk moves within both datasets and to understand the probability of transitioning from one talk move to another. Our research questions are concerned with academically productive discourse patterns. We defined academically productive discourse patterns as those which result in student contributions that could be characterized as highly rigorous, such as under and fully theorized science explanations. This can be contrasted with unproductive discourse that results in less rigorous student contributions.

Talk move patterns were filtered to identify outcomes relevant to this study's parameters. We specifically looked for teacher-initiated patterns, as this was the only dynamic possible in the simulated classroom. We also filtered for patterns that included both teacher and student talk. Last, we prioritized patterns that were at minimum three talk moves in length, although we also report shorter patterns specific to each dataset.

4 Results

4.1 Data descriptives

A descriptive analysis of talk moves as percentages of total talk from both datasets is shown in Figure 3. A visual inspection reveals similarities and differences across talk move occurrences. Teachers from in situ classrooms used revoicing (Trv) almost twice as much as preservice teachers did in the simulated classroom. In-service teachers also acknowledged contributions (Tac) about 8% more than preservice teachers did in the simulation. Preservice teachers tended to use minilecture (Tml) and evaluating correctness (Tec) more than teachers from in situ classrooms.

As for the students, human students contributed about 17% more under-theorized science explanations (Sue) than the AI-powered student agents did. Human students also made more descriptions and observations (Sdo) than the agents. The AI-powered agents

Table 1: Teacher and student talk moves adapted from Barnes et al. [1].

Role	Category	Talk move	Code	Example from simulated classroom
Teacher	Ambitious	Acknowledge contribution	Tac	"*Interesting idea Karen.*"
		Counterclaim	Tcc	"*But what happened to the pressure in the tanker car?*"
		Distribute participation	Tdp	"*Does anyone have any ideas on why that happened?*"
		Press for explanation	Tpe	"*What do you mean by that?*"
		Probing question	Tpq	"*What generalizations can we make about gasses and liquids from that?*"
		Revoice	Trv	"*John said that the steam got smaller.*"
	Conservative	Display question	Tdq	"*So do trees turn green in the summer or in the winter?*"
		Evaluate correctness	Tec	"*Karen is right.*"
		Minilecture	Tml	"*When steam turns into water we call that condensation – it's a state change.*"
	Other	Pragmatic	Tpm	"*Today we will be learning how to multiply fractions*"
Student	Less rigor	Silent (no student talk)	Sst	"(Silent)"
		Definition/ Fact	Sdf	"*They break down organic material into simple sugars*"
		Description/ Observation	Sar	[I've seen fungi] "*In the trees in my backyard*"
		Question	Sqn	"*And then what about the other two molecules?*"
	More rigor	Under theorized science explanation	Sue	"*Like, they're pulling in the sides of the tank, and so they get pushed in somehow.*"
		Fully theorized science explanation	Sfe	[No fully theorized science explanations were observed in the data]
	Other	Pragmatic	Spm	"*I'm doing very, very well.*"

asked questions (Sqn) much more than real students did and tended to make more pragmatic remarks (Spm).

All other talk moves, including teacher probing questions (Tpq), pressing for explanations (Tpe), display questions (Tdq), counterclaims (Tcc), and student silences (Sst), and definitions/facts (Sdf), were similar between the in situ and simulated classrooms, with less than 3.5% differences.

Fully theorized science explanations did not occur in either dataset. This was expected, as the data were based on lesson stages where teachers were introducing and building consensus on topics.

4.2 Discourse patterns

The GSP algorithm identified 19 talk move patterns that were common between both the in situ and simulated classrooms. The results of the analysis are shown in Tables 2, 3 and 4, along with support, a metric indicating the proportion of sequences that contain the specified pattern in the dataset, and confidence, a measure of reliability with values closer to 1 indicating a higher probability of the pattern tail appearing given the first talk move of the sequence.

Table 2 shows the 19 talk move patterns that were common between both datasets. Among these common patterns, nine included students' under theorized science explanations (Sue), indicating academically productive discourse. Teacher probing questions (Tpq) were the most common antecedents to high-rigor student contributions, occurring in eight of the nine academically productive discourse patterns, with teacher pressing for explanation (Tpe) being the ninth. Teacher probing question (Tpq) followed by student under theorized science explanation (Sue) and then teacher acknowledging contribution (Tac) had the highest support in the in situ classroom among the patterns common to both datasets, occurring in 21.4% of the identified sequences. This was much lower in the simulated classroom, with the pattern occurring in only 5.9% of sequences.

Another important note from Table 2 is that preservice teachers using probing questions were met with low-rigor descriptions/observations (Sdo) from the AI-powered students more commonly than occurred with students in the in situ classroom. The talk move pattern 'Tpq → Sdo → Tpq' occurred in 19.6% of simulated classroom sequences with a confidence score of .438, but only occurred in 7.9% of in situ classroom sequences and confidence at .185. In-service teachers from in situ classrooms were more likely to use the revoice (Trv) or acknowledge contribution (Tac) talk moves given the 'Tpq → Sdo' antecedent, instead of following up with another probing question like in the simulated classroom.

Looking at Tables 3 and 4, which contain teacher-led talk move sequences with the highest support, we can see some commonality between the two datasets. Namely, the 'Tpq → Sdo' pattern was the most common discourse exchange in both datasets, occurring in 42.7% of in situ classroom and 44.9% of simulated classroom discourse sequences. Also, 'Tpq → Sue' was the second most common sequence, occurring in 40.6% and 30.8% of sequences in the in situ and simulated classrooms respectively.

Certain salient differences between in situ and simulated classroom teacher-student talk are also notable. Preservice teachers in the simulated classroom commonly used conservative talk moves, such as minilectures (Tml), which resulted in a high frequency of lower-rigor student contributions such as descriptions/observations (Sdo) and questions (Sqn). In-service teachers from in situ classrooms avoided conservative talk moves in their top five most

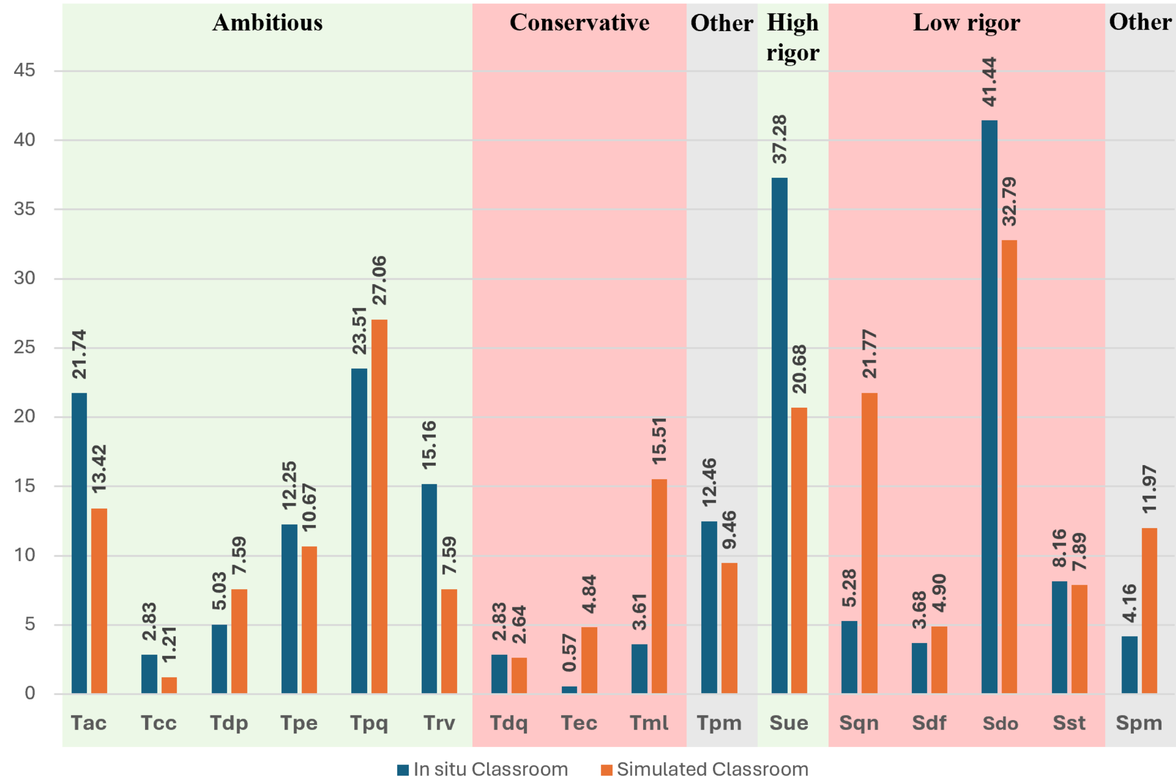

Figure 3: Comparison of talk moves as a percentage of all talk moves per dataset and speaker role (student or teacher).

Table 2: Frequent talk move patterns common in simulated and in situ classrooms, ranked by support of in situ classroom talk moves.

Talk move sequences common in both datasets	In situ classroom		Simulated classroom	
	Support	Confidence	Support	Confidence
Tpq, Sue, Tac	0.214	0.526	0.059	0.192
Tpq, Sar, Tac	0.164	0.383	0.084	0.188
Tac, Tpq, Sue	0.116	0.396	0.037	0.353
Tac, Tpq, Sar	0.108	0.369	0.034	0.324
Trv, Tpq, Sue	0.087	0.434	0.028	0.300
Tpq, Sar, Trv	0.084	0.198	0.034	0.076
Tpq, Sue, Tpe	0.082	0.201	0.056	0.182
Tpq, Sar, Tpq	0.079	0.185	0.196	0.438
Trv, Tpq, Sar	0.066	0.329	0.040	0.433
Tpq, Sar, Tpe	0.061	0.142	0.050	0.111
Tpe, Sar, Tpq	0.053	0.253	0.047	0.326
Tpq, Sue, Tac, Tpq	0.053	0.247	0.022	0.368
Tpq, Sue, Trv	0.050	0.123	0.025	0.081
Tpq, Sar, Tpq, Sue	0.040	0.500	0.050	0.254
Tpq, Sue, Tpq	0.034	0.084	0.100	0.323
Tpq, Sar, Tpm	0.029	0.068	0.022	0.049
Tpq, Sar, Trv, Tpq	0.026	0.313	0.028	0.818
Tpq, Sar, Tpq, Sar	0.024	0.300	0.087	0.444
Tpe, Sue, Tpq	0.021	0.114	0.028	0.500

Table 3: Top five most supported sequences from in situ classrooms.

Talk move sequence	Support	Confidence
Tpq, Sar	0.427	0.488
Tpq, Sue	0.406	0.464
Tpq, Sue, Tac	0.214	0.526
Tpe, Sar	0.208	0.457
Tpe, Sue	0.185	0.405

Table 4: Top five most supported sequences from simulated classrooms.

Talk move sequence	Support	Confidence
Tpq, Sar	0.449	0.585
Tpq, Sue	0.308	0.402
Tml, Sqn	0.240	0.546
Tpq, Sar, Tpq	0.196	0.438
Tml, Sar	0.184	0.418

frequent patterns and were rewarded with higher-rigor student contributions. Also of note, teachers in the in situ classroom followed high-rigor student contributions with acknowledgements (Tac), whereas preservice teachers did not do this with the AI-powered student agents. The fourth most supported pattern in simulated classrooms shows preservice teachers using probing questions, an ambitious talk move, but getting definitions/observations (Sdo) from the AI-powered student agents, which they then followed up with another probing question. This shows that preservice teachers persisted in attempts to guide the AI-powered student agents towards higher-rigor contributions.

4.3 Talk move transition probability

The transition matrices in Figures 4 and 5 show the probability of a consequent talk move following an antecedent for the simulated and in situ classrooms, respectively. Transition probabilities were omitted if they fell below .25 to make the figures easier to read and to highlight the most probable transitions.

Some commonalities between groups are evident. For example, in both figures teacher probing questions (Tpq) featured prominently, attracting the most talk moves transitions. Another similarity between Figures 4 and 5 is that ambitious teacher talk moves were more likely to lead to high-rigor student contributions compared to conservative talk moves.

Differences between the figures are conspicuous. For instance, preservice teachers rarely acknowledged the contributions of the AI-powered students, whereas teachers in the in situ classrooms did so for most student contributions. Additionally, teachers from in situ classrooms tended to transition from conservative talk moves to ambitious talk moves. For example, evaluating correctness (Tec), minilectures (Tml), and pragmatics (Tpm) each had over 30% probability of shifting to a probing question next. This means that although the in-service teachers used some conservative talk moves, they avoided allowing students to respond to them by tacking ambitious talk moves to the end of their turns at talk. In the simulated classroom, however, conservative talk moves tended to transition to student contributions. When using the minilecture (Tml) talk move in the simulated classroom, preservice teacher had a 41% probability of receiving a student question (Sqn), which in turn had a 39% probability of transitioning to another minilecture, creating a cycle of unproductive academic discourse.

However, cycles of productive academic discourse were also evident in both groups. In the simulated classroom, teacher probing questions (Tpq) had a 28% probability of leading to under-theorized science explanations (Sue) from students, which in turn had a 31% probability of leading to additional probing questions. A similar cycle was found with in situ classrooms, although teachers acknowledging student contributions (Tac) was an intermediary following Sue and prior to probing questions.

5 Discussion

Our research questions inquired after the patterns of productive academic discourse observed between preservice teachers and AI-powered students in teacher training simulations. We used data from in situ classrooms with teachers professionally trained in ambitious science teaching as a model comparison. Results indicated that preservice teachers were able to use ambitious science talk moves to foster academically productive discourse with AI-powered student agents during teaching simulations.

Discourse data from the simulated classrooms showed that teacher probing questions were used during most dialogic interactions with the AI-powered students. Probing questions tended to lead to high-rigor student contributions, which in our data meant under-theorized science explanations. When students gave less rigorous responses, such as descriptions and observations (Sdo), preservice teachers were most likely to follow up with ambitious probing questions, encouraging the AI-powered students to move toward higher-rigor contributions. These findings are encouraging and highlight the utility of AI-integrated teacher training simulations for practicing ambitious science teaching.

The results also revealed opportunities for improving preservice teacher's dialogic teaching. Cycles of unproductive discourse in the simulated classrooms showed that at times preservice teachers acted as the sole source of information in the classroom, responding to AI-powered student questions with minilectures which in turn generated more questions. Instances such as this can become coachable moments in simulation-based training, or instances for reflection where teachers can review how such unproductive cycles develop and perpetuate.

The quality of the discursive behavior of the AI-powered student agents reflected well for the design approach used in developing them. When participants in the simulation-based training used ambitious talk moves, these were likely to be followed by higher-rigor student responses, just as is seen with human students. This realism reinforces the preservice teachers' ambitious science teaching practice, demonstrating the utility of AI-integrated simulation-based teacher training.

By comparing the simulation discourse data with the discourse of in situ classrooms we can see some elements that may inform

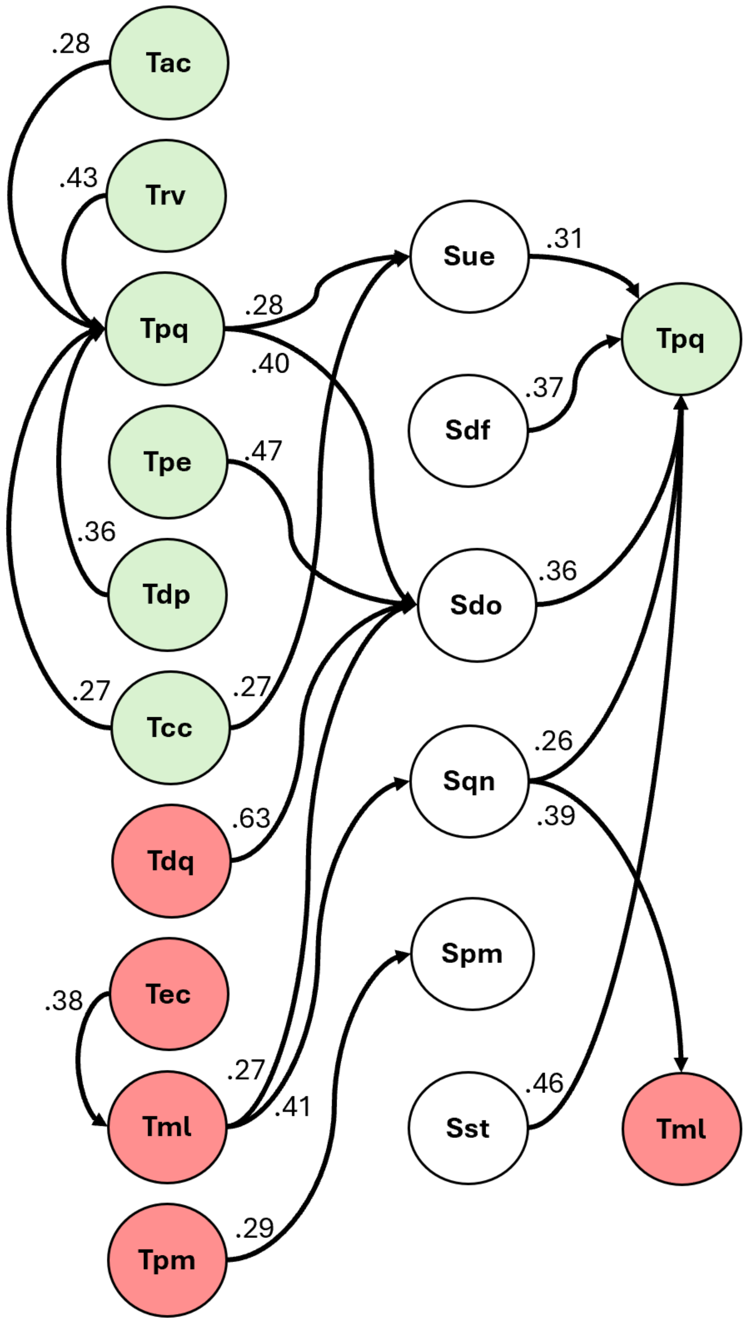

Figure 4: Talk move transition probabilities for simulated classroom discourse between preservice teachers and AI-powered student agents. Ambitious talk moves are in green, conservative talk moves are in red, and student talk moves are in white.

design considerations for teaching simulations. Namely, preservice teachers hardly acknowledged or revoiced the contributions of the AI-powered student agents, which was a highly probable talk move transition in the in situ classrooms. This may be due to the text-based format of the simulation-based dialogue, where responses were not as immediate as would occur during spoken discourse. Allowing for spoken interactions in the simulator could address this. Another possibility is that preservice teachers had a low sense of social presence, the feeling of being in the simulation with another [6], which could influence their behavior [13]. In essence, if preservice teachers viewed the AI-powered students with more legitimacy, they might have been prone to use more talk moves designed to encourage participation and clarify understanding such as ‘Tac’ and ‘Trv’. This might be remedied by including design

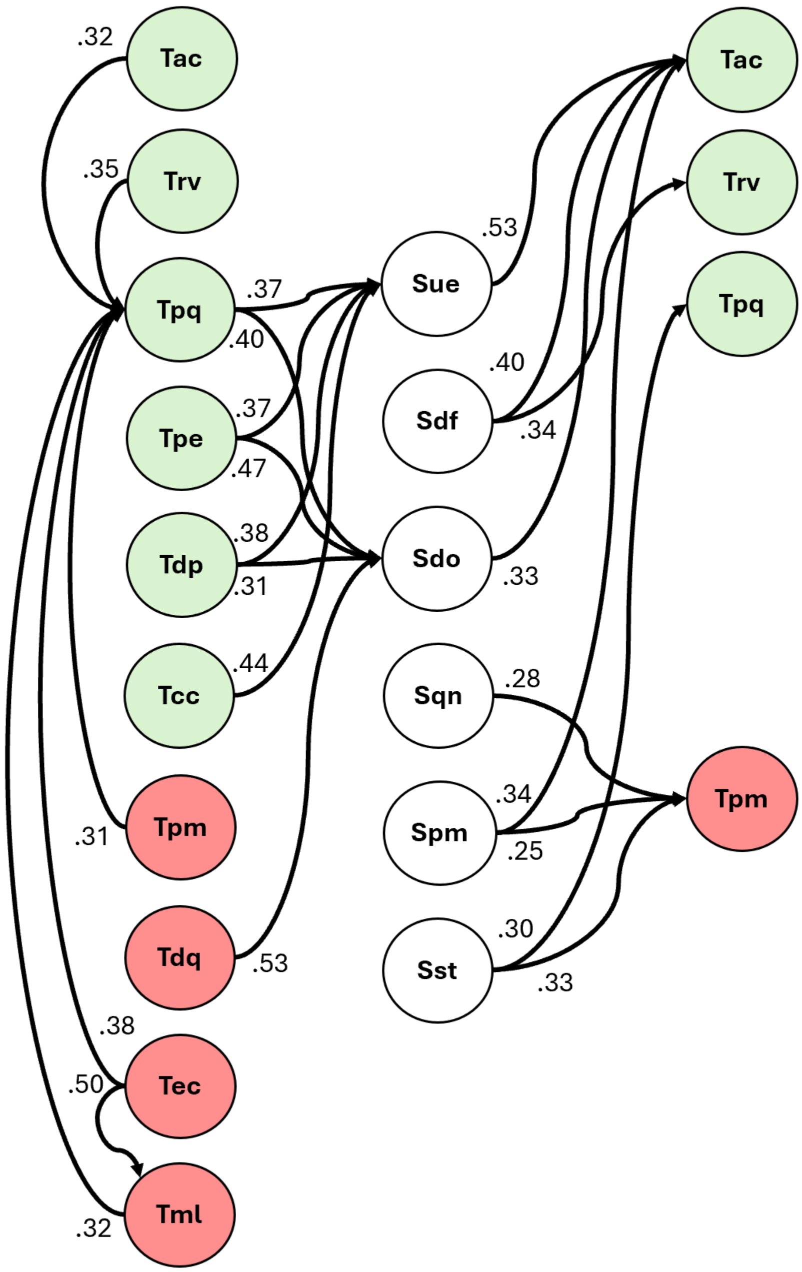

Figure 5: Talk move transition probabilities for in situ classroom discourse between in-service teachers and human students. Ambitious talk moves are in green, conservative talk moves are in red, and student talk moves are in white.

features that increase the users' sense of social presence such as increased sensory immersion [17].

6 Conclusion

AI-powered student agents increase the accessibility, affordability, and authenticity of teacher training simulations. In this study, preservice teachers were able to practice eliciting academically productive discourse and develop their ambitious science teaching talk moves. Additionally, analyzing the discourse patterns that occurred during the simulation-based training revealed instances where preservice teachers required support as well as possibilities to improve on the design of the simulator. Future research should explore the implementation of adaptive supports for preservice teachers to guide them toward more academically productive discourse. Incorporating design features in the simulator such as voice interaction should also be explored as a strategy to further develop preservice teachers' ambitious science teaching.

Acknowledgments

This work was supported by the National Science Foundation, grant no: [2110777]

References

[1] Evan R. Barnes, Ron Gray, and Anna S. Grinath. 2022. Talk moves as pedagogical tools for eliciting and working with student ideas in an undergraduate general biology laboratory. *Science Education, 107*, 89-123. https://doi.org/10.1002/sce.21762

[2] Alex Barrett, Fengfeng Ke, Chih-Pu Dai, Luke West, Saptarshi Bhowmik, and Nuodi Zhang. 2023. AI-integrated virtual students for teacher training: Comparing simulation-based classroom dialogue with the real thing. In Blikstein, P., Van Aalst, J., Kizito, R., & Brennan, K. (Eds.), *Proceedings of the 17th International Conference of the Learning Sciences (ICLS)*, (pp. 1797-1798). https://doi.org/10.22318/icls2023.112492

[3] Alex Barrett, Fengeng Ke, Nuodi Zhang, and Chih-Pu Dai. 2024. Comparing the science talk of AI and human students. In Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences (ICLS)*, (pp. 2073-2074). https://doi.org/10.22318/icls2024.963739

[4] Saptarshi Bhowmik, Alex Barrett, Fengfeng Ke, Xin Yuan, Sherry Southerland, Chih-Pu Dai, Luke West, and Zhaihuan Dai. 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)*, (pp. 1972-1973). https://dx.doi.org/10.22318/icls2022.1972

[5] Saptarshi Bhowmik, Luke West, Alex Barrett, Nuodi Zhang, Chih-Pu Dai, Zlatko Sokolikj, Sherry Southerland, and Xin Yuan. 2024. Evaluation of an LLM-powered student agent for teacher training. In R. F. Mello, N. Rummel, I. Jivet, G. Pishtari & A. Ruipérez Valiente (Eds.), *Technology enhanced learning for inclusive and equitable quality education (EC-TEL 2024)*. Springer. https://doi.org/10.1007/978-3-031-72312-4_7

[6] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence, 12*(5), 456-480. https://doi.org/10.1162/105474603322761270

[7] Chih-Pu Dai, and Fengfeng Ke. 2022. Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers and Education: Artificial Intelligence*, 100087. https://doi.org/10.1016/j.caeai.2022.100087

[8] Chih-Pu Dai, Fengfeng Ke, Nuodi Zhang, Alex Barrett, Luke West, Saptarshi Bhowmik, Sherry Sutherland, and Xin Yuan. 2024. Designing conversational agents to support student teacher learning in virtual reality simulation: A case study. In the *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (*CHI 2024)*. https://doi.org/10.1145/3613905.3637145

[9] Lisa A. Dieker, Carrie Straub, Michael Hynes, Charles E. Hughes, Caitlyn Bukathy, Taylor Bousfield, and Samantha Mrstik. 2019. Using virtual rehearsal in a simulator to impact the performance of science teachers. *International Journal of Gaming and Computer-Mediated Simulations*, *11*, 1–20. https://doi.org/10.4018/IJGCMS.2019100101

[10] Sarah E. Finch, and Jinho D. Choi. 2020. Towards unified dialogue system evaluation: A comprehensive analysis of current evaluation protocols. In the *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, (pp. 236-145). https://doi.org/10.18653/v1/2020.sigdial-1.29

[11] Paulo Freire. 1970. *Pedagogy of the oppressed*. Seabury Press.

[12] Anna S. Grinath, and Sherry A. Southerland. 2018. Applying the ambitious science teaching framework in undergraduate biology: Responsive talk moves that support explanatory rigor. *Science Teacher Education, 103*, 92-122. https://doi.org/10.1002/sce.21484

[13] Eugy Han, and Jeremy N. Bailenson. 2024. Social interaction in VR. *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.1489

[14] Fengfeng Ke, Xin Yuan, and Sherry Southerland. 2021. *Teaching Practices with Multiplayer Mixed Reality Simulations and Virtual Students*. National Science Foundation (Award Abstract # 2110777). Retrieved from https://www.nsf.gov/awardsearch/showAward?AWD_ID\protect\TU\textdollar$=$\protect\TU\textdollar2110777

[15] Junliang Luo, Tianyu Li, Di Wu, Michael Jenkin, Steve Liu, and Gregory Dudek. 2024. Hallucination detection and hallucination mitigation: An investigation. arXiv:2401.08358. Retrieved from https://doi.org/10.48550/arXiv.2401.08358

[16] Jamie Mikeska, Heather Howell, and Devon Kinsey. 2023. Do simulated teaching experiences impact elementary preservice teachers' ability to facilitate argumentation-focused discussions in mathematics and science? *Journal of Teacher Education, 74*(5), 422-436. https://doi.org/10.1177/00224871221142842

[17] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI, 5*(114). https://doi.org/10.3389/frobt.2018.00114

[18] Jackson Antonio do Prado Lima. 2020. *GSP-Py – Generalized sequence pattern algorithm in Python*. https://doi.org/10.5281/zenodo.3333987

[19] Chrysi Rapanta, Andrea Miralda-Banda, Mercè Garcia-Milà, Maria Vrikki, Fabrizio Macagno, and Maria Evagorou. 2023. Detecting the factors affecting classroom dialogue quality. *Linguistics and Education, 77*(101223). https://doi.org/10.1016/j.linged.2023.101223

[20] Alina Reznitskaya, and Maughn Gregory. 2013. Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist, 48*(2), 114-133. https://doi.org/10.1080/00461520.2013.775898

[21] Giorgio A. Spedicato, Tae S. Kang, Sai B. Yalamanchi, Mildenberger Thoralf, Deepak Yadav, Ignacio Cordón, Vandit Jain, Toni Giorgino, Richèl J. C. Bilderbeek, Daniel Ebbert, Shreyash Maheshwari, and Reinhold Koch. 2023. *Markovchain: Easy handling discrete time Markov chains*. https://doi.org/10.32614/CRAN.package.markovchain

[22] Peter Teo. 2016. Exploring the dialogic space in teaching: A study of teacher talk in the pre-university classroom in Singapore. *Teaching and Teacher Education, 56*, 47-60. http://dx.doi.org/10.1016/j.tate.2016.01.019

[23] Jessica Thompson, Sara Hagenah, Hosun Kang, David Stroupe, Melissa Braaten, Carolyn Colley, and Mark Windschitl. 2016. Rigor and responsiveness in classroom activity. *Teachers College Record, 118*(050303), 1-58.

[24] Lev Vygotsky. 1962. *Thought and language*. MIT Press.

[25] Mark Windschitl, Jessica Thompson, and Melissa Braaten. 2018. *Ambitious Science Teaching*. Harvard Education Press.

[26] Andy Xie, Camille Bruckmann, and Tracy Qian. 2022. Sensitivity analysis on transferred neural architectures of BERT and GPT-2 for financial sentiment analysis. arXiv:2207.03037. Retrieved from https://doi.org/10.48550/arXiv.2207.03037

[27] Nuodi Zhang, Fengfeng Ke, Chih-Pu Dai, Sherry A. Southerland, Alex Barrett, Saptarshi Bhowmik, Luke West, and Xin Yuan. 2024. Exploring Preservice Teachers' Perceptions and Experiences of Teaching AI Students in Virtual Simulations. In Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences (ICLS)*, (pp. 2417-2418). https://doi.org/10.22318/icls2024.484241

[28] Nuodi Zhang, Fengfeng Ke, Chih-Pu Dai, and Alex Barrett. 2024. Supporting preservice teachers' responsive teaching in AI-integrated simulations. In Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences (ICLS)*, (pp. 2051-2052). https://doi.org/10.22318/icls2024.502726