Supporting Teacher Learning with Large Language Models Simulated Student Dialogues

Shunki Kure, School of Education, University of North Carolina at Chapel Hill, United States, shk@unc.edu
Ha Nguyen, School of Education, University of North Carolina at Chapel Hill, United States, ha.nguyen@unc.edu
Jie Cao, School of Education, University of North Carolina at Chapel Hill, United States, jiecao@unc.edu

Abstract: Facilitating classroom discussion can be challenging for teachers, as they are not always able to anticipate and respond to students' ideas in the moment. We examine the potential of LLM-simulated student dialogues for teacher learning in science education. We find that LLMs can successfully generate discussions that leverage various dialogue moves and sequences to reflect classroom discourse. These simulations offer benefits for teacher sense-making, helping teachers identify potential student misconceptions and everyday reasoning and improve their lesson plans. We discuss the promise of LLM-based simulations for teacher preparation and the need for future research to enhance dialogue authenticity and variability.

Keywords: large language models, student dialogues, teacher learning, science education

Introduction

Providing students with opportunities to engage in peer discussion is crucial in science education, as discussion invites students to tap into prior knowledge, develop and refine ideas, and build on others' ideas (Shemwell & Furtak 2010). However, facilitating classroom discussion can be challenging for teachers, who are not always able to anticipate and respond to students' ideas in the moment (Davis & Simmt 2006). Various teacher training tools, from teaching rehearsals to video analysis of classroom situations, aim to provide teachers with opportunities to analyze, elicit, and build on students' ideas (Sherin & van Es 2009). However, these tools are costly; they rely on experienced teachers or teacher educators to enact the rehearsal and provide feedback to teachers. Additionally, the content of the teaching rehearsals or video analysis might not always align with the lessons that the teachers are preparing. In this research, we investigate the use of Large Language Models (LLMs) to provide an accessible and less costly solution (Barrett et al. 2025) to simulate students' science understanding in peer discussion. While researchers have explored the use of LLMs to simulate individual students' learning (Barrett et al. 2025; Zheng et al. 2025), our examination of student peer discussion offers unique affordances through surfacing student ideas and participation dynamics. The LLM-simulated dialogues serve as a tool for reflection for teachers to (a) illustrate student understanding building on their everyday and scientific explanations, and (b) help teachers derive insights to inform instruction, such as revising lesson plans to better reflect student thinking. This simulation tool allows teachers to practice diagnosing students' reasoning patterns, identify knowledge gaps, and design effective scaffolds. For the simulated dialogues to be useful, they need to authentically reflect student thinking and seed productive insights for teachers. To these ends, we investigate two research questions (RQs):

RQ1: How diverse are the talk moves, as demonstrated by their frequencies and sequences, in the LLM-simulated student dialogues?

RQ2: What do teachers notice through making sense of the LLM-simulated dialogues, and how can those insights inform instruction?

Literature Review

Simulating Student Dialogues with Large Language Models

Researchers have increasingly explored the use of LLMs to "roleplay" as students or teachers and generate text that simulates student thinking and student-teacher dialogues, as one way to reveal gaps in student understanding to improve lesson planning (Hu et al. 2025), highlight common errors that students can make in a learning task (MacNeil et al. 2024), generate better tutoring interactions (Scarlatos et al. 2025), and support teachers in practicing teaching moves (Barrett et al. 2025). For example, Barrett et al. (2025) found that preservice teachers were able to leverage talk moves that acknowledge students' contributions, press for explanations, and ask probing questions to increase the rigor in student thinking when interacting with LLM-powered student agents, similar to interactions in real-world classrooms. These findings highlight the potential of LLM simulations to support teaching practices.

Simulating Student Dialogues to Support Responsive Science Teaching

While prior research has simulated individual students' responses (e.g., within tutoring systems) or student-teacher dialogues (Xu et al. 2025; Zheng et al. 2025), we focus on student-student peer discussions as an important aspect of science education. This emphasis is grounded in *responsive teaching* frameworks that highlight the importance of eliciting, noticing, and pursuing the substance of students' ideas in individual, peer, and whole-class discussions to deepen science knowledge (Hammer et al. 2012; Robertson et al. 2015). Student discussion plays a particularly important role in promoting responsive science teaching, as it allows students to engage with scientific content, develop learning motivation, and most importantly, makes student ideas visible to peers and teachers (Gillies et al. 2012). Teachers can listen in and take up ideas from students' group dialogues, connect these ideas to a prior contribution, and invite knowledge-building moves from other students.

For LLMs to successfully simulate students' dialogues in ways that are meaningful for instruction, the dialogues need to reflect characteristics of authentic student interactions. Students can employ multiple *dialogue moves*: initiating or extending an idea, agreeing or disagreeing with one another, and asking for information or confirmation (Kaartinen & Kumpulainen 2002). They might also leverage different forms of explanation, such as formal explanation that reflects scientific reasoning, everyday explanation from their own experiences and everyday language, and a mix of formal and everyday explanations (Lee & Irving 2018). Additionally, as student dialogues are *dynamic* and science understanding can evolve through interactions (Coleman et.al. 2009; Huang et al 2019), it is also important to understand how the interactions unfold and how students generate, refine, and extend ideas throughout the dialogues. In summary, the simulated dialogues should reflect both the diversity of students' talk moves and the dynamic sequence in which these moves evolve.

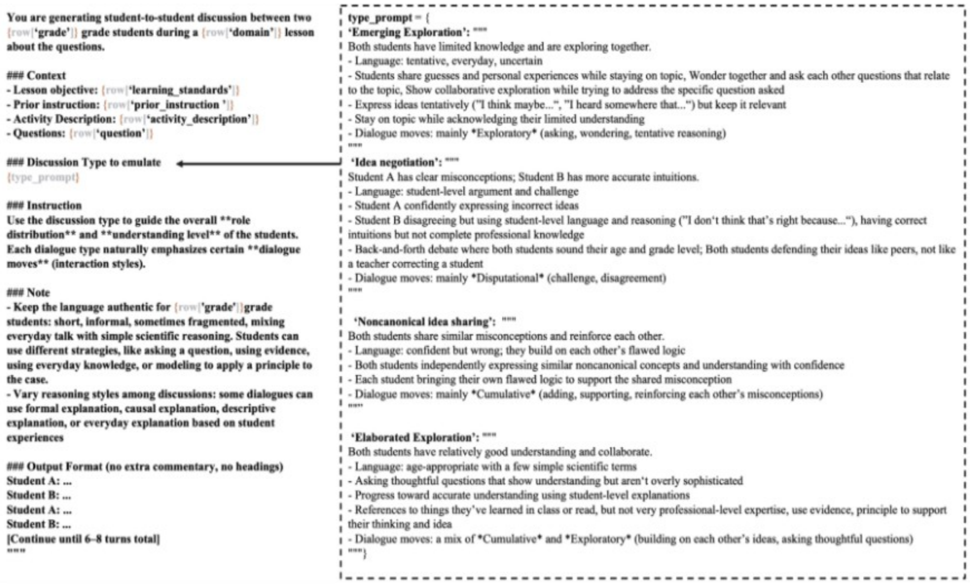
Methods

Overall, our research can be broken down into four steps. First, we created a dataset of simulated student dialogues, in response to learning activities from 45 publicly available OpenSciEd lesson plans. Second, we developed prompts to an LLM (OpenAI's GPT-5) to generate the dialogues. Third, we analyzed the simulated dialogues for the diversity and sequence of talk moves (RQ1). Finally, we interviewed four middle and high school teachers, who reviewed the simulated dialogues and articulated potential instructional insights that emerged (RQ2).

Data Sources

To seed the LLM dialogues, we selected discussion questions from 45 publicly available OpenSciEd lesson plans equally divided across grade levels and domains (chemistry, life sciences, and physics). These lesson plans were written by teams of experienced teachers and teacher educators; aligned with the Next Generation Science Standards; and broadly implemented in real-world science classrooms (Edelson et al. 2021). More importantly, the discussion questions that we included aligned with responsive science teaching frameworks to elicit student reasoning and peer discussion. Example questions in an 8th-grade life science lesson are: "Why are some baby cattle born with big muscles while others aren't? Where do the babies with big muscles get that trait variation?" These "how" and "why" questions invite students to make observations, gather evidence, and co-construct reasoning.

We then generated the LLM pair dialogues. We provided details about the learning tasks, including the students' grade levels, learning standards, prior instruction, task description, and discussion questions. We also defined four distinct dialogue types to simulate different dynamics of science discussion (Kaartinen & Kumpulainen 2002; Robertson et al. 2015). These included *Emerging Exploration*, where the simulated students generated initial hypotheses based on everyday explanations; *Elaborated Exploration*, where students co-constructed ideas leveraging both everyday and formal scientific explanations; *Idea Negotiation*, where students had different ideas and were disagreeing with one another; and *Noncanonical Idea Sharing*, where both students had incorrect scientific ideas. Figure 1 shows the prompt for dialogue generation. In total, we generated 120 simulated dialogues (*n* = 30 per dialogue type).



```
You are generating student-to-student discussion between two
{row[‘grade’]} grade students during a {row[‘domain’]} lesson
about the questions.

### Context
- Lesson objective: {row[‘learning_standards’]}
- Prior instruction: {row[‘prior_instruction ’]}
- Activity Description: {row[‘activity_description’]}
- Questions: {row[‘question’]}

### Discussion Type to emulate
{type_prompt}

### Instruction
Use the discussion type to guide the overall **role
distribution** and **understanding level** of the students.
Each dialogue type naturally emphasizes certain **dialogue
moves** (interaction styles).

### Note
- Keep the language authentic for {row['grade']}grade
students: short, informal, sometimes fragmented, mixing
everyday talk with simple scientific reasoning. Students can
use different strategies, like asking a question, using evidence,
using everyday knowledge, or modeling to apply a principle to
the case.
- Vary reasoning styles among discussions: some dialogues can
use formal explanation, causal explanation, descriptive
explanation, or everyday explanation based on student
experiences

### Output Format (no extra commentary, no headings)
Student A: ...
Student B: ...
Student A: ...
Student B: ...
[Continue until 6–8 turns total]
"""
```

```
type_prompt = {
‘Emerging Exploration’: """
Both students have limited knowledge and are exploring together.
- Language: tentative, everyday, uncertain
- Students share guesses and personal experiences while staying on topic, Wonder together and ask each other questions that relate
to the topic, Show collaborative exploration while trying to address the specific question asked
- Express ideas tentatively ("I think maybe...", "I heard somewhere that...") but keep it relevant
- Stay on topic while acknowledging their limited understanding
- Dialogue moves: mainly *Exploratory* (asking, wondering, tentative reasoning)
"""
 ‘Idea negotiation’: """
Student A has clear misconceptions; Student B has more accurate intuitions.
- Language: student-level argument and challenge
- Student A confidently expressing incorrect ideas
- Student B disagreeing but using student-level language and reasoning ("I don‘t think that’s right because...“), having correct
intuitions but not complete professional knowledge
- Back-and-forth debate where both students sound their age and grade level; Both students defending their ideas like peers, not like
a teacher correcting a student
- Dialogue moves: mainly *Disputational* (challenge, disagreement)
"""


  ‘Noncanonical idea sharing’:  """
Both students share similar misconceptions and reinforce each other.
- Language: confident but wrong; they build on each other’s flawed logic
- Both students independently expressing similar noncanonical concepts and understanding with confidence
- Each student bringing their own flawed logic to support the shared misconception
- Dialogue moves: mainly *Cumulative* (adding, supporting, reinforcing each other’s misconceptions)
"""

 ‘Elaborated Exploration’: """
Both students have relatively good understanding and collaborate.
- Language: age-appropriate with a few simple scientific terms
- Asking thoughtful questions that show understanding but aren‘t overly sophisticated
- Progress toward accurate understanding using student-level explanations
- References to things they’ve learned in class or read, but not very professional-level expertise, use evidence, principle to support
their thinking and idea
- Dialogue moves: a mix of *Cumulative* and *Exploratory* (building on each other’s ideas, asking thoughtful questions)
"""}
```

Figure 1. Prompt for the LLM to Simulate Student Dialogues

Consider the following example of an LLM-generated discussion among high school students, in response to the questions: “If yeast is a decomposer, how does it interact with matter and energy? Where is the matter and energy in sugar? In mapping the yeast model to the zombie fire system: What does the yeast represent? What does the water represent? What does the sugar represent? What do the bubbles represent? What does the beaker represent?”. The discussion represents natural back-and-forth that directly responds to the science inquiries and demonstrates knowledge co-construction.

Student A: I think yeast is acting like a tiny decomposer, right? Eating sugar like it would eat dead plant.

Student B: Yeah, that sounds right. The sugar has matter — like carbon and hydrogen atoms — and energy in its chemical bonds. Yeast breaks those bonds to get energy.

Student A: So in our model, the yeast would be the microbes or fungi in the peat that do decomposition. Makes sense. What about the water we add — is that just to keep the yeast alive?

Student B: Probably. I guess the water is like the wet conditions in the peat or the moisture that lets microbes do their work. It’s the medium where the reactions happen.

Student A: Okay. And the bubbles we saw — those are gas from the yeast, right? I think they’re carbon dioxide from breaking down the sugar.

Student B: I think so too. The bubbles = CO2 leaving the system. So sugar matter turns into CO2 and some stays in the yeast as biomass, and energy goes into making more yeast or doing work.

Analytic Procedures

To examine the diversity of the talk moves (RQ1), we coded for the moves in the simulated dialogues (*n*=915 moves within 120 dialogues), following Kaartinen & Kumpulainen (2002; Table 1). We also coded for the nature of explanation (everyday, formal, or both), in alignment with responsive teaching frameworks that emphasize students’ everyday thinking as productive for science discussion (Hammer et al. 2012). Two coders met to discuss and establish the coding principles. They independently coded 10% of the data, reached substantial inter-coder reliability (Cohen’s k = .84 for dialogue moves and .90 for nature of explanation), and divided the remaining data in half to code. We reported the descriptive statistics of the talk moves by dialogue types (Emerging Exploration, Elaborated Exploration, Idea Negotiation, Noncanonical Idea Sharing).

Table 1. Codebook for Talk Moves

Code	Description
Diversity of moves	
initialize	Initialize a discussion topic at the beginning of the conversation
ask questions	Ask for information or confirmation
respond	Respond to others’ questions
extend/continue	Build on others’ ideas
disagree	Express disagreement with peers
conclude	Draw conclusions from the discussion
Nature of explanation	
everyday explanation	Use everyday experiences and intuition
formal explanation	Use scientific terms, concepts, and models
everyday/formal	Contain both everyday and formal explanations

Additionally, we used process mining (Janssenswillen et al 2019) to summarize the typical interaction sequences of each dialogue type using process maps (i.e., directly-follows graphs that represent the sequenced talk moves). These analyses allowed us to evaluate both the diversity (indicated by frequency) and dynamics (shown by sequence) of LLM-simulated dialogues.

To answer what teachers noticed through the simulated dialogues (RQ2), we interviewed four teachers from Utah and California (50% female; years of teaching experience range 3-15 years). We selected the teachers to represent different subjects (physics, biology) and grade levels (middle and high school). The interviews were approved by the Institutional Review Board. They were audio-recorded and lasted 30-45 minutes each. To make the simulated dialogues more relevant and meaningful to teachers, we asked teachers to share their own lesson plans (instead of the OpenSciEd lessons) and applied the same prompt template (Figure 1) to generate student dialogues.

During the interviews, we asked teachers to review the dialogues and respond to several prompting questions about the dialogues’ usefulness. Example questions included: *What do you think about these students’ dialogues?; How well (or not) do the dialogues reflect what your students might say?; How can these dialogues be useful to your instruction?; How can we improve these dialogues?* In this paper, we report on two interview vignettes to illustrate potential ways in which the dialogues helped surface student ideas that teachers were previously unaware of and informed instructional improvement.

Findings

RQ1: Diversity of the Simulated Dialogues

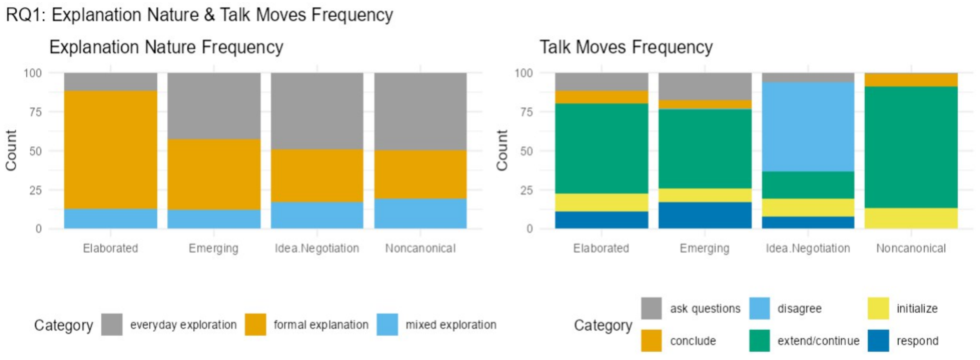


Figure 2. Talk Move Frequency by Dialogue Types

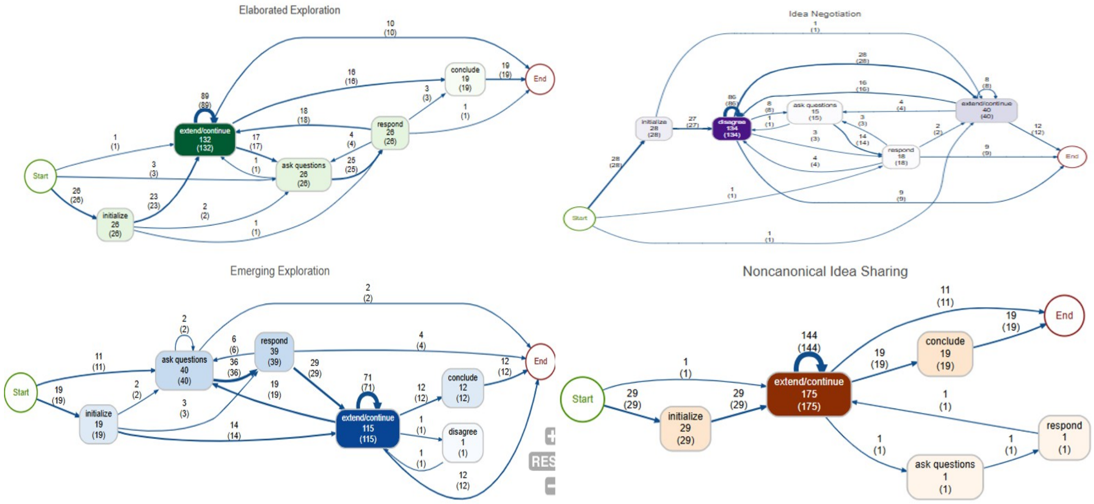


Figure 3. Discussion Process Maps

Our analysis indicated *diversity* in the simulated dialogues: the different dialogue types employed different talk moves to initiate, build on, and show agreement or disagreement with peers and showed variation in the substance of explanation. For example, the Elaborated Exploration dialogues were high in extending/continuing moves (57.6%), while Idea Negotiation showed higher frequencies of disagreeing moves (57%). Meanwhile, Emerging Exploration showed somewhat balanced nature of explanation (everyday, formal, and a mix of everyday and formal explanation); while Elaborated Exploration mostly showed formal explanation.

The process maps illustrated the different flows of each dialogue type. In the process maps, each node represents a talk move (e.g., initialize, extend, disagree), and arrows indicate the transitions between moves. Arrow thickness indicates frequency, and loops indicate repeated moves. Elaborated Exploration and Emerging Exploration showed similar dialogue structures, expanding the topic by continuing or extending others' ideas. In comparison, Idea Negotiation shows a lot of disagreement loops, and Noncanonical Idea Sharing did not involve refinement of ideas. These patterns are somewhat expected. We expect students engaging in idea negotiation to demonstrate disagreement with one another while extending the discussion, while those in idea sharing groups, who similarly shared partial understanding, might not critically engage in knowledge construction moves.

Together, findings suggest that the **LLM generally followed the prompts to simulate the different dialogue characteristics and dynamics**, reflecting what might be expected in student discussion.

RQ2: Teachers' Use Cases for the Simulated Dialogues

We illustrate two use cases for the LLM-simulated dialogues, drawing from teacher interviews. A first use case that we observed across interviews was that **the dialogues surfaced students' partial or incorrect understanding** that teachers might not have thought about, and these observations allowed teachers to generate insights to follow up with students. Consider the following example, where teacher Ramy (pseudonym) was creating a new physics lesson on modeling bird movement for 6th-grade students. The generated dialogues for Idea Negotiation incorrectly connected bird movement to how helicopters spin their blades.

> **Student A**: Scientists are into birds 'cause if you flap wings harder, you can fly higher and faster, like how helicopters spin their blades really fast.
>
> **Student B**: Yeah! And if we make our model flap really fast, it should go super high too. More flapping means more power.

After reviewing these dialogues, teacher RS noted: "I don't have a reference, understanding of children's thinking about birds. So this is helpful to see misconceptions, and I'd have follow-up prompts to connect their thinking. So another analogy, like helicopter, which connects to the previous example". Here, the LLM-simulated dialogues helped the teacher realize a potential misconception that the students might have and consider the everyday analogy as part of his future scaffolding for students.

A second use case is that **the simulated dialogues reflected how students might reach the learning goals in their discussion**, validating the learning activities and highlighting potential ways for teachers to build on students' ideas. Consider the following vignette from teacher Allan, who was working on a physics lesson on energy conservation for 8th-graders. The simulated dialogues for Emerging Exploration highlighted how students might approach an understanding of potential and kinetic energy.

> **Student A:** Then when it goes down, it gets faster, so that's the other energy—kinetic?
> **Student B:** Could be. But then, does it lose the first kind completely or just… switch it?
> **Student A:** I don't know. Like, maybe it's trading the height energy for speed?
> **Student B:** That makes sense, but then when it goes back up, how does it get more height again if it already used it?

Allan perceived this dialogue "interesting and useful". He stated that "If you ask students to converse in this manner, it shows that they understand the switching (in the word that they used)". Later in the interview, he mentioned that it is "useful for me to see which ideas to emphasize when they are having the discussion to see the relationship between energy, so they don't switch together." While the teacher noted that the language in the dialogue appeared more polished than what would be expected for students at this grade level, he reflected that reviewing the simulated dialogues helped him observe how students might describe the physics concepts in their own terms and considered concepts for which students might need more exploration or scaffolding.

Discussion

Our findings reveal that LLMs such as GPT-5 demonstrate strong adherence to the dialogue structure in the prompt strategy. We found that the dialogue types displayed diverse moves and sequences that were typical of their dialogue nature—emerging exploration, elaborated exploration, idea negotiation, and noncanonical idea sharing—to surface student thinking. For example, the *Noncanonical Idea Sharing* dialogues typically began with an initiating move, proceeded through an extending move, and concluded with a concluding statement that did not allow students to co-construct knowledge. Insights from teacher interviews revealed that these different discussion patterns helped teachers to better understand students' everyday thinking, highlight potential misconceptions that they can address or build on, and consider instructional areas to focus on in lesson planning and enactment.

Our research has practical implications for teacher education, by illustrating how LLM-simulated student dialogues can support teachers' diagnostic reasoning and responsive teaching to follow up on students' ideas. Engaging with simulated dialogues allows teachers to practice focusing on key information in real student discussions, such as diagnosing students' reasoning flow and anticipating points of conceptual difficulty. Effective scaffolding can reduce students' frustration and enhance their motivation, leading to more active classroom participation (Van der Stuyf 2002). We propose that teacher education programs and professional development opportunities can incorporate LLM simulations as complementary tools to enhance teachers' diagnostic practice in coursework and instructional coaching. By analyzing the simulated discussions, teachers and pre-service teachers can anticipate potential student ideas and reflect on which ideas the simulations may or may not capture. While these simulations cannot replace live student discourse that fosters teachers' adaptive expertise, they provide a safe and easily accessible way to anticipate and pursue potential students' reasoning. Through practicing with LLM simulated discussions, teachers can identify where knowledge gaps tend to occur in students and improve their professional vision to notice students' difficulties, interpret students' reasoning more accurately, and design more targeted instructional support, particularly as they prepare new learning activities.

Even though the consistency of LLM-generated student dialogues offers valuable opportunities for systemic exploration of discussion patterns, it also reveals the limitations of the simulations. The model's adherence to the LLM prompt may constrain the dynamic of dialogue moves, resulting in less spontaneity than is typical in authentic student discussions. Real classroom interactions might involve unpredictable talk moves and reasoning paths that are not yet fully captured in the simulated discourse. Additionally, the simulated dialogues often appeared more fluent and complete than what might be expected for student discussions. Future research should investigate and enhance the authenticity of the generated dialogues, for instance, by investigating additional dimensions of student discussion (e.g., language, content) and drawing on related work that simulates student-teacher dialogues (Barrett et al. 2025) to finetune LLMs with authentic student discussion data.

Limitations and Future Research

Our study offers initial insight into the diversity and usefulness of LLM-simulated discussions. Several limitations of the research should be considered in guiding future research. First, the findings are based on a small interview sample of four teachers, so they might have limited generalizability across grade levels, content areas, and teacher experience. We encourage future work to involve a larger and more diverse range of teachers and contexts. Second, the prompt design and the curricular focus of OpenSciEd materials shape the structure and content of the dialogues and might limit the range and spontaneity of students' idea evolvement. Third, while the findings highlight some advantages of LLM-simulated dialogues, we do not empirically compare the simulations with authentic student-student discussion due to limited data availability. Direct comparisons with real classroom discussions, where students experience the same curricula, can illustrate which dialogue moves, interactional dynamics, and student reasoning are missing from the simulated versions. Fourth, future research should incorporate real student discussion to further finetune the LLM and evaluate the extent to which the dialogues reflect the variability and complexity of classroom discourse. Another direction is to examine the simulations' affordances for supporting teaching practice in real classrooms. For example, researchers can explore how teachers might attend to specific aspects of student discussion decide on how to follow up on students' ideas, following their use and interpretation of the simulations.

Conclusion

In this study, we examined the potential of LLM-simulated student dialogues for teacher learning in science education. Findings indicate that LLMs can successfully generate discussions that leverage various dialogue moves and sequences to reflect classroom discourse. These simulations offer meaningful benefits for teacher sense-making, helping teachers identify potential student misconceptions and everyday reasoning and plan more effective scaffolding. Overall, these findings highlight the promise of LLM-simulated dialogues as tools for developing teachers' instructional reasoning and promoting responsive teaching practices to elicit and respond to student ideas in science discussion. To enhance their training value, future improvements should aim to balance prompt adherence with greater variability and naturalness and compare the simulated dialogues with real classroom data. Additionally, follow-up studies can implement the simulations in real classroom settings and examine which insights from the simulations teachers take up and incorporate into their lesson planning and enactment.

References

Barrett, A., Ke, F., Zhang, N., Dai, C.-P., Bhowmik, S., & Yuan, X. (2025). Pattern analysis of ambitious science talk between preservice teachers and AI-powered student agents. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 761–770. https://doi.org/10.1145/3706468.3706570

Campbell, T., & Lee, O. (2021). Instructional materials designed for *A Framework for K-12 Science Education* and the *Next Generation Science Standards*: An introduction to the special issue. *Journal of Science Teacher Education, 32*(7), 727–734. https://doi.org/10.1080/1046560X.2021.1975359

Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *Journal of the Learning Sciences, 6*(4), 347–365. https://doi.org/10.1207/s15327809jls0604_1

Davis, B., & Simmt, E. (2006). Mathematics-for-teaching: An ongoing investigation of the mathematics that teachers (need to) know. *Educational Studies in Mathematics, 61*(3), 293–319. https://doi.org/10.1007/s10649-006-2372-4

Edelson, D. C., Reiser, B. J., & Krajcik, J. S. (2021). *OpenSciEd instructional units: Designing NGSS-aligned materials for middle school science.* OpenSciEd. https://www.openscied.org

Freire, P. (1970). *Pedagogy of the oppressed.* New York, NY: Bloomsbury Academic. https://doi.org/10.5040/9781350183641

Gillies, R. M., Nicholson, P., & Burgh, G. (2012). Promoting reasoned argumentation, problem solving and learning during small-group discussions. *Teaching and Teacher Education, 28*(5), 683–692. https://doi.org/10.1016/j.tate.2012.02.004

Hammer, D., Goldberg, F., & Fargason, S. (2012). Responsive teaching and the beginnings of energy in a third-grade classroom. *Science Education, 96*(6), 879–900. https://doi.org/10.1002/sce.21014

Hu, B., Zhu, J., Pei, Y., & Gu, X. (2025). Exploring the potential of large language models to enhance teaching plans through teaching simulation. *NPJ Science of Learning, 10*(1). https://doi.org/10.1038/s41539-025-00300-x

Huang, C.-Q., Han, Z.-M., Li, M.-X., Jong, M. S., & Tsai, C.-C. (2019). Investigating students' interaction patterns and dynamic learning sentiments in online discussions. *Computers & Education, 140*, 103589. https://doi.org/10.1016/j.compedu.2019.05.015

Janssenswillen, G., Depaire, B., Swennen, M., Jans, M., & Vanhoof, K. (2019). bupaR: Enabling reproducible business process analysis. *Knowledge-Based Systems, 163*, 927–930. https://doi.org/10.1016/j.knosys.2018.10.018

Kaartinen, S., & Kumpulainen, K. (2002). Collaborative inquiry and the construction of explanations in the learning of science. *Learning and Instruction, 12*(2), 189–212. https://doi.org/10.1016/S0959-4752(01)00004-4

Lee, S. C., & Irving, K. E. (2018). Development of two-dimensional classroom discourse analysis tool (CDAT): Scientific reasoning and dialog patterns in the secondary science classes. *International Journal of STEM Education, 5*(1). https://doi.org/10.1186/s40594-018-0100-0

MacNeil, R., Suresh, T., & Park, E. (2024). Identifying common student errors through large language-model simulations of learner reasoning. *Computers & Education Open, 5*, 100143. https://doi.org/10.1016/j.caeo.2024.100143

Robertson, A. D., Scherr, R. E., & Hammer, D. (2015). Responsive teaching in science and mathematics. *Journal of Research in Science Teaching, 52*(8), 1072–1099. https://doi.org/10.1002/tea.21211

Scarlatos, A., Liu, N., Lee, J., Baraniuk, R., & Lan, A. (2025). Training LLM-based tutors to improve student learning outcomes in dialogues. *arXiv.org.* https://arxiv.org/abs/2503.06424

Shemwell, J. T., & Furtak, E. M. (2010). Science classroom discussion as scientific argumentation: A study of conceptually rich (and poor) student talk. *Educational Assessment, 15*(3-4), 222–250. https://doi.org/10.1080/10627197.2010.530563

Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*(1), 20–37. https://doi.org/10.1177/0022487108328155

Tanner, K. D. (2009). Talking to learn: Why biology students should be talking in classrooms and how to make it happen. *CBE—Life Sciences Education, 8*(2), 89–94. https://doi.org/10.1187/cbe.09-03-0021

Teo, P. (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. https://doi.org/10.1016/j.tate.2016.01.019

Van der Stuyf, R. R. (2002). Scaffolding as a teaching strategy. *Adolescent Learning and Development, 52*, 5–18. https://www.scirp.org/reference/referencespapers?referenceid=3825007

Xu, H., Liu, H., Leng, Y., & Ding, Y. (2025). Adapting to large language models: How insiders and outsiders reshape scientific knowledge production. *arXiv.org*. https://arxiv.org/abs/2505.12666

Zheng, H., Han, F., Huang, Y., Wu, Y., & Wu, X. (2025). Factors influencing behavioral intention to use e-learning in higher education during the COVID-19 pandemic: A meta-analytic review based on the UTAUT2 model. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-13299-2