education sciences

MDPI

Article

Using Artificial Intelligence to Support Peer-to-Peer Discussions in Science Classrooms

Kelly Billings [1,*], Hsin-Yi Chang [2], Jonathan M. Lim-Breitbart [1] and Marcia C. Linn [1]

[1] Berkeley School of Education, University of California Berkeley, Berkeley, CA 94720, USA; breity@berkeley.edu (J.M.L.-B.); mclinn@berkeley.edu (M.C.L.)
[2] Program of the Learning Sciences, National Taiwan Normal University, Taipei 10610, Taiwan; hychang@ntnu.edu.tw
* Correspondence: kelly_billings@berkeley.edu

Abstract: In successful peer discussions students respond to each other and benefit from supports that focus discussion on one another's ideas. We explore using artificial intelligence (AI) to form groups and guide peer discussion for grade 7 students. We use natural language processing (NLP) to identify student ideas in science explanations. The identified ideas, along with Knowledge Integration (KI) pedagogy, informed the design of a question bank to support students during the discussion. We compare groups formed by maximizing the variety of ideas among participants to randomly formed groups. We embedded the chat tool in an earth science unit and tested it in two classrooms at the same school. We report on the accuracy of the NLP idea detection, the impact of maximized versus random grouping, and the role of the question bank in focusing the discussion on student ideas. We found that the similarity of student ideas limited the value of maximizing idea variety and that the question bank facilitated students' use of knowledge integration processes.

Keywords: natural language processing; peer discussion; knowledge integration

check for updates

Citation: Billings, K.; Chang, H.-Y.; Lim-Breitbart, J.M.; Linn, M.C. Using Artificial Intelligence to Support Peer-to-Peer Discussions in Science Classrooms. *Educ. Sci.* **2024**, *14*, 1411. https://doi.org/10.3390/educsci14121411

Academic Editor: Randall S. Davies

Received: 29 August 2024
Revised: 29 October 2024
Accepted: 11 December 2024
Published: 23 December 2024

1. Introduction

Peer discussion of ideas has been shown to support learning [1,2]. Successful peer discussions require students to respond to each other, and scaffolds that enable participants to discuss the content of one another's ideas are crucial [3]. Creating real-time groups and providing questions to facilitate discussion in the classroom could be streamlined with technology. Artificial intelligence (AI) tools show promise in helping lessen a teacher's workload by providing support for tasks teachers have, including assessment and feedback [4–6]. AI can be further utilized to support teachers with other daily tasks in the classroom.

We developed and tested a natural language processing (NLP)-informed chat tool to help address these challenges. The chat tool uses NLP technology to create collaborative groups and assign adaptive question prompts. The technology does this by first detecting ideas in students' written science explanations to a prompt that asks students to use their knowledge of geology concepts, such as plate tectonics, to explain how Mt. Hood, a volcano located on the west coast of Oregon, was formed. These ideas were then used to group students based on two conditions for the chat activity: groups with as many different ideas from each other as possible or randomly assigned groups. We hypothesized that groups with differing ideas would have higher learning gains because their wealth of ideas would lead to more productive conversations that would support students in distinguishing between the many ideas they hold about plate tectonics. The ideas identified through NLP were also used to assign adaptive question prompts based on the ideas held by students in the chat group for all groups. These adaptive prompts were designed using the Knowledge Integration framework (KI) [7] to elicit students' ideas about plate tectonics or support students in distinguishing between the different ideas they hold about plate tectonics. Similar to other KI-informed supports, we hypothesized that these prompts would facilitate knowledge integration while students discussed their ideas with their peers [8–10].

We contribute to a growing area of research around the use of AI tools in the classroom. Recent research has highlighted both the potential and limitations of using AI-informed instruction in classrooms [8,11]. AI is a powerful tool that can help reduce teachers' workloads and provide insights into student learning for teachers [12–14]; however, teachers are still unsure of how to utilize AI in their practice and how useful AI technologies can be in the classroom [15,16]. We contribute to a growing need for classroom-based research on the use of AI to support learning by utilizing NLP to support students' peer discussion of ideas, which has not been conducted previously. We document the success of our tool, implications for future research, and ways that this work can inform future integration of AI in the classroom in partnership with teachers. We investigate:

1. How accurately did the NLP model detect students' ideas?
2. How effective was NLP in grouping students with disparate ideas? Moreover, how did NLP-formed groups compare to randomly-formed groups on progress in KI?
3. How did students' use of the KI-inspired adaptive prompts assigned by the NLP impact students' progress in sorting out ideas, forming KI explanations, and using KI processes in their chat conversations?

2. Literature Review and Theoretical Framework

2.1. Use of NLP Technology in Science Education and Open Questions

Rapid developments in AI techniques have opened the door for AI-informed educational interventions in recent years. The ways AI technologies are used are wide in scope, ranging across grade levels and subject areas [5,6,17]. In secondary STEM education, some applications of AI technologies include scoring students' science explanations in English [18–23] and other languages [24,25]. This work shows that AI tools can score or code student responses with satisfactory accuracy as compared to human scorers, which is a key first step in creating AI that is useful for teachers in classroom contexts [22]. Other studies focus on generating automated guidance for students similar to that of their teachers [26,27], using generative AI [28], or on supporting student learning through intelligent tutoring systems or chatbots [29,30]. These tools can provide students with personalized feedback, taking some of the weight off of educators for large classes or providing more specific instruction in large online classes where direct interactions with educators are difficult. These technologies show promise in supporting student learning and personalizing instruction.

Natural language processing (NLP), which we utilize in this study, has become a commonly utilized AI technology in education settings and shows promise in assisting automated assessment and identification of student ideas in written work [31–33]. NLP can provide insights into students' thinking more quickly than if a teacher were reading students' explanations on their own. This can be leveraged to support student learning. For example, Gerard et al. [33] found that NLP idea detection and automated scoring can quickly show how students are making sense of multiple ideas when learning. Insights, such as the ideas students hold or how their ideas change over time, can be provided to teachers to inform their future teaching [12–14]. NLP idea detection and automated scoring can also inform technologies such as AI dialogues and adaptive guidance that are responsive to the ideas students hold [4,33,34]. These technologies show promise in supporting teacher responsiveness by providing personalized instruction to students to support their STEM learning.

However, there are still challenges and open questions about using AI in the classroom. While AI shows promise in detecting ideas, the accuracy of these technologies still needs improvement [33]. While NLP has been utilized in facilitating student-to-bot interactions, directing students to particular online activities that are responsive to their explanations, and providing teachers additional information on student learning, using NLP to scaffold student-to-student interactions is a less explored area of NLP use in the classroom. We seek to utilize NLP to facilitate students' collaborative discussions, addressing this major gap in current research.

2.2. Scaffolding Group Discussions and Online Chat Environments

Leveraging collaborative groups can be a powerful way to support students to elaborate and clarify their ideas [2,35]. Group work is a common structure implemented by teachers in the classroom and often occurs in face-to-face contexts; however, with the onset of the COVID-19 pandemic and online learning, creating online-based collaborative structures has proved crucial, and many online collaborative tools have continued to be leveraged even after schools have moved back to in-person learning. We leverage NLP idea detection to support students' online chat conversations. Several studies on online chat environments compare face-to-face collaboration and online chat collaboration. For example, Jonassen and Kwon [36] found that there were no statistically significant differences in how face-to-face groups and online chat groups engage with one another when solving STEM problem-solving activities. Sins et al. [37] found similar results when comparing face-to-face and online chat collaboration for modeling tasks. These results indicate that certain chat-facilitated collaboration activities can be as effective as face-to-face collaboration.

However, creating collaborative groups and simply instructing students to talk about their ideas does not necessarily lead to productive, equitable conversations [38,39]. Teachers facilitate collaboration by providing scaffolds to support students' interaction with one another. These include guidance prompts [40], sentence frames [41], and discussion facilitation protocols or norms [42]. These tools help students to have productive conversations about their ideas, allowing them to clarify and elaborate on their ideas [2,35]. Similarly to face-to-face discussions, online discussions can also benefit from scaffolding. For example, Lazonder et al. [43] investigated how students utilized sentence starters in an online chat environment while working on an ecology task. The authors examined students' online interactions and used students' language to create sentence frames that mimicked the language they used in their conversations. Though these sentence frames were created with the intention to be both supportive and familiar to students, they were not well utilized by students in the online chat environment. This indicates that there is room for clarifying the criteria for useful scaffolds (such as sentence frames) in online chat environments.

We seek to utilize NLP idea detection to improve the facilitation and scaffolding of online chat environments with the hope that they can provide complementary support to face-to-face teacher-facilitated environments. We use ideas identified by the NLP to group students for the chat activity. Strategic grouping strategies are often utilized by teachers to help support student-to-student discussions; however, creating real-time groups in the classroom can be challenging and time-consuming and can disrupt the flow of a lesson. We leverage NLP to extend these teacher strategies. We also utilize NLP-informed idea detection to provide adaptive question bank prompts for students to utilize in their conversations. Though Lazonder et al. [43] found that students did not utilize provided sentence starters, we posit that adaptive question bank prompts tailored to students' ideas are more useful and better utilized by the students during their chat conversations.

2.3. Theoretical Framework and Implications for NLP Technology

Knowledge Integration (KI) informs the design of the curricular unit, NLP features, and online chat tool. KI supports students to integrate their varied, disconnected ideas about a scientific phenomenon to form a coherent understanding. KI leverages students' initial ideas from their everyday experiences and prior learning to support them in discovering and connecting new ideas to their initial understanding. This helps them form coherent explanations of scientific phenomena. The KI framework has had a long history of being leveraged to support students' inquiry learning [7] and the integration of technology into the STEM curriculum [44]. The framework informs both the design of the curriculum and the design of assessments. These qualities make KI a promising framework to lean on when introducing NLP-informed technologies into classroom learning environments.

KI provides support through a four-step cycle. This cycle can inform the design of learning environments and learning scaffolds. The cycle includes eliciting students' initial ideas, providing opportunities for students to discover new ideas to add to their repertoires, supporting students to distinguish among the different ideas, and finally, providing opportunities for students to reflect on and connect ideas [7]. For example, in the plate tectonics unit used in this study, students' initial ideas about what types of plate movement occur at certain plate boundaries are elicited. Students then discover how plate density determines how different plates interact and which geologic features occur at certain plate boundaries. Students then use the density factors and geologic features to distinguish how plates are moving in an interactive computer model. Finally, students connect their ideas to explain how Mt. Hood formed.

KI has been leveraged to support the integration of NLP technologies previously, including validating the accuracy of using NLP for automated KI scoring [41,45], informing adaptive, automated guidance [8,10,40], and automated thought-partners [4,33,34,46]. These technologies leverage NLP to detect student ideas, which we also perform in this study; however, the guidance or thought-buddy prompts are not written by an AI model. They are created by teachers and researchers, informed by the KI framework, and assigned using NLP idea detection. For example, prompts can be designed to elicit new or elaborated ideas from students. Other prompts direct students back to specific points in a unit to discover new ideas that would help them extend their thinking. Prompts can also be designed to support students in distinguishing between different ideas they hold about a topic. This results in adaptive, NLP-informed tools that provide similar support to in-the-moment scaffolds and interventions created by teachers in the classroom.

Studies have shown promise in this approach. For example, Gerard et al. [9] showed that adaptive, KI-inspired guidance informed by students' automated KI scores was more effective in supporting students' revisions of science explanations than generic guidance. Tansomboon et al. [10] built on this work by examining what types of adaptive guidance are more effective in supporting students to revise their science explanations. The authors compared two types of KI-informed guidance: revisiting evidence and planning changes to writing. Both types of guidance had significant learning gains, though there was no significant difference between the types of guidance.

Automated thought partners informed by NLP-assisted idea detection and KI have been shown to support students' integration of their ideas about climate change [47], thermodynamics [34], and genetics [33,46]. Each guidance prompt posed by the thought-partner to the student is informed by Knowledge Integration pedagogy. The findings from one study found that certain prompts are especially successful at eliciting more ideas about climate change from students during the chatbot conversation [47]. The findings from these studies indicate that the NLP-informed chatbot increases students' knowledge integration levels upon revision, indicating the promise of pairing NLP idea detection with pedagogically informed prompts to support student learning and revision of science explanations.

We take a similar approach to the work in this study. We leverage NLP to identify ideas and use KI as a pedagogical framework to create scaffolds (such as strategic grouping and question-bank prompts) that mimic a teacher's toolbox to support students in discussing their ideas with one another. While merging KI and NLP technologies has been deeply explored as a framework to inform students' independent learning (via computer guidance or adaptive thought-partner conversations), it has yet to be utilized in ways that scaffold group discussions. This paper seeks to address this new space, building on KI's initial promise as a framework to inform the implementation of NLP tools in the classroom.

3. Methods

We leverage design based research methodology to explore the impact of the chat tool on student learning [48]. Our research uses both quantitative and qualitative methodologies to assess the effectiveness of our designed intervention (in this case, the chat tool) in supporting student learning. We also use results from these methods to identify the next steps for improving our design.

3.1. Participants

The chat tool was embedded in a plate tectonics unit that was used in two grade 7 science classrooms (2 teachers and 256 students) in a school located in a metropolitan area of the western United States. The school has roughly 70% minority enrollment, and 26% of students receive free or reduced-price lunch. One of the teachers has used the unit over multiple years and has helped to refine the unit. The other teacher is her partner teacher, who is new to the school. Students worked in pairs (with some groups of one or three) during the duration of the unit, sharing one computer between them (N = 131 pairs). As students worked, they could talk about their ideas and take turns typing into the online learning environment using their shared computer. During the chat activity, each pair was grouped with another pair, resulting in a conversation group of 4 to 5 students.

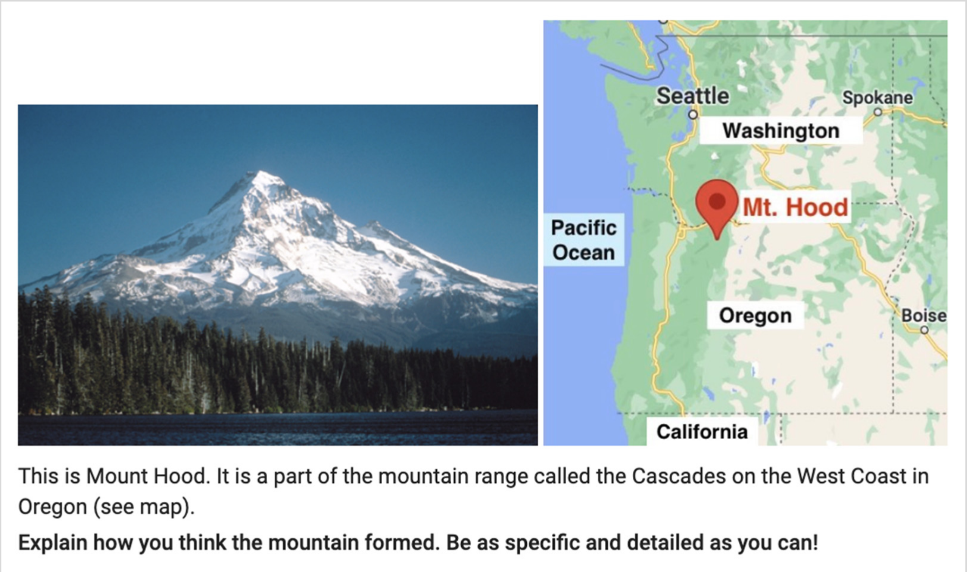
We removed any student pairs that did not answer both the pre and post-chat explanations (see below). Then, we removed conversations and the respective pairs who did not engage in the chat activity. These included conversations where no pair engaged or where only one pair engaged and their conversation partners did not respond. Our final data corpus had 102 pairs of students and 56 conversation groups.

3.2. Curriculum and Embedded Assessments

The NLP technology at the center of this study was incorporated into the existing plate tectonics unit to help facilitate an online peer-to-peer chat activity. The unit is hosted on the Web-Based Inquiry Science Environment "https://wise.berkeley.edu (accessed on 18 December 2024)" and is informed by Knowledge Integration [7] pedagogy. This unit and the refinement of several unit activities have been studied previously [12,13,49]. The unit supports students in integrating their ideas about earth science concepts such as plate tectonics, convection, and slab pull to explain how different geologic features formed, including rift valleys, deep sea trenches, and volcanoes such as Mt. Hood. Students engage with interactive maps and computer models to support them in integrating their ideas. This unit has been extensively customized by one of the participating teachers.

The peer-to-peer chat activity occurs at the end of the first lesson, which introduces students to concepts about plate boundaries, including the types of boundaries that exist and the mechanisms that drive plate interactions (such as density). By the end of this lesson students have discovered many ideas about plate interactions that can help them explain how a mountain has formed. The chat activity was added to the end of this section in hopes that it would support students to distinguish among the ideas they had discovered during the lesson in the chat with their peers.

An assessment item was embedded into the unit. The item asks pairs to explain how they think the volcano called Mt. Hood formed. The item includes a picture of the mountain and a screenshot of a map showing the mountain's location in Oregon, other states adjacent to Oregon, and the Pacific Ocean (Figure 1). This assessment item is scored along a KI rubric (Table 1) and an idea rubric (Table 2). The KI rubric scores students on a scale of 1–5 and rewards students for connecting ideas about a scientific phenomenon. The idea rubric contains fourteen different ideas that range from off-topic, irrelevant earth science (erosion) or vague ideas (plates move on earth) to increasingly more specific mechanistic ideas about plate tectonics (such as how density drives the creation of subduction zones) to explain the formation of Mt. Hood.



This is Mount Hood. It is a part of the mountain range called the Cascades on the West Coast in Oregon (see map).

Explain how you think the mountain formed. Be as specific and detailed as you can!

Figure 1. Mt. Hood explanation item.

Table 1. Mt. Hood Knowledge Integration rubric.

KI Level	Description	Example
1	Off task—student writes an answer, but it is not related to the question being asked.	IDK We do not need to revise our explanation
2	Vague or non-normative ideas. Includes other non-target earth science topics.	Over millions of years, these mountains have probably been formed by erosion. I think Mt. Hood was formed by avalanches and old volcano eruptions.
3	Partial link or one complete idea in isolation.	Mt. Hood was created from two continental tectonic plates pushing against each other. I think mount hood formed when an oceanic and continental plate converged.
4	One full link between ideas.	We think Mt. Hood formed by one oceanic plate and one continental plate. We think that the oceanic plate got subducted and that formed the mountain.
5	Two full links between ideas.	For a volcano like Mt. Hood to form, two plates move towards each other at a convergent boundary. The plates move because of convection currents inside the mantle. These currents push plates together, making the convergent boundary. One of the plates subducts under the other plate, and the rocks turn into magma. The magma rises and forms a volcano.

Table 2. Final Mt. Hood idea rubric and F scores for each idea. Specific ideas are underlined in student examples. Off-topic, divergent boundaries and repeats of the question were dropped because the F-score was 0.32 or below. Underlined text in the examples column indicate the section of the example that the idea applies to.

Idea Type	Name	F Score	Examples	Adaptive Question Bank Prompts
Specific Mechanistic	Correct Density	0.75	Mt. Hood was formed by the tectonic plates, the oceanic and continental crust collide, forming a convergent boundary. The oceanic crust is denser than the continental crust, so it subducts under the continental crust.	How does convection affect plate movement where Mt. Hood is located? Mt. Hood is on a plate boundary; what is happening at that location specifically.
	Subduction	0.70	Mt. Hood was formed when the oceanic plate went under the continental plate.	How do the characteristics of different plates lead to subduction? What makes one plate go under the other?

Table 2. *Cont.*

Idea Type	Name	F Score	Examples	Adaptive Question Bank Prompts
Specific Mechanistic	Continental/ Oceanic plates	0.74	For a volcano to form, like Mt. Hood, an oceanic crust and a continental crust would move to each other because of the tectonic plate movement.	When continental and oceanic plates collide, a volcano often forms. Why? How does a collision between continental and oceanic plates create a mountain like Mt. Hood?
	Convection	0.81	Mt. Hood was caused by convection currents moving towards each other. What happened was these currents were caused by magma moving around in Earth's mantle due to the temperature difference of the crust and core. Plates move because of convection currents in the mantle.	How does convection affect plate movement where Mt. Hood is located? Mt. Hood is on a plate boundary; what is happening at that location specifically?
General Mechanistic	Convergent	0.79	Mt. Hood was made when two plates collided.	What types of plates are colliding? What happens to the two plates when they collide?
	Plates Move	0.66	Mt. Hood is a part of the Cascades range, and it was formed by the moving of 2 tectonic plates.	What causes plates to move? Tell me more about what kinds of plates these are and how they moved.
	Volcano	0.63	I think that the mountains are formed by volcanoes that erupt underground.	What caused a volcano to form at this location? What processes in the earth cause the volcano to form?
Incorrect Mechanistic	Incorrect or general Density	0.47	When the Tectonic plates collide, the crash locations rise, creating a mountain. Since neither plate is denser and since they are convergent, the plates have no choice but to right. Density affects how plates move.	How does density have an effect on how plates interact with one another? Mt. Hood is next to the Pacific Ocean. What does this information tell you about how density impacts the formation of Mt. Hood?
	Continental/ continental plates	0.59	a mountain is formed when two pieces of continental crust collide, and then that makes the land go up which forms a mountain.	Mt. Hood is near the Pacific Ocean. How does this feature affect plate types? Mt. Hood is near the Pacific Ocean. What does this information tell you about how it was formed?
Other Earth Science	Earthquakes	0.7	Long ago, in a state far, far away, there was a huge earthquake! The cause of this earthquake is the tectonic plates in the mantle of the earth rub together, which then sprouts up from the surface.	What is causing the earthquakes to happen?
	Erosion	0.52	I think it might've been formed by erosion because the sides of the mountain are kind of going inwards. So I think the erosion shaped the mountain and made it form like that.	How did Mt. Hood get to be higher than the surrounding area before it eroded? How did the mountain get there for it to be eroded away?
	Rocks and snow pile up	0.58	I think Mt. Hood or mountain from the terrain of sun, rain, and possibly the waves that collided and hit rocks piling up to make a mountain.	How do you think Mt. Hood's formation connects with what is happening inside the earth? Please explain in more detail how a mountain as large as Mt. Hood could have formed in this specific location.

3.3. NLP Technology and Chat Interface

The KI rubric and the idea rubrics were used to create a KI scoring model and an idea detection model. The two models were trained on the same data set of 1170 student responses to the Mt. Hood assessment item collected from previous research studies. For more information on how the KI model was created, see [49]. The NLP idea detection model uses a token classification approach [50]. After creating the initial rubric, two researchers performed two rounds of agreement coding, using 10% of data for each agreement round, and achieved a Cohen's Kappa above 0.85. These remaining data were split in half and coded by the two researchers. The resulting idea model had an overall micro-averaged F-score of 0.7206. Six ideas had an F-score of 0.7 or above. Five ideas had an F-score of 0.45–0.7. One idea had an F-score of 0.32 (off-topic), and two ideas had an F-score of 0, as they were not found in a high enough frequency to train the model (divergent boundary and repeats the question). These three ideas were dropped from the idea detection model because the model could not accurately identify these ideas.

The idea detection model was then used to help facilitate the peer-to-peer chat activity. Student pairs answered the Mt. Hood explanation item. After answering the pre-chat explanation, students pairs were put into conversations of two to three pairs. Two conditions were used to determine conversations: NLP-informed and random (Figure 2). In the NLP-informed condition, the idea detection model identified students' ideas in their pre-chat explanation and then grouped pairs into conversations to maximize the number of different ideas among them. For example, if the idea detection model detected convergent, subduction, and volcano for one pair and convection and plates move for another pair, the condition would ideally group these pairs because they have five different ideas between them. The random condition simply grouped pairs at random, uninformed by the NLP idea detection. Different periods were assigned the NLP-informed condition (five periods) or the random condition (five periods).

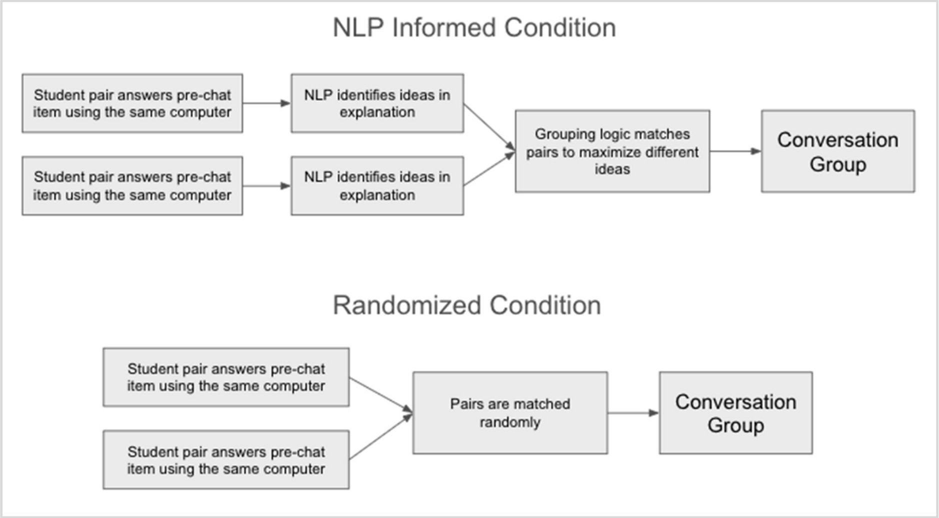
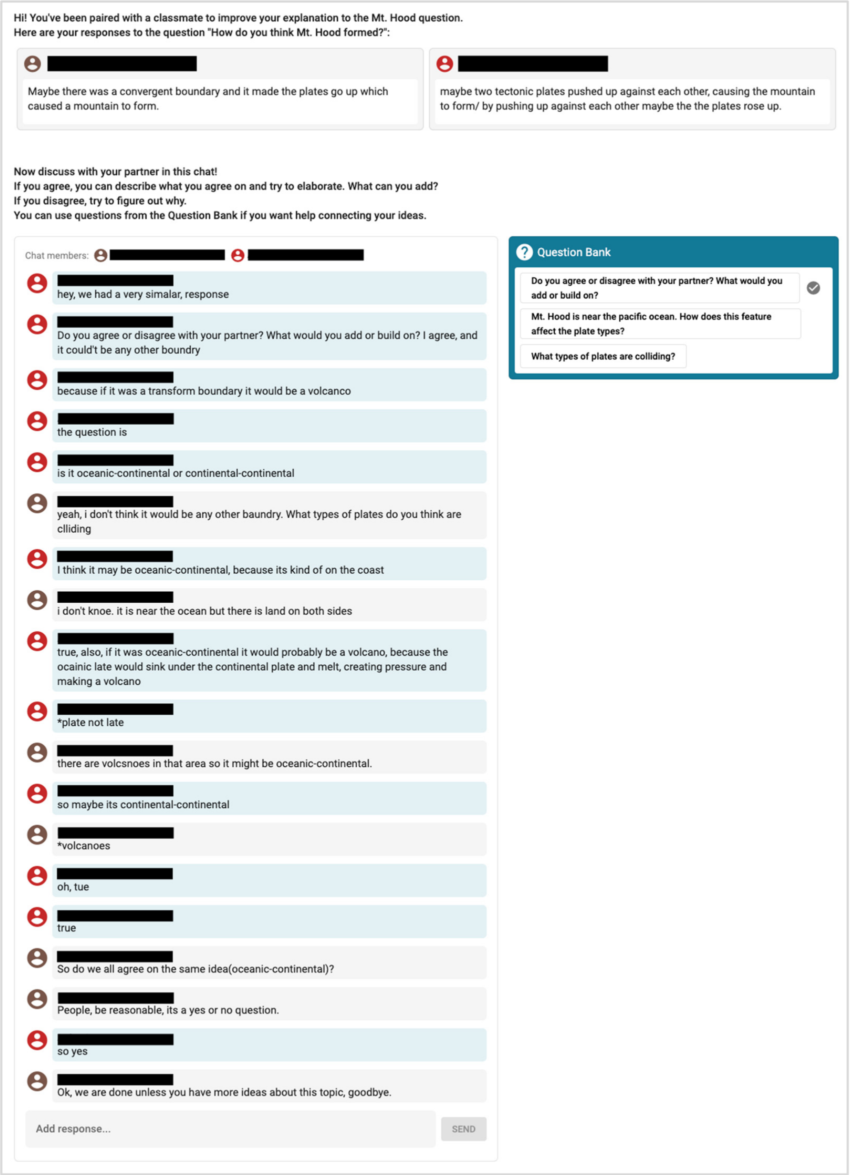

NLP Informed Condition
Student pair answers pre-chat item using the same computer
NLP identifies ideas in explanation
Student pair answers pre-chat item using the same computer
NLP identifies ideas in explanation
Grouping logic matches pairs to maximize different ideas
Conversation Group
Randomized Condition
Student pair answers pre-chat item using the same computer
Student pair answers pre-chat item using the same computer
Pairs are matched randomly
Conversation Group

Figure 2. Chat grouping logic for the NLP-informed condition and randomized condition.

After the conversation groups were created, students were prompted to chat in their conversation groups. The chat screen contains each pair's explanation so that they are accessible to the conversation group as a starting point for discussion. The chat interface, which is set up similarly to online social media chats students might be familiar with, makes it clear to students that they are chatting with peers rather than an AI chatbot or tutor. A list of adaptive question bank prompts next to the chat interface helps facilitate students' discussion in the chat (Figure 3). Students could add these prompts into the chat interface to use them in their conversation. All students were provided a generic question bank prompt: Do you agree or disagree with your partner? What would you add or build on? Groups were also provided two adaptive question bank prompts based on all the ideas present in both pairs' explanations (Table 2). All prompts were informed by the knowledge integration framework [47] and were designed to support students to think more deeply about their ideas and clarify their thinking. Some prompts were designed to elicit more ideas from students (What types of plates are colliding?) and others were designed to support students to distinguish between their ideas about plate tectonics (What characteristics of different plate types lead to subduction?). The two adaptive questions were assigned according to a ranking logic created by researchers that prioritized responding to certain ideas over others if more than one idea was present [34]. For example, if the NLP identified a subduction idea and a plate moves idea in a pair's response, an adaptive question bank prompt associated with subduction would be provided because it is ranked higher in the ranking logic. After completing the chat, pairs were prompted to revise their answer to the Mt. Hood explanation item. This revised explanation is the post-chat explanation.



Hi! You've been paired with a classmate to improve your explanation to the Mt. Hood question.
Here are your responses to the question "How do you think Mt. Hood formed?":

[redacted]
Maybe there was a convergent boundary and it made the plates go up which caused a mountain to form.

[redacted]
maybe two tectonic plates pushed up against each other, causing the mountain to form/ by pushing up against each other maybe the the plates rose up.

Now discuss with your partner in this chat!
If you agree, you can describe what you agree on and try to elaborate. What can you add?
If you disagree, try to figure out why.
You can use questions from the Question Bank if you want help connecting your ideas.

Chat members: [redacted] [redacted]

[redacted]
hey, we had a very simalar, response

[redacted]
Do you agree or disagree with your partner? What would you add or build on? I agree, and it could't be any other boundry

[redacted]
because if it was a transform boundary it would be a volcanco

[redacted]
the question is

[redacted]
is it oceanic-continental or continental-continental

[redacted]
yeah, i don't think it would be any other baundry. What types of plates do you think are clliding

[redacted]
I think it may be oceanic-continental, because its kind of on the coast

[redacted]
i don't knoe. it is near the ocean but there is land on both sides

[redacted]
true, also, if it was oceanic-continental it would probably be a volcano, because the ocainic late would sink under the continental plate and melt, creating pressure and making a volcano

[redacted]
*plate not late

[redacted]
there are volcsnoes in that area so it might be oceanic-continental.

[redacted]
so maybe its continental-continental

[redacted]
*volcanoes

[redacted]
oh, tue

[redacted]
true

[redacted]
So do we all agree on the same idea(oceanic-continental)?

[redacted]
People, be reasonable, its a yes or no question.

[redacted]
so yes

[redacted]
Ok, we are done unless you have more ideas about this topic, goodbye.

Question Bank

- Do you agree or disagree with your partner? What would you add or build on? ✓
- Mt. Hood is near the pacific ocean. How does this feature affect the plate types?
- What types of plates are colliding?

Figure 3. Chat interface and sample student discussion. Students' initial answers to the Mt. Hood assessment item are displayed above the chat environment. Question bank prompts are displayed next to the chat environment, and students can select questions they want to add to the chat. *Students used * to indicate corrections in spelling*.

3.4. Data Sources and Analysis

3.4.1. Analysis and Accuracy of the NLP and Group Creation

A total of 100 randomly selected responses from these classroom data were human-coded by one of the researchers who created the idea rubric and coded the initial data set to train the NLP model. We assessed the accuracy of the human–machine agreement on this subset of these data using a response-level micro-averaged F1 score.

To assess the accuracy of the NLP-informed grouping strategy, we identified the ideas present in each pair's pre-chat response and the conversations created in the NLP-informed grouping condition and the randomized grouping condition. We then identified how many different ideas were present in each conversation to see if the NLP-informed condition created more conversations with more different (2+) ideas than the randomized condition.

3.4.2. Analysis of Pre and Post-Chat Responses

A human coder who was trained on the Mt. Hood KI rubric previously scored the pre and post-chat responses along the KI rubric. Two additional human scorers coded pre and post-chat responses for ideas. The human coders completed two rounds of agreement using 5% of these data for each round. They then split these remaining data and coded for ideas.

After removing pairs who did not answer the pre-test, post-test, or engage in the chat, there were 52 pairs in the NLP-informed condition and 50 pairs in the randomized condition. To assess for changes in KI between the NLP-informed and randomized grouping conditions we used ANCOVA to control for pre-chat scores and test for statistically significant differences in KI scores post-test between conditions. We used a *t*-test to analyze if there were any significant differences in changes in the average number of ideas between pre and post-chat explanations between the two conditions. Finally, we used a chi-square test to see if there were any statistically significant changes in the number of specific mechanistic ideas between pre and post-chat explanations.

3.4.3. Analysis of Student Group's Chat Conversations

We coded students' chat conversations to analyze how conversation groups engaged in KI processes and utilized the adaptive question bank prompts in the chat. Chats were first segmented into "turns" which were defined as a unit of talk in the form of a text message that was produced by one pair. Repeated messages from the same pair (for example, a student spamming another student's name in the chat) were considered one turn. The first three authors created a rubric to code the turns. This rubric contained several different types of processes and codes (including cognitive/KI processes, regulatory processes, social processes, and off-task behavior). Each turn in the chat was coded according to a coding rubric. The first three authors used an initial codebook on 10% of these data, then met to discuss areas of agreement and disagreement. We updated the codebook to address issues that arose during the conversation. Then, the three authors coded an additional 10% of these data, discussing and revising the code book as needed. The first author then coded the remaining chat conversations. For this analysis, we focus only on the KI/cognitive processes (Table 3).

Table 3. Chat codebook for KI/cognitive processes.

Code	Definition	Example
KI—Elicit	Students' conversion prompts their peers to articulate and explicate their existing ideas.	Can you explain your response? What does your answer mean? The question is, is it oceanic–continental or continental-continental?

Table 3. *Cont.*

Code	Definition	Example
KI—Add/Discover	Students' conversation shows that they encounter new ideas and add to existing repertoire. Students Share a new idea in the chat that has not been discussed previously or elaborate on their previous idea. Respond to an eliciting question by adding ideas.	I think it may be oceanic–continental because it's kind of on the coast. I believe that some of the plates will be oceanic plates, and once oceanic and continental plates push against each other, there will be volcanoes.
KI—Distinguish	Students compare and evaluate different ideas. Disagreement about ideas	I like how you added that you said the continental plates with low density added up. In response to an idea shared in a previous turn: It is about in the middle of Oregon, so there is no way it is oceanic and continental.
KI—Connect/reflect	Students connect their ideas with other peers' ideas and/or reflect on elaboration and connection; students connect ideas to answer QE questions.	Do you agree or disagree with your partner? What would you add or build on? I agree, and it couldn't be any other boundary. (Question bank followed by connect/reflect) I said it was continental/continental; you said it was continental/oceanic. After reading yours, I think yours is correct.
Use of Question Bank	Students use the question bank in their chat conversations.	

3.4.4. Analysis of Students' Use of Adaptive Question Bank Prompts and Impact on KI and Ideas

We then identified which conversations and which participating pairs used the adaptive question bank prompts and which did not. We separated student data into these two groups to see if the use of the adaptive question bank had any impact on KI scores and the use of KI in the chat. We used chi-square tests to look for statistically significant differences in how each group used KI processes. We controlled for pre-chat scores using ANCOVA to see if there were any statistically significant differences in KI scores post-test between pairs that used the adaptive question bank in their conversations and pairs that did not. We used a *t*-test to analyze if there were any significant differences in changes in the average number of ideas between pre and post-chat explanations between the two groups. Finally, we used a chi-square test to see if there were any statistically significant changes in the number of specific mechanistic ideas between pre and post-chat explanations.

4. Results

4.1. Research Question 1

The agreement between the idea detection model and the human scorer who created the rubric on the 100 randomly selected responses from these classroom data was reasonable. The overall F1 score for the model was 0.755, and the model achieved good precision (0.712) and recall (0.804). F1 scores were similar or higher to these training data for eight ideas. Two ideas (Isolated volcano and Incorrect density as the mechanism for subduction) had lower F1 scores than these training data (0.2–0.45 range) indicating it was challenging for the idea detection model to detect those ideas. Two ideas (convection and rocks and snow pile-up) did not occur in the data sample, so they did not have an F1 score. Overall, the idea detection model was reasonably successful at identifying students' ideas. This indicates that the NLP model identified ideas with enough accuracy to use the model to support a classroom task, such as creating groups and assigning adaptive question bank prompts.

4.2. Research Question 2

While identifying ideas in students' explanations was successful, creating conversations based on similar or different ideas proved more challenging. The grouping logic created 56 total conversations for the chat activity. A total of 28 were in the NLP-informed condition, and 28 were in the randomized condition. A total of 54% of NLP-informed (optimizing for more different ideas) conversations had two or more different ideas between pairs. Similarly, 45% of the randomized conversations had two or more different ideas. This indicated that this grouping strategy might have been challenging to execute in real-time in the classroom.

We then sought to understand why creating NLP-informed conversations proved challenging. One possible reason for this could be that students' explanations had similar ideas, so there were not many different ideas for the grouping strategy to use to create student groups. To test this hypothesis, we further analyzed the ideas identified by the NLP to see if there were enough different ideas in students' explanations to make conversations with different ideas.

To conduct this analysis, we analyzed all pre-chat explanations (115) regardless of whether the pair participated in the chat or completed the post-chat explanation. We chose to perform this for our analysis because the grouping logic used all 115 of these explanations to create conversations. On average, pre-chat explanations had 1.82 different ideas. 80.8% of pre-chat explanations had one to two ideas. Of the remaining explanations that had more than two ideas (19.2%), only one explanation had four ideas. No explanations had more than four ideas. This indicates that, overall, each pair's explanations did not contain many distinct ideas, making grouping by maximizing the number of ideas more challenging. Many students also had the same ideas in their pre-chat explanations (Table 4). Convergent, the idea that plates move towards each other, was found in 89.6% of responses, making it highly likely that groups would share this idea. Seven of the twelve ideas were found in zero to five percent of explanations. These results show that, in addition to there being few ideas in the data corpus, there was also a homogeneity in the types of ideas that occurred in pre-chat explanations. This likely made it challenging to create NLP-informed conversations that maximized the number of different ideas amongst pairs. This resulted in the NLP-informed grouping strategy making many conversations that had one or zero different ideas.

Table 4. Pair's ideas pre-chat explanation.

Idea Name	Percentage (Out of N = 115)
Incorrect or general density	10.4
Correct density	3.5
Subduction	13.9
Continental-oceanic or oceanic-oceanic	21.7
Continental-Continental	35.6
Convergent	89.6
Convection	0
Plates move	0
Isolated Mountain/Volcano	4.3
Earthquakes	0
Erosion	1.7
Rocks and snow pile up	1.7

We also considered additional classroom contexts that could make utilizing NLP-idea detection to group students challenging. One factor that likely contributed to making the grouping strategy challenging was the difficulty in timing the activity to match pairs into conversation groups. In order to optimally make groups that maximize differences in ideas among pairs, all student pairs would ideally answer the Mt. Hood prompt and then wait until all other pairs have answered the prompt before moving to the chat step. This proved challenging, as real-life classroom factors such as absent students and classroom management issues can make it hard to coordinate students so that everyone is submitting around the same time. These challenges highlight the messiness of using NLP technology in the classroom. Though a powerful tool, NLP might not be flexible enough to adjust to these challenges without adjustments to how we apply the technology.

Though we acknowledge challenges in making the NLP-informed groups, we still sought to understand if the groups that were made had any impact on students' KI progress or the ideas they expressed. When controlling for pre-chat KI scores, we found that the grouping condition factor (NLP-informed or randomized) had no significant effect on post-chat KI scores (F = 0.722, *p* = 0.394). This indicated that the different conditions did not have different impacts on the KI score.

We examined whether the conditions had an impact on the number of ideas expressed by students from pre-chat to post-chat explanations. Both conditions started at a similar mean number of ideas, with NLP-informed starting at 1.84 ideas on average (SD = 0.766) and randomized starting at 1.86 ideas on average (SD = 0.736). Both conditions averaged 1.98 ideas for the post-chat response (Max SD = 0.736, randomized SD = 1.000). Condition factors did not have an impact on the change in the number of ideas.

We then examined whether there were differences in changes in ideas that show a more sophisticated understanding of plate tectonics between conditions. These ideas were labeled as specific mechanistic ideas in the idea rubric. At the pre-chat, there were 15 specific mechanistic ideas across all pair explanations from the NLP-informed condition. At the post-chat, there were 28 mechanistic ideas across all pair explanations. This increase in specific mechanistic ideas was statistically significant (*p* = 0.047 *). In the randomized condition there were 27 specific mechanistic ideas at the pre-chat and 39 specific mechanistic ideas at the post-chat. This increase was not statistically significant (*p* = 0.140). This suggests that the condition factor had an impact on the addition of specific mechanistic ideas, which demonstrate a more complex understanding of plate tectonics.

Though we see this difference in specific mechanistic ideas, we do not see any other statistically significant differences by grouping factor. This indicates that, overall, the grouping factor did not have an impact on students' knowledge integration or change in ideas between pre-chat and post-chat responses. It showed an impact in facilitating student development of mechanistic ideas. Considering the challenges identified in creating the conversations themselves, there are potential areas of growth in using NLP technology to create real-time student groups. We elaborate more on this in the discussion.

4.3. Research Question 3

NLP technology was also used to facilitate the assignment of question bank prompts. In this section, we examine the use and success of those prompts in supporting student learning. Of the 56 total conversations, 33 (59%) used at least one prompting question from the adaptive question bank. All of these conversations used at least one NLP-informed adaptive question bank prompt tailored to the student's ideas, and 19 of the conversations used the generic question in addition to one of the NLP-informed adaptive questions. No conversations groups used only generic questions. This indicates that the KI-inspired adaptive question bank prompts were more compelling to students than the generic prompts. 13 of 28 (46%) conversations in the NLP-informed conditions used the question bank, and 20 of the 28 (71%) conversations in the randomized condition used the question bank.

To measure the question bank's impact on student learning, we analyzed whether there were differences in the KI score from pre-chat to post-chat explanations between pairs that were in conversations where the adaptive question bank was used versus pairs that were in conversations where the adaptive question bank was not used. When controlling for KI pre-chat scores, we did not find any statistically significant differences in KI post-chat scores between the two conversation groups [F = 2.302, *p* = 0.132]. The mean number of ideas from pre-chat to post-chat explanations did not indicate major differences between the two groups. Pairs that used the question bank had 1.85 ideas on average (SD = 0.783) pre-chat and an average of 1.93 ideas (SD = 1.0) post-chat. Those who did not use the question bank had, on average, 1.86 ideas pre-chat (SD = 0.710) and 2.09 ideas on average post-chat (SD = 8.68). These are not statistically significant differences.

We examined differences in specific mechanistic ideas from pre-chat to post-chat explanations between pairs that used the question bank and pairs that did not use the question bank. Both student pairs who used the question bank prompts and pairs who did not use the question bank prompts had statistically significant increases in specific mechanistic ideas. There were 23 specific mechanistic ideas pre-chat and 41 specific mechanistic ideas post-chat for those who used the question bank prompts (p = 0.024). There were 19 specific mechanistic ideas pre-chat and 26 specific mechanistic ideas post-chat for those who did not use the question bank (p = 0.011). While both improved, this suggests that using the question bank or not did not have an impact on student use of specific mechanistic ideas.

Results in idea changes and KI changes indicate that, while many students used the question bank prompts, using these prompts did not result in any major differences as compared to those who did not use the prompts from pre-chat to post-chat explanations. Though we did not see any major differences, we were curious if using the question bank had any impact on pairs' KI processes while discussing ideas with one another in the chat. We next move to examining students' chat conversations more closely.

Of the 33 groups that used at least one question bank prompt, 31 of the 33 conversations (94%) engaged in at least one KI process necessary for developing an integrated understanding during the chat activity [7]. 19 of the 23 conversations that did not use the prompts (82.6%) engaged in at least one KI process; however, these overall differences in KI processes were not statistically significant (Chi-Square, X^2 = 0.224, p = 0.638). On average, conversations that used the adaptive question bank engaged in 2.21 different KI processes in the chat. Conversation groups that did not use the adaptive question bank engaged in 1.76 different KI processes on average in the chat. This difference is also not statistically significant (p = 0.241). Notably, none of the conversation groups who did not use the adaptive question bank engaged in all four KI processes, and 5 groups (15.15%) who used the adaptive question bank engaged in all four KI processes (Figure 4). This indicates that the question bank may support conversation groups to engage in the full KI cycle, which helps to support an integrated understanding of scientific phenomena [7].

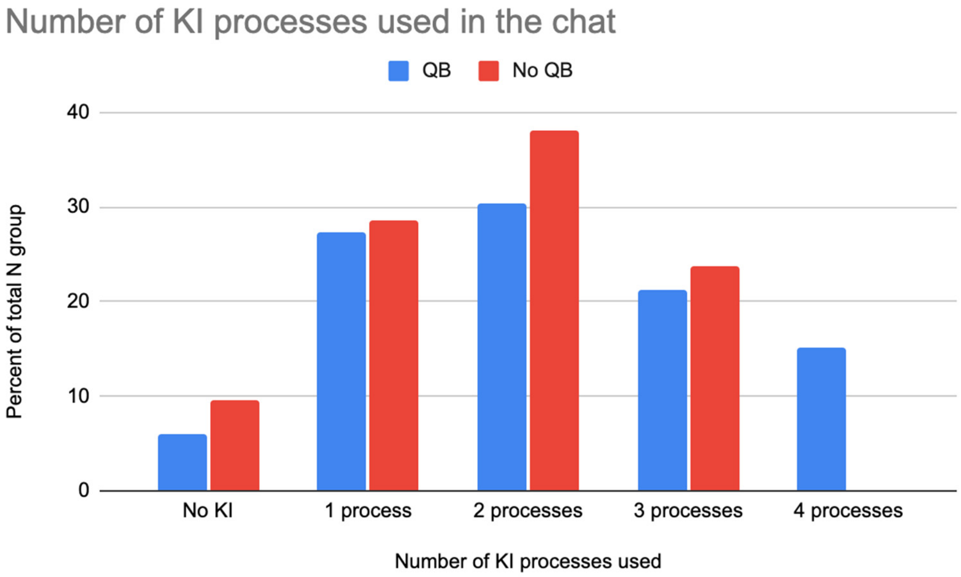


Figure 4. Number of KI processes groups engaged in during the chat was split into groups that used the adaptive question bank and groups that did not use the adaptive question bank.

When broken down by specific KI processes (Table 5), we found that two processes had a statistically significant difference in use between conversations that used the adaptive question bank and those that did not use the adaptive question bank. Elicit was used in 51.5% of groups that used the adaptive question bank prompts but was only used in 23.8% of groups that did not use the adaptive question bank prompts (Chi-Square, X^2 = 4.080,

p = 0.043 *). Add/discover was used in 90.9% of groups that used the adaptive question bank prompts and in 61.9% of groups that did not use the adaptive question bank prompts (Chi-Square, X^2 = 6.656, p = 0.010 **). We did not see any statistically significant differences in distinguishing (Chi-Square, X^2 = 3.021, p = 0.082) between conversations where the question bank was not used and those who used the question bank. Connect/reflect was low for both groups (Table 5), with no statistically significant differences (Chi-square, X^2 = 0.001, p = 0.971).

Table 5. Frequencies, percentages, and chi-square test results for KI processes between QB and No-QB groups. * and ** indicates statistically significant findings.

KI Process	Used QB (% = N/33)	Did not Use QB (% = N/23)	Chi-Square
Some KI	31 (93.9%)	19 (90.5%)	$X^2 = 0.224$, $p = 0.636$
Elicit	17 (51.5%)	5 (23.8%)	$X^2 = 4.080$, $p = 0.043$ *
Discover	30 (90.9%)	13 (61.9%)	$X^2 = 6.656$, $p = 0.010$ **
Distinguish	14 (42.4%)	14 (66.7%)	$X^2 = 3.021$, $p = 0.082$
Connect/Reflect	8 (24.2%)	5 (23.8%)	$X^2 = 0.001$, $p = 0.971$
	33 (100%)	23 (100%)	

These findings suggest that the question bank supports students to elicit ideas from their peers (ex, Explain more about your thinking pls) and creates opportunities for discovery by sharing new ideas in the chat (ex, I believe that some of the plates will be oceanic plates, and once oceanic and continental plates push against each other, there will be volcanoes). Adding new ideas into the chat is not surprising, as the question bank prompts were designed to elicit new ideas from students related to the ideas that they shared in their pre-chat response [33]. Prompting pairs to elicit ideas from their peers, often in the form of student-generated questions, was less expected. One hypothesis is that using the question bank prompts modeled asking questions for students, prompting them to perform the task themselves. The question bank prompts do not appear to support students in distinguishing between their ideas. The distinguishing step in the knowledge integration process is key for supporting integrated understanding [7], so future research using the question bank should examine how best to write prompts to support students in distinguishing between ideas, not just eliciting peers' ideas or sharing new ideas in the chat.

Finally, we examine if specific types of prompts, adaptive and generic, led to certain KI processes in the chat. When pairs used adaptive question bank prompts in the chat, the next turn was often a new idea, either shared by the other pair or by the pair that used the original adaptive question bank prompt. Tables 6 and 7 show two examples of add/discover processes that follow the use of an adaptive question bank prompt. In Table 6, pair 005 uses an adaptive question prompt that asks students to consider the location of Mt. Hood and what this location might tell them about the types of plates involved. Pair 006 shares a new idea about the ring of fire and then elaborates on that idea to explain what types of plates are involved in Mt. Hood's formation. In Table 7, one pair (009) asks questions using the adaptive prompts and their own question, and the other pair (008) shares new ideas in the chat in response. At the end of this excerpt, 009 finally shares a new idea (about density), but it is not taken up further in the chat conversations, and both pairs move to a brief off-task conversation before ending the conversation; however, this pattern of utilizing adaptive question bank prompts and providing new ideas in the chat in response to the prompt is common across chat conversation groups.

This is in contrast to responses to the generic prompt, which asks students to compare their pre-chat responses. This results in students distinguishing between ideas in response to the prompt by comparing pre-chat answers. In Table 8, both pairs utilize the generic question bank, then immediately move to distinguish why they agree, though for different reasons. Pair 020 focuses on evaluating the level of detail of the explanations, while 018 centers on a specific idea that occurs in both explanations (convergent). While this conversation group elaborated on their ideas, other conversations often responded to the generic question by simply stating agreement or disagreement without any further elaboration as to why (ex, Yes, we agree or I agree completely with them).

Table 6. An example of common add/discover responses to an adaptive question bank prompt.

Turn	Pair	Chat Text	Code
1	005	Mt. Hood is near the pacific ocean. How does this feature affect the plate types?	Adaptive Question Bank Prompt
2	006	maybe that's why the ring of fire is here	Add/Discover
3	006	I believe that some of the plates will be oceanic plates, and once oceanic and continental plates push against each other, there will be volcanoes.	Add/discover

Table 7. An example of common turns in conversation and KI processes (elicit and add/discover) in response to adaptive question bank prompts.

Turn	Pair	Chat Text	Code
1	009	When continental oceanic plates collide, a volcano often forms. Why?	Adaptive Question Bank Prompt
2	008	The oceanic plate might move down, which causes magma to rise and cause a volcano.	Add/Discover
3	009	yess	Other (agreement, no elaboration)
4	009	How do the characteristics of different plates lead to subduction?	Adaptive Question Bank Prompt
5	008	One plate might have moved down to create the subduction.	Add/Discover
6	009	yes, why do you think that happened?	Elicit
7	009	explain more about your thinking pls	Elicit
8	008	Gravity might have pulled the ocean plate down.	Add/Discover
9	009	WE thought it was density?	Add/Discover

Table 8. An example of common turns in conversation and KI processes (distinguish) in response to generic question bank prompts.

Turn	Pair	Chat Text	Code
1	020	Do you agree or disagree with your partner? What would you add or build on?	Generic Question Bank
2	018	Do you agree or disagree with your partner? What would you add or build on?	Generic Question Bank
3	020	yes I agree with my partner because their answer is very similar to ours, although I would add more detail to your guy's explanation	Distinguish
4	018	Yes because the two plates come together	Distinguish

Table 9 provides an example of a conversation that utilized both adaptive and generic prompts. The conversation begins with pair 003 utilizing the generic prompt, which they utilize to distinguish similarities between the two explanations (convergent). The pairs then utilize two adaptive prompts. The first prompts the pairs to have a conversation about plate types, which consists of Add/discover, Elicit, and Distinguish turns. The pairs appear to use the second adaptive prompt to continue addressing the first. They end the conversation by using the generic prompt to help restate the agreement without any further elaboration.

Table 9. An example of a conversation that utilized both adaptive and generic question bank prompts. This conversation group engages in elicit, add/discover, and distinguish KI processes.

Turn	Pair	Chat Text	Code
1	003	Do you agree or disagree with your partner? What would you add or build on?	Question bank (generic)
2	003	we both agree that they are converging	Distinguish
3	004	we agree, too, because when a convergent boundary happens, it makes a mountain	Distinguish
4	004	Mt. Hood is near the pacific ocean. How does this feature affect the plate types?	Question bank (NLP-adaptive)
5	003	it would make the plate types convergent and oceanic, so it would change them	Discover
6	004	What types of plates are colliding?	Question bank (NLP-adaptive)
7	004	Do you think that the plate types are oceanic and continental?	Elicit Add/Discover
8	003	yes, we do think they are continental and oceanic	Distinguish
9	003	Do you agree or disagree with your partner? What would you add or build on?	Question bank (generic)
10	004	nothing else. I completely agree with them	Distinguish

Examining chat conversations highlights some features and improvements to the question bank prompts. Adaptive prompts, which often ask students to elaborate or clarify thinking, prompt the sharing of new ideas in the chat. The generic prompts, which ask pairs to compare their answers, prompt distinguishing between ideas through discussions of agreement or disagreement. Framing of prompts (i.e., eliciting new ideas or asking students to compare ideas) contributes to how the scaffold will push students to discuss their ideas in the chat. Future work can seek to adjust adaptive prompts to push students to consider more comparisons between ideas, such as the generic prompts in this study, rather than simply asking them to share new ideas.

Overall, students utilized question bank prompts (particularly adaptive prompts) frequently, indicating that students felt that this type of scaffold was useful in supporting their conversations. While there were no major differences in KI gains between those who used the question bank prompts in their conversations and those who did not use them, our analysis of KI processes within chat conversations reveals statistical differences in the types of KI processes used in the chat between those that used the question bank prompts and those that did not. Adaptive prompts support the use of all KI processes, especially elicit and discover. Meanwhile generic prompts support students to distinguish more often. This highlights the importance of prompt framing in targeting specific KI processes to support student thinking. These findings reveal future ways we can adjust the prompts to support KI processes such as distinguishing, which are especially important in supporting students to form coherent understandings of scientific phenomena [7].

5. Discussion

This study explores the impact of an NLP-informed chat tool on students' knowledge integration of earth science topics. In answering our first research question, we discovered that the NLP idea detection proved reliable for most ideas when compared to human scoring. This shows the promise of using NLP idea detection to support classroom decision-making and prompting; however, our investigation into our second research question revealed that utilizing detected ideas to assign students to conversation groups that maximized different ideas proved challenging as students had many similar ideas going into peer chat activity. We found that the KI score and the average number of ideas in pre-chat and post-chat responses were similar across conditions. We found a statistically significant increase in specific mechanistic ideas for those in the NLP-informed condition. This indicates that when students have fairly similar ideas, grouping students based on their ideas does not impact outcomes.

Several future directions for studying grouping using AI technologies have emerged. The first is to explore logic that allows more flexibility in utilizing idea detection to create groups, leading to more impactful outcomes. For example, grouping pairs with specific differences in ideas, such as contrasting ideas about the plate types involved in the creation of Mt. Hood, could prompt students to distinguish between these different perspectives. This logic builds on the knowledge integration literature that suggests that supporting students to consider and compare additional ideas can support more coherent scientific understanding [8,51].

The second is to study how teachers use differences in ideas to group students. Research could investigate the relative role of ideas versus other factors, such as student social dynamics. This could create NLP that includes data teachers use to form groups, such as long-term assessment data or knowledge of social interactions. Using diverse data sources could help teachers promote equity and build peer respect in conversation groups, which is often a goal during group work [3,33].

Though creating productive groups is important, students also need support to engage in productive conversation about their ideas. Well-designed scaffolds can push students to think more deeply about one another's ideas rather than simply confirming or supporting one another's ideas [49]. Our findings related to research question three highlight how NLP can be leveraged to create supportive scaffolds that push students to engage in KI processes when discussing ideas with one another.

The KI dialog prompts in prior work show promise in eliciting student ideas in the chat-bot conversations [33,34,46,47]. The NLP-informed language scaffolds in the form of adaptive question bank prompts, inspired by the dialog prompts in previous research, are also effective in scaffolding student collaboration. This suggests the value of the KI pedagogy for guiding prompt design compared with other approaches [43]. Adaptive, pedagogy-informed scaffolds are crucial to ensuring that each student can participate in a dialog.

Designing an effective adaptive question bank deserves further study. We found a tendency for students who used the adaptive question bank prompts to engage in knowledge integration, specifically the elicit and discover steps of the KI cycle. The adaptive question bank prompts supported students to share new ideas with one another and modeled asking probing questions so that students felt invited to create their own questions and use them in the chat. Redesigning the question bank could increase support for cognitive processes, such as distinguishing between ideas [7]. While our work begins to highlight broader implications for KI-informed prompts, more analysis is needed to gain a deeper understanding of which specific KI prompts are most effective. A deeper analysis could inform best design practices for KI prompts that could inform a variety of NLP-informed KI scaffolds.

Future work regarding our question bank prompts includes examining what types or features of the adaptive prompts are more successful in supporting certain KI processes or supporting students to add specific, targeted ideas to the chat conversation. Adaptive prompts could be adjusted to respond to the set of ideas held by the pair, ideally identifying opportunities to compare ideas. Teachers can also inform this work, as they prompt groups while circulating around the classroom. Recordings of the types of prompts teachers use to push student discussions forward could serve as models for future chat prompts.

Though our study did not show major gains in KI, our assessments and study design offer a novel approach to examining the success of chat tools to support learning. In much of the prior literature, the success of the chat tool is often measured by students' ability to complete a specific task or project using the chat as a way to communicate with one another [36,37,43]. Instead, we used a pre and post-assessment design to measure students' learning. Future research is needed to identify indicators of success for short collaboration activities such as the chat tool. This includes questions about when to administer post-assessments and questions about what types of assessments, scoring rubrics, or codebooks best measure student learning [52,53].

Further work is also needed to determine how and when online chat tools contribute to learning. Some studies evaluate these tools on their own, while others embed them into coherent learning units [37,43]. More clarity is needed to understand how these tools can be utilized by students and teachers within the context of the existing curriculum. In this study, the chat activity was placed at the end of the lesson after the class had engaged with the same content and discovered similar ideas about plate tectonics. As a result, many students came into the chat activity with similar ideas about how Mt. Hood formed. Future directions include exploring how adding the chat activity at different points in the unit might impact the ideas shared within chat conversations. For example, the chat could be used as a first activity to prompt students to elicit the initial ideas they and their peers hold about a topic. This could reveal students' initial understandings of a scientific phenomenon and make those ideas public to peers and the teacher so that they can be interrogated and built upon. This could contribute to understanding the role of chat activities in supporting student learning in classroom-based settings.

The chat tool was used in a middle school science classroom; however, with the increase in online learning opportunities since the COVID-19 pandemic, an effective chat environment could support students across grade levels. More work is needed to understand how online chat tools can support high school and college level students, where collaborative online work is more common.

This study shows how NLP can be used beyond an assessment tool that can inform teachers about the ideas that their students hold. We leverage idea detection in a novel way, using it to inform tasks such as grouping and the assignment of scaffolds that are normally conducted by the teacher. Though we do not seek to replace teachers who have vast knowledge of their students and practices to support student learning [49,54,55], this application of NLP helps to lessen the tasks that teachers must complete during a single class period. A way to build on this work to provide teachers with more insights is to use NLP to identify cognitive processes, such as KI processes, in students' chat conversations. This information could be used to assign or plan learning activities to students that build on the processes students are already engaging in or support students to engage in different KI processes.

As AI and NLP become more integrated into classroom technologies, the accuracy of these tools must continue to be evaluated and improved in order to ensure reliability [18,32]. Though our NLP technology proved accurate for most ideas, some ideas (Isolated volcano and Incorrect density as the mechanism for subduction) were challenging to detect. Future work includes continuing to improve the NLP model as we acquire more training data from classrooms that use the unit and NLP tools.

Finally, as we continue to explore the use of AI in education, looking for ways that AI can be used to help make teachers' jobs easier, rather than framing AI as a tool to replace teacher's expert knowledge, is important. One way to accomplish this is by inviting teachers and other educators into the conversation about AI [56] and involving them in design-based research focusing on developing and refining AI-based classroom technologies [33,57].

6. Conclusions and Limitations

We applied AI technologies by utilizing NLP to help facilitate student-to-student online chat conversations. Supporting students to discuss ideas with one another is a vital tool that supports student learning. Appropriate scaffolding is key in supporting students to engage in productive discussions. This work charts new territory in the application of the use of NLP to help facilitate peer discussion in real-life classrooms, taking some of the facilitation and scaffolding tasks off of the teacher. While the NLP idea detection proved accurate, our design still needs ongoing refinement in supporting students' knowledge integration; however, our findings point to many ways that the technology can be improved. These include improvements in grouping logic and question bank prompts to be more responsive to the ideas students hold and facilitate more specific cognitive processes in chat conversations. Teachers' input and ideas are key in moving the technology forward. Further research is needed to test these design improvements.

We identified three limitations for this study. The first limitation concerns possible issues with the grouping logic. When the computer grouped pairs, not all pairs were available to use for the grouping logic due to timing issues, so the computer grouped pairs that were ready. Any pairs that moved to the chat late were paired with others who were late or other pairs that were ungrouped. To address this issue, future directions include creating grouping logics that are more flexible so that they can still work even if students are not joining the chat at the exact same time. Another is to create a more flexible activity that provides students who arrive to the chat early with an alternative activity while others catch up. As we work towards solutions, it is important to partner with teachers who have a wealth of knowledge about these real-life classroom constraints and challenges.

Second, a hybrid approach to group collaboration (with students working in pairs in person and collaborating in conversation groups using the online chat tool) was employed in this study. This limited our sample size, and our results may only apply to this hybrid approach. Additionally, this approach also means that we were unable to measure every student's learning, as some pairs might have one student typing or sharing ideas while another watches. Future studies may investigate different approaches to group collaboration with the chat tool, including having students work independently and be paired with one other student to chat or groups students or pairs in different classes at the same school.

Finally, there is room for improvement in using pre- and post-assessments in this way, as we did not see significant differences in KI levels for both grouping assignments or the use of the adaptive question bank. We posit this could be because of the relatively short length of time students were engaged in the chat activity between generating the pre-chat and post-chat explanations. Previous research has suggested that students are less likely to increase KI levels when instructional time for a specific unit or topic is shortened [58]. Though we did not shorten instructional time for the unit as a whole, the chat activity itself was relatively short. Both teachers used one 50-min class period to complete the activity, and no groups used the entire 50 min to answer the explanation item, complete the chat activity, and answer the post-chat explanation item. Though the Mt. Hood assessment item is well aligned with instruction, which is a key factor in creating an accurate KI assessment item [9,35], this short period of time between pre and post-chat responses might not make it an ideal measure of student learning from the chat activity because students need a longer period of time to truly integrate their ideas [16].

Future research should push to implement chat tools and AI in real-life classrooms in novel ways that attend to the complexities of classroom environments. As stated above, teachers are important collaborators in this work, and their expertise can help inform the design of chat environments and accompanying scaffolds. A deeper analysis of how teachers facilitate and scaffold chat technology in their classrooms is needed, as their facilitation can provide ideas and examples to improve the chat interface itself. Leveraging partnerships with teachers is key to moving the practical application of AI in classrooms forward.

Author Contributions: All authors contributed to the study conception and design. K.B. and J.M.L.-B. contributed to the design of the chat environment and NLP tools. K.B., H.-Y.C. and J.M.L.-B. contributed to the methodology and data analysis, which M.C.L. provided feedback on. K.B. and H.-Y.C. contributed to the original draft preparation and all authors contributed to reviewing and editing the draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation, grant number 2101669.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the University of California Berkeley (Approval Code: 2021-06-14389; Approval date: 27 February 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study and was obtained in alignment with IRB approval.

Data Availability Statement: The datasets presented in this article are not readily available due to our approved Institutional Review Board's Human Subjects protocol for conducting research with a vulnerable population (children, under the age of 18). The authors upon emailed request will provide an explanation of the analytic syntax and discuss how to adapt it for others' logged data.

Acknowledgments: We thank the two teachers' involved in this study for their contributions and Thank you to the two teachers who were part of this study. Their work and feedback is key in improving the technology to better support student learning. Thank you to the TELS (Technology Enhanced Learning in Science) research team at the Berkeley School of Education for the feedback while we prepared the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chi, M.T.H.; Siler, S.A.; Jeong, H.; Yamauchi, T.; Hausmann, R.G. Learning from human tutoring. *Cogn. Sci.* **2001**, *25*, 471–533. [CrossRef]
2. Sampson, V.; Clark, D. The impact of collaboration on the outcomes of scientific argumentation. *Sci. Educ.* **2009**, *93*, 448–484. [CrossRef]
3. van de Pol, J.; Mercer, N.; Volman, M. Scaffolding Student Understanding in Small-Group Work: Students' Uptake of Teacher Support in Subsequent Small-Group Interaction. *J. Learn. Sci.* **2018**, *28*, 206–239. [CrossRef]
4. Guo, S.; Zheng, Y.; Zhai, X. Artificial intelligence in education research during 2013–2023: A review based on bibliometric analysis. *Educ. Inf. Technol.* **2024**, *29*, 16387–16409. [CrossRef]
5. Ouyang, F.; Zheng, L.; Jiao, P. Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Educ. Inf. Technol.* **2022**, *27*, 7893–7925. [CrossRef]
6. Zhai, X.; Neumann, K.; Krajcik, J. AI for Tackling STEM Education Challenges. *Front. Educ.* **2023**, *8*, 1183030. [CrossRef]
7. Linn, M.C.; Eylon, B.S. *Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration*; Routledge: London, UK, 2011.
8. Gerard, L.; Linn, M.C. Computer-based guidance to support students' revision of their science explanations. *Comput. Educ.* **2022**, *176*, 104351. [CrossRef]
9. Gerard, L.; Linn, M.C.; Madhok, J. Examining the Impacts of Annotation and Automated Guidance on Essay Revision and Science Learning. In *Transforming Learning, Empowering Learners: Proceedings of the International Conference of the Learning Sciences (ICLS) 2016, Singapore, 20–24 June 2014*; Looi, C.K., Polman, J.L., Cress, U., Reimann, P., Eds.; International Society of the Learning Sciences: Bloomington, IN, USA, 2016; Volume 1.
10. Tansomboon, C.; Gerard, L.F.; Vitale, J.M.; Linn, M.C. Designing automated guidance to promote productive revision of science explanations. *Int. J. Artif. Intell. Educ.* **2017**, *27*, 729–757. [CrossRef]
11. Zhai, X.; Yin, Y.; Pellegrino, J.W.; Haudek, K.C.; Shi, L. Applying machine learning in science assessment: A systematic review. *Stud. Sci. Educ.* **2020**, *56*, 111–151. [CrossRef]
12. Dolenc, K.; Aberšek, B.; Aberšek, M.K. Online Functional Literacy, Intelligent Tutoring Systems and Science Education. *J. Balt. Sci. Educ.* **2015**, *14*, 162–171. [CrossRef]
13. Billings, K.; Gerard, L.; Linn, M.C. Improving Teacher Noticing of Students' Science Ideas with a Dashboard. In Proceedings of the 15th International Conference of the Learning Sciences—ICLS 2021, Bochum, Germany, 8–11 June 2021; de Vries, E., Hod, Y., Ahn, J., Eds.; International Society of the Learning Sciences: Bochum, Germany, 2021; pp. 1027–1028.
14. Wiley, K.; Dimitriadis, Y.; Linn, M. A human-centred learning analytics approach for developing contextually scalable K-12 teacher dashboards. *Br. J. Educ. Technol.* **2024**, *55*, 845–885. [CrossRef]
15. Choi, S.; Jang, Y.; Kim, H. Influence of Pedagogical Beliefs and Perceived Trust on Teachers' Acceptance of Educational Artificial Intelligence Tools. *Int. J. Hum. Comput. Interact.* **2022**, *39*, 910–922. [CrossRef]
16. Chounta, I.A.; Bardone, E.; Raudsep, A.; Pedaste, M. Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K-12 education. *Int. J. Artif. Intell. Educ.* **2022**, *32*, 725–755. [CrossRef]
17. Zafari, M.; Bazargani, J.S.; Sadeghi-Niaraki, A.; Choi, S.M. Artificial intelligence applications in K-12 education: A systematic literature review. *IEEE Access* **2022**, *10*, 61905–61921. [CrossRef]
18. Jescovitch, L.N.; Scott, E.E.; Cerchiara, J.A.; Merrill, J.; Urban-Lurain, M.; Doherty, J.H.; Haudek, K.C. Comparison of machine learning performance using analytic and holistic coding approaches across constructed response assessments aligned to a science learning progression. *J. Sci. Educ. Technol.* **2021**, *30*, 150–167. [CrossRef]

19. Kaldaras, L.; Li, T.; Haudek, K.; Krajcik, J. Developing Rubrics for AI Scoring of NGSS Learning Progression-based Scientific Models. In Proceedings of the 2024 American Education Research Association (AERA) Annual Meeting, Philadelphia, PA, USA, 11–14 April 2024; AERA: Philadelphia, PA, USA, 2024.
20. Kaldaras, L.; Yoshida, N.R.; Haudek, K.C. Rubric development for AI-enabled scoring of three-dimensional constructed-response assessment aligned to NGSS learning progression. *Front. Educ.* **2022**, *7*, 983055. [CrossRef]
21. Li, H.; Gobert, J.; Dickler, R. Automated Assessment for Scientific Explanations in On-Line Science Inquiry. In Proceedings of the International Conference on Educational Data Mining Society (EDM), Wuhan, China, 25–28 June 2017.
22. Liu, O.L.; Rios, J.A.; Heilman, M.; Gerard, L.; Linn, M.C. Validation of automated scoring of science assessments. *J. Res. Sci. Teach.* **2016**, *53*, 215–233. [CrossRef]
23. Nehm, R.H.; Ha, M.; Mayfield, E. Transforming biology assessment with machine learning: Automated scoring of written evolutionary explanations. *J. Sci. Educ. Technol.* **2012**, *21*, 183–196. [CrossRef]
24. Schleifer, A.G.; Klebanov, B.B.; Ariely, M.; Alexandron, G. Transformer-based Hebrew NLP models for short answer scoring in biology. In Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), Toronto, ON, Canada, 13 July 2023; pp. 550–555.
25. Wang, C.; Liu, X.; Wang, L.; Sun, Y.; Zhang, H. Automated scoring of Chinese grades 7–9 students' competence in interpreting and arguing from evidence. *J. Sci. Educ. Technol.* **2021**, *30*, 269–282. [CrossRef]
26. Hicke, Y.; Masand, A.; Guo, W.; Gangavarapu, T. Assessing the efficacy of large language models in generating accurate teacher responses. In Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications, Toronto, ON, Canada, 13 July 2023; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 745–755.
27. Huber, T.; Niklaus, C.; Handschuh, S. Enhancing Educational Dialogues: A Reinforcement Learning Approach for Generating AI Teacher Responses. In Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), Toronto, ON, Canada, 13 July 2023; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 736–744.
28. Kurtz, G.; Amzalag, M.; Shaked, N.; Zaguri, Y.; Kohen-Vacs, D.; Gal, E.; Zailer, G.; Barak-Medina, E. Strategies for Integrating Generative AI into Higher Education: Navigating Challenges and Leveraging Opportunities. *Educ. Sci.* **2024**, *14*, 503. [CrossRef]
29. Nye, B.D.; Graesser, A.C.; Hu, X. AutoTutor and family: A review of 17 years of natural language tutoring. *Int. J. Artif. Intell. Educ.* **2014**, *24*, 427–469. [CrossRef]
30. Paladines, J.; Ramirez, J. A systematic literature review of intelligent tutoring systems with dialogue in natural language. *IEEE Access* **2020**, *8*, 164246–164267. [CrossRef]
31. Zhai, X. Practices and theories: How can machine learning assist in innovative assessment practices in science education. *J. Sci. Educ. Technol.* **2021**, *30*, 139–149. [CrossRef]
32. Zhai, X.; Krajcik, J.; Pellegrino, J.W. On the validity of machine learning-based Next Generation Science Assessments: A validity inferential network. *J. Sci. Educ. Technol.* **2021**, *30*, 298–312. [CrossRef]
33. Gerard, L.; Holtmann, M.; Riordan, B.; Linn, M.C. Impact of an Adaptive Dialog that Uses Natural Language Processing to Detect Students' Ideas and Guide Knowledge Integration. *J. Educ. Psychol.* **2024**. *Advance online publication*. [CrossRef]
34. Li, W.; Gerard, L.; Lim-Breitbart, J.; Bradford, A.; Linn, M.C.; Riordan, B.; Steimel, K. Explaining thermodynamics: Impact of an adaptive dialog based on a natural language processing idea detection model. In Proceedings of the 17th International Conference of the Learning Sciences—ICLS 2023, Montreal, QC, Canada, 10–15 June 2023; Blikstein, P., Van Aalst, J., Kizito, R., Brennan, K., Eds.; International Society of the Learning Sciences: Bloomington, IN, USA, 2023; pp. 1306–1309. [CrossRef]
35. Patterson, A.D. Equity in groupwork: The social process of creating justice in a science classroom. *Cult. Stud. Sci. Educ.* **2019**, *14*, 361–381. [CrossRef]
36. Jonassen, D.H.; Kwon, H. Communication patterns in computer mediated versus face-to-face group problem solving. *Educ. Technol. Res. Dev.* **2001**, *49*, 35–51. [CrossRef]
37. Sins, P.H.M.; Savelsbergh, E.R.; van Joolingen, W.R.; van HoutWolters, B.H.A.M. Effects of face-to-face versus chat communication on performance in a collaborative inquiry modeling task. *Comput. Educ.* **2011**, *56*, 379–387. [CrossRef]
38. Bianchini, J.A. Where knowledge construction, equity, and context intersect: Student learning of science in small groups. *J. Res. Sci. Teach.* **1997**, *34*, 1039–1065. [CrossRef]
39. Webb, N.M. Information processing approaches to collaborative learning. In *The International Handbook of Collaborative Learning*; Routledge: New York, NY, USA, 2013; pp. 19–40.
40. Gerard, L.; Kidron, A.; Linn, M.C. Guiding collaborative revision of science explanations. *Int. J. Comput. Support. Collab. Learn.* **2019**, *14*, 291–324. [CrossRef]
41. Moore, K.; Anthony, H.G. Using Sentence Frames and Question Cards to Scaffold Discourse and Argumentation in Science. In *Teaching Science Students to Communicate: A Practical Guide*; Rowland, S., Kuchel, L., Eds.; Springer: Cham, Switzerland, 2023. [CrossRef]
42. Cohen, E.G.; Lotan, R.A. *Designing Groupwork: Strategies for the Heterogeneous Classroom*, 3rd ed.; Teachers College Press: New York, NY, USA, 2014.
43. Lazonder, A.W.; Wilhelm, P.; Ootes, S.A. Using sentence openers to foster student interaction in computer-mediated learning environments. *Comput. Educ.* **2003**, *41*, 291–308. [CrossRef]

44. Linn, M.C.; McElhaney, K.; Gerard, L.; Matuk, C. Inquiry learning and opportunities for technology. In *International Handbook of the Learning Sciences*; Fischer, F., Hmelo-Silver, C.E., Goldman, S.R., Reimann, P., Eds.; Routledge: New York, NY, USA, 2018; pp. 221–233.
45. Lee, H.S.; Gweon, G.H.; Lord, T.; Paessel, N.; Pallant, A.; Pryputniewicz, S. Machine learning-enabled automated feedback: Supporting students' revision of scientific arguments based on data drawn from simulation. *J. Sci. Educ. Technol.* **2021**, *30*, 168–192. [CrossRef]
46. Holtmann, M.; Gerard, L.; Li, W.; Linn, M.C.; Riordan, B.; Steimel, K. How does an adaptive dialog based on natural language processing impact students from distinct language backgrounds? In Proceedings of the 17th International Conference of the Learning Sciences—ICLS 2023, Montreal, QC, Canada, 10–15 June 2023; Blikstein, P., Van Aalst, J., Kizito, R., Brennan, K., Eds.; International Society of the Learning Sciences: Bloomington, IN, USA, 2023; pp. 1350–1353. [CrossRef]
47. Bradford, A.; Li, W.; Riordan, B.; Steimel, K.; Linn, M.C. Adaptive dialog to support student understanding of climate change mechanism and who is most impacted. In Proceedings of the 17th International Conference of the Learning Sciences—ICLS 2023, Montreal, QC, Canada, 10–15 June 2023; Blikstein, P., Van Aalst, J., Kizito, R., Brennan, K., Eds.; International Society of the Learning Sciences: Bloomington, IN, USA, 2023; pp. 1350–1353. Available online: https://repository.isls.org/handle/1/10333 (accessed on 10 December 2024).
48. Bell, P. On the theoretical breadth of design-based research in education. *Educ. Psychol.* **2004**, *39*, 243–253. [CrossRef]
49. Gerard, L.F.; Spitulnik, M.; Linn, M.C. Teacher use of evidence to customize inquiry science instruction. *J. Res. Sci. Teach.* **2010**, *47*, 1037–1063. [CrossRef]
50. Riordan, B.; Wiley, K.; Chen, J.K.; Bradford, A.; Bichler, S.; Mulholland, M.; Gerard, L.F. Automated scoring of science explanations for multiple NGSS dimensions and knowledge integration. In Proceedings of the Annual Meeting of the American Educational Research Association (AERA), San Francisco, CA, USA, 17 April 2020.
51. McNeil, K.L.; Krajcik, J. Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *J. Res. Sci. Teach.* **2008**, *5*, 53–78. [CrossRef]
52. Liu, O.L.; Lee, H.S.; Hofstetter, C.; Linn, M.C. Assessing knowledge integration in science: Construct, measures, and evidence. *Educ. Assess.* **2008**, *13*, 33–55. [CrossRef]
53. Ryoo, K.; Linn, M.C. Designing and validating assessments of complex thinking in science. *Theory Into Pract.* **2015**, *54*, 238–254. [CrossRef]
54. Goldman, S.R.; Hmelo-Silver, C.E.; Kyza, E.A. Collaborative Design as a context for teacher and researcher learning: Introduction to the special issue. *Cogn. Instr.* **2022**, *40*, 1–6. [CrossRef]
55. Philip, T.M.; Pham, J.H.; Scott, M.; Cortez, A. Intentionally addressing nested systems of power in schooling through teacher solidarity co-design. *Cogn. Instr.* **2022**, *40*, 55–76. [CrossRef]
56. Mah, C.; Walker, H.; Phalen, L.; Levine, S.; Beck, S.W.; Pittman, J. Beyond CheatBots: Examining Tensions in Teachers' and Students' Perceptions of Cheating and Learning with ChatGPT. *Educ. Sci.* **2024**, *14*, 500. [CrossRef]
57. Celik, I.; Dindar, M.; Muukkonen, H.; Järvelä, S. The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends* **2022**, *66*, 616–630. [CrossRef]
58. Clark, D.; Linn, M.C. Designing for Knowledge Integration: The Impact of Instructional Time. *J. Learn. Sci.* **2003**, *12*, 451–493. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.