Title
Designing an Adaptive Dialogue to Promote Science Understanding

ISBN
9781737330653

Authors
Gerard, L
Bichler, S
Bradford, A
et al.

Publication Date
2022

Citation: Gerard, L., Bichler, S., Bradford, A., Linn, M. C., Steimel, K., & Riordan, B. (2022). Designing an Adaptive Dialogue to Promote Science Understanding. In C. Chinn, E. Tan, C. Chan, & Y. Kali (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences—ICLS 2022* (pp. 1653–1656). International Society of the Learning Sciences.

Designing an Adaptive Dialogue to Promote Science Understanding

Libby Gerard, Sarah Bichler, Allison Bradford, Marcia C. Linn
libbygerard@berkeley.edu, sbichler@berkeley.edu, allison_bradford@berkeley.edu, mclinn@berkeley.edu,
University of California, Berkeley

Kenneth Steimel, Brian Riordan
ksteimel@ets.org, briordan@ets.org
ETS

Abstract: We used Natural Language Processing (NLP) to design an adaptive computer dialogue that engages students in a conversation to reflect on and revise their written explanations of a science dilemma. We study the accuracy of the NLP idea detection. We analyze how 98 12-13 year-olds interacted with the dialogue as a part of a Diagnostic Inventory. We study students' initial and revised science explanations along with their logged responses to the dialogue. The dialogue led to a high rate of student revision compared to prior studies of adaptive guidance. The adaptive prompt encouraged students to reflect on prior experiences, to consider new variables, and to raise scientific questions. Students incorporated these new ideas when revising their initial explanations. We discuss how these adaptive dialogues can strengthen science instruction.

Objective

Through a partnership among university education researchers, computer scientists, and science teachers from six middle schools, we exploit advances in Natural Language Processing (NLP) to detect the rich ideas that students use to explain science dilemmas. We test designs for computer-generated adaptive dialogues to affirm students' ideas and support them to refine and extend their understanding. We build on prior work demonstrating the value of adaptive guidance that models behaviors for integrated revision such as making connections between ideas (Chi et al., 2017; Gerard & Linn, 2022). The adaptive dialogue in this study promotes equitable science teaching by assisting teachers in encouraging each student to express and elaborate their ideas about the target science topic including ideas grounded in their lived experiences.

Adaptive dialogue for revising science explanations

We design and test an NLP based adaptive dialogue that elicits the different ideas students bring to science problems and models a path to developing those ideas. The adaptive dialogue may assist teachers by ensuring that each student receives guidance which identifies a spark in their explanation to build on. The dialogue is designed to guide students to engage in integrated revision. Integrated revision is grounded in a knowledge integration (KI) perspective on learning, suggesting that as students reformulate the connections among their ideas, distinguishing which ideas are supported by evidence and how they link together, they deepen their understanding and develop durable insights (Gerard & Linn, 2022; Linn & Eylon, 2011). Central to the process of integrated revision is eliciting the different ideas one holds about a problem. These initial ideas often reflect prior experiences or previous science instruction. Eliciting a learner's ideas promotes agency by enabling the students to generate their expectations before they test them, increasing their interest in the outcomes. In repeated studies across disciplines and age groups, retrieval practice, compared to review and retrieval, has shown learning advantages (Karpicke & Roediger, 2008). Articulating one's ideas allows learners to reflect on their many different ideas and distinguish which among their ideas are relevant and supported by evidence.

Eliciting a learner's ideas in science is tricky since students may be reluctant to express ideas that differ from those of others or that they feel will be rejected as incorrect. Research suggests students with non-dominant views may be hesitant to express these ideas in a classroom context. When the culture of school encourages students to "do school" rather than "do science", students are likely to dismiss their ideas developed from years of experience or prior courses and adopt the idea presented by instruction without distinguishing between the two (Carlone et al., 2015). Research shows that when instruction neglects the ideas students hold while introducing new ideas, students develop a fragmented understanding by holding both perspectives rather than evaluating their knowledge (diSessa, 1998). We build on these findings to design an adaptive dialogue that engages each student in integrated revision by affirming students' initial ideas, prompting them to add new ideas as they reflect on the topic, and supporting them to consider the valuable reasoning and evidence underlying their initial ideas.

Methods

We investigate two research questions: (1) How accurate is the NLP technology for detecting ideas students hold about an open-ended science dilemma?, and (2) How do students respond to adaptive guidance that promotes integrated revision on a Diagnostic Inventory?

Participants, Diagnostic Inventory, and NLP-based Adaptive Dialogue

We administered a Diagnostic Inventory with one teacher's 98 7th and 8th graders (50 7th graders; 47 8th graders). The majority of the students identify as a part of historically marginalized racial, linguistic and socioeconomic groups (97% Latinx; 95% of parents speak a language other than English at home; 85% eligible for a free or reduced price lunch). We report on the *Car on a Cold Day* problem, described in Table 1.

To train the NLP scoring model, a corpus of 1000+ student explanations collected in prior research from 5 schools, including demographics similar to the school in this study, was used. Two researchers developed a 5-point Knowledge Integration rubric to score the student explanations for 'Car on a Cold Day', which rewards students for linking normative, relevant ideas. The researchers also identified the distinct ideas expressed by students (21 ideas, Table 1). One student explanation could include multiple ideas. The researchers refined the rubrics until they achieved inter-rater reliability (Cohen's Kappa > .85). Then, each researcher coded 50% of the explanations assigning a KI score, and a tag for each distinct idea within each explanation. Data was used to train an NLP model.

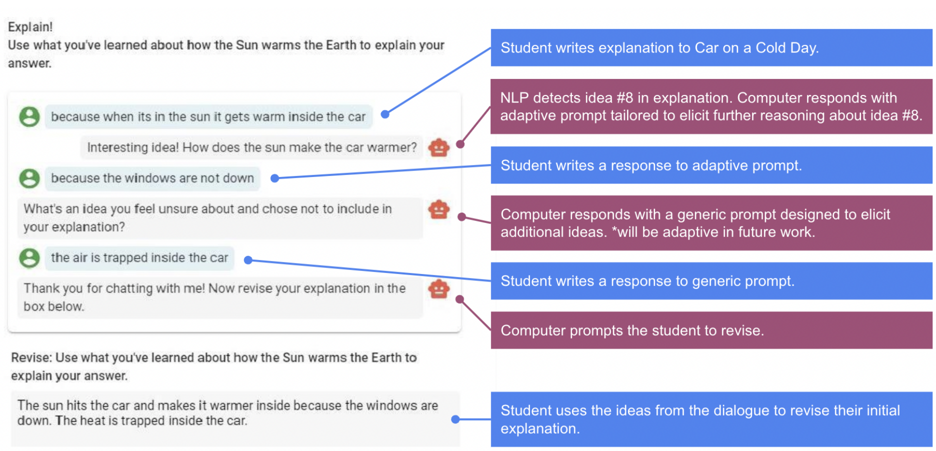
Table 1. *Sample of "Car on a Cold Day" Scoring Rubric and Guidance for Adaptive Dialogue Round 1*

Item prompt: *On a cold day Akbar walks to his car that is parked in the sun. The sun has been shining on the car for the whole day. How will the temperature inside the car feel? Use what you've learned about how the Sun warms the Earth to explain.*

#	Sample Ideas	Assigned Guidance in Dialogue Round 1
3	car same temperature as outside	Can you tell me more about how the temp. inside the car changes the longer it sits in the sun?
4	car warmer than outside	Can you tell me more about how the Sun makes the inside of the car warmer than outside?
7	Heat comes from the sun	Can you tell me more about the forms of energy that come from the sun?
8	Direct sun warms the car	Interesting idea! How does the sun make the car warmer?
9	Metal/car attracts heat/cold	Can you tell me more about how the material impacts how the car gets energy from the sun?

To build the NLP model, we used a word classification or sequence labeling approach (Riordan et al., 2020). First, the words in each response are transformed into numerical vectors with state-of-the-art language models that have been previously trained to capture latent relationships in the English language via the self-attention mechanism. Second, for each word, idea detection is cast as a multi-label classification problem. Consecutive word-level predictions of a targeted idea form a predicted idea span. The idea detection model was trained, evaluated, and deployed with 10-fold cross validation for hyperparameter tuning and evaluation. Final model building used all available data. Adaptive dialogue prompts for each idea were designed to engage students in conversation, modeling an integrated revision process and promoting science understanding (See Figure 1).

Figure 1*: Screenshot of a Sample Guidance Dialogue from the Car on a Cold Day item on Diagnostic Inventory.*



Data Analysis

We determined the most frequently detected idea across all students' initial explanations (idea 8). We then identified the subset of students (N=53) who expressed idea 8. Focusing on these dialogues enabled comparison of the students' paths of reasoning rather than the impact of different adaptive prompts. Next, we selected a sample of 10 students who would be likely to express unique paths of reasoning in the dialogue based on criteria that they had varied levels of English and prior instruction about the greenhouse effect. Three researchers individually annotated each response to characterize students' revision moves. Researchers discussed annotations until reaching agreement. One researcher used the resulting categories [Fig.2] to code the 53 dialogues.

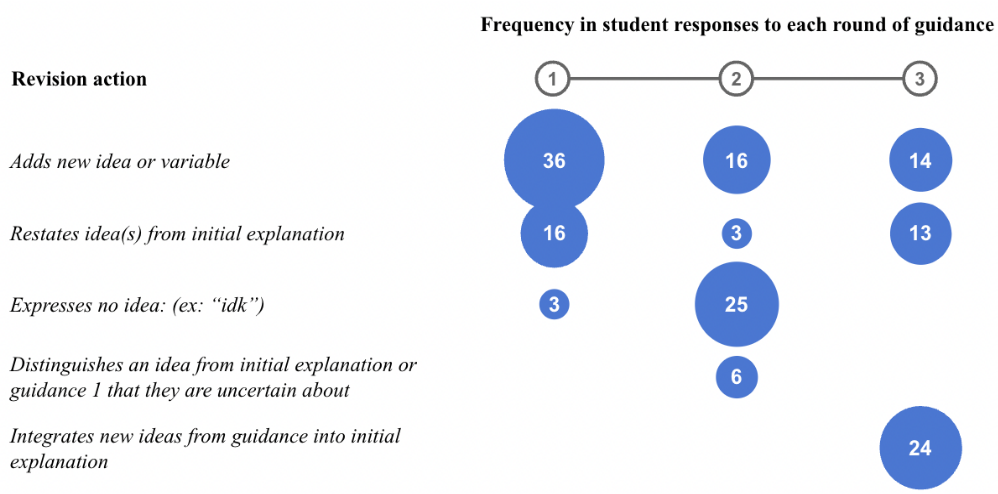
Results

To evaluate the NLP model accuracy, word token-level micro-averaged precision, recall, and F1 scores were used. These metrics account for the imbalance in idea classes. Despite often inaccurately predicting the length of idea spans, the model accurately predicted spans for many of the idea classes. Aggregating across all targeted ideas, we obtained a word-level micro-averaged precision of 0.71, and a recall of 0.51, yielding an F1-score of 0.60. The F1 score (0.60) suggests the difficulty of achieving high accuracy across all idea classes from our relatively small dataset. The model is relatively conservative in predicting idea spans, resulting in higher precision (0.71) but lower recall of ideas (0.51). Some idea classes were sparsely represented in the data (e.g., idea was tagged with less than 300 words of the 27000 words in the data set), thus the model tended not to predict idea spans for these ideas.

Analysis of students' revisions indicated that overall, the dialogue supported students to revise their explanations: 91% of the students (87/96) revised their initial explanation after engaging in two rounds of dialogue. In contrast, in our prior research only 69% of students revised in response to two rounds of knowledge integration adaptive guidance (Gerard & Linn, 2022). Additional studies report that under 50% of students engage in sustained revision when the activity calls for it (e.g. Zhu et al., 2020).

The analysis of student revision moves suggests that the dialogue effectively elicited student ideas (Fig. 2). In response to the *adaptive* first prompt in the dialogue, 36 of the students (68%) added a new idea to elaborate the mechanism underlying their initial explanation. In response to the *generic* second prompt (same for all students), 16 (30%) of these students added another new variable they considered relevant, and 6 (11%) distinguished a part of the mechanism from their initial explanation that they were uncertain about. Twenty-five (48%) of the students however struggled to recognize another relevant idea in response to the generic prompt (e.g., "I'm not sure; IDK"). This could be due to the generic nature of the prompt or because it is the second prompt. After the dialogue, students were encouraged to consolidate the ideas they raised in the dialogue and revise their initial response. When revising, 24 (46%) of the students integrated an idea that they had raised in the dialogue to explain more fully the mechanism they had initially expressed. Fifteen (28%) expressed another new variable, in addition to those raised in the dialogue, in their revised explanation. Overall, the dialogue elicited an increasing number of ideas each student held about the problem. The adaptivity of the prompt in the first round was more effective than the generic prompt in the second round in supporting students to recognize and express the relevant ideas they held.

***Figure 2**. Student revision sequences in the dialogue [N=53]; **1**= Response to adaptive guidance in round 1; **2** = Response to generic guidance in round 2; **3**=Revising initial explanation*



<u>Case analysis</u>. We describe the dialogues for three students who started with a similar idea, yet had distinct sequences of revision to illustrate the revision moves. Each of the students started out by expressing a vague idea that the car will be warm inside because of the sun. The computer accurately detected this idea (idea 8: sun warms the car) and started the conversation with, "Interesting idea! How does the sun make the car warmer?". Each student responded by retrieving a new idea they held, likely by reflecting on their time spent in a hot car and distinguishing relevant evidence to elaborate their view (e.g. (student 1) sun is hot; (2) car windows up; (3) metal material of the car). The computer then asked the generic prompt (what's an idea you felt uncertain about?). At this point student 1 distinguished their uncertainty about how the mechanism worked (how does the inside of the car get hot). Student 2 recognized an additional relevant new idea (air is trapped inside the car) to elaborate the idea they raised in the previous round of dialogue. Student 3 explored the feasibility of their initial idea (will metal overheat and explode?).After these two rounds of dialogue, each of the students revised their initial explanation. Student 1 incorporated new details to elaborate their initial explanation [*When you leave your car parked in the sun it is usually warm when you get in and when you park in the shade it is cool inside.→The sun is very hot and when it shines on the car the car eventually heats up for many hours that it is parked on the road*]. Student 2 linked the new ideas they raised in the dialogue and integrated them with their initial ideas to create a more complete and accurate mechanism [*when its in the sun it gets warm inside the car→The sun hits the car and makes it warmer inside because the windows are [not] down. The heat is trapped inside the car*.] Student 3 abandoned their initial ideas in favor of new ideas they devised to consider and generated an alternative explanation [*sun warms the earth by its hotness Its kinda like a warm blanket wrapped around earth but its actually the sun hotness→The car will feel cold on the inside even though it was in the sun for a long time, it'll be the same temperature as outside*]. The analysis illustrates how the dialogue supported students to recognize the multiple ideas they held relevant to solving the problem, and to reflect on how these ideas link together to refine their science explanation.

Conclusions & Next Steps

This study suggests that asking adaptive knowledge integration oriented questions in a computer generated dialogue can encourage students to take steps toward integrated revision of their ideas. The dialogue supported students to generate many ideas, developed from prior experiences as consideration for solving the science dilemma. Once students added new ideas they either distinguished among these, or used the new idea to buttress or modify their initial ideas. Guiding students to reflect on the different ideas they have, initiates a process in which students raise scientific questions and consider new variables. This may position students to further explore the variables or seek answers in subsequent instruction. This analysis aligns with previous findings on retrieval practice, where prompts to retrieve more ideas from a passage fostered learning (Karpicke & Roediger, 2008). Future work will embed dialogs in instructional units to more closely test impacts of retrieval practice on learning, and to explore guidance strategies to support use of evidence for distinguishing among ideas. We will also explore how adaptive dialogues may support teachers to tailor their instruction to student's specific ideas, and to affirm the paths that lead the student to their ideas. This is particularly valuable when the student expresses non-normative ideas that may otherwise have been treated as simply incorrect or demonstrative of a lack of effort.

References

Carlone, H. B., Johnson, A., & Scott, C. M. (2015). Agency amidst formidable structures: How girls perform gender in science class. *Journal of Research in Science Teaching, 52*(4), 474-488.

Chi, M., Kang, S.,Yaghmourian, D. (2017). Why students learn more from dialogue-than monologue-videos: Analyses of peer interactions. *Journal of the Learning Sciences*, 26 (1), pp. 10-50

diSessa. A. (1988). Knowledge in pieces in G. Forman, P. Pufall (Eds.), Constructivism in the computer age, Erlbaum, Hillsdale, NJ (1988), pp. 49-70

Gerard, L. & Linn, M.C. (2022). Computer-based guidance to support student revision of their science explanations. *Computers & Education, 176*

Karpicke, J.D., & Roediger, H.L. (2008). The critical importance of retrieval for learning. *Science, 319(5865),* 966-968.

Linn, M.C., Eylon, B-S (2011). Science learning and instruction: Taking advantage of technology to promote knowledge integration. Routledge, New York (2011)

Riordan, B., Cahill, A., King Chen, J., Wiley, K., Bradford, A., Gerard, L., & Linn, M.C. (2020). Identifying NGSS-Aligned Ideas in Student Science Explanations. In Workshop on Artificial Intelligence for Education.

Zhu, M., Liu, O.L.,Lee, H-S (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, 143