Enhancing Scientific Argumentation with Automated Feedback in the Context of Two Earth Science Curriculum Units

Ou Lydia Liu, Amy Pallant, Hee-Sun Lee
Margarita Olivera, Vinetha Belur, Mengxiao Zhu– Educational Testing Service
Trudi Lord, Sarah Pryputniewicz – The Concord Consortium

Study 1. There are differences between the contextualized and generic HASbot feedback.

❖Classrooms were randomly assigned to two groups: 168 students (42%) received the generic HASbot and 230 students (58%) received the contextualized HASbot.

Generic	Contextualized
You made a claim without an explanation. Can you include scientific evidence and reasoning that support your claim?	You made a claim without an explanation. Can you explain how energy packets interact with greenhouse gases in the model? Provide evidence that shows how the temperature changes when carbon dioxide is removed.

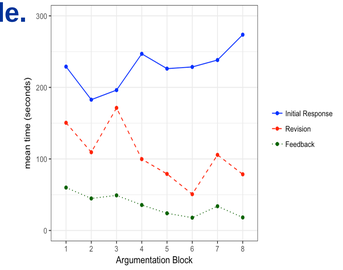
❖ No significant difference on the completion rates, revision rates, time spent on the argumentation blocks, and score changes when averaging the uncertainty and explanation tasks of all modules.

❖ **Significant difference on the number of revisions**

- ❖ Students in the generic HASbot condition made more revisions ($M = 2.20, SD = 1.29$) than students in the contextualized HASbot condition ($M = 1.77, SD = 1.08$); $t(242) = 3.18, p < 0.01$.
- ❖ More revisions under the generic HASbot feedback condition achieved similar score changes with fewer revisions under the contextualized feedback condition.

The amount of time students spend writing their initial argument increase over the course of the module.

- ❖ Log data tracks all online activity by students and is automatically collected by server.
- ❖ Over the course of the module, the time spent responding to argument prompts increased.
- ❖ Over time, students spent less time on revisions of their arguments.



Study 2: Overall, students find the HASbot feedback useful and motivating.

- ❖ 363 students from 8 high schools who completed the Climate Change Module in year 3 of the project were surveyed on their perceptions of the HASbot feedback.
- ❖ 3 out of 5 students felt the HASbot feedback was the right amount.
- ❖ Nearly half of the students felt the HASbot feedback made it clear how to revise their answer.
- ❖ 50% of students would like additional activities with the HASbot feedback.

Reasons for wanting additional activities with HASbot:

Know what areas to work on (28%)	*"It help[s] me to know what to improve on or helps my understanding of what was wrong"*
Chance to improve answers (22%)	*"I liked how I got to revise my work to make it better"*
Helpful in general (20%)	*"I found them really helpful"*
Increases understanding of material (14%)	*"I feel that understanding concepts comes easier with this kind of feedback, and more assignments with this structure could benefit our ability to learn"*
Better grades / scores (13%)	*"I would like it because it shows what I can fix on my work and will help me get a higher score on my work"*
Quick feedback/ immediate correction (10%)	*"I would really love to have automated feedback…so that I can be corrected right away"*
Greater confidence in my work (5%)	*"I would feel much more confident in the work that I turn in"*
Other (15%)	Motivating to receive feedback, independent learning, better arguments, easier to remember correct answers and mistakes

❖ 16% of students reported not wanting additional activities with HASbot feedback.

Reasons for not wanting additional activities with HASbot:

Unhelpful / Doesn't fit (31%)	*"Sometimes, I thought my answer fulfilled all the criteria for the answer to receive a top score, but the automated response would give a very low score."*
Unspecific / Unclear (25%)	*"It was not as personalized as it could have been. If it gave more specific feedback, maybe I would enjoy it."*
Disliked It (no explanation) (20%)	*"I do not like this kind of work"*
Prefers teacher feedback (16%)	*"I would prefer to learn through the teacher because its easier to find out where I went wrong"*
Too tedious (13%)	*"I did not enjoy doing these activities. I thought they were very repetitive and it took forever to complete"*

PROJECT TIMELINE

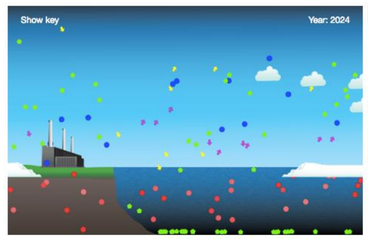
Year 1: Validate automated scoring models for uncertainty-infused scientific argumentation tasks in Climate Change Module (Mao et al., 2018).
Year 2: Validate automated scoring models for uncertainty-infused scientific argumentation tasks in the Water Module; conduct a small scale implementation study of automated feedback with Climate Change Module (Zhu et al., 2017).
Year 3: Compare the effect of contextualized vs. generic feedback using within-teacher, class-level random assignment (**Study 1** & **Study 2**); conduct an implementation study of automated, contextualized feedback with Water Module (**Study 3**).
Year 4 (Ongoing): Conduct a large-scale implementation study on contextualized feedback with Climate Change Module; conduct a small-scale design study with contextualized feedback and model feedback in Water Module.

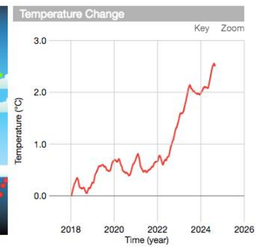
Two Earth Science Modules with Automated Feedback (HASbot) on Written Argumentation

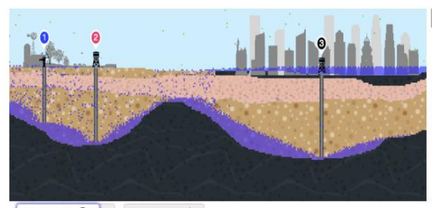
We have added automated feedback to two curriculum modules that enable students in grades 6-12 to explore important and timely questions about our planet: "Will there be enough fresh water?" and "What is the future of Earth's climate?" In each module, students encounter eight scientific argumentation tasks. Each argumentation task is designed as a four-part item set, including: 1) a multiple-choice claim, 2) an open-ended explanation, 3) a certainty rating on a five-point Likert scale (from very uncertain to very certain), and 4) an open-ended rationale for the certainty rating. The modules allow students to run experiments and collect data using interactive simulations. Students are asked to write scientific arguments supported by both the data that they have generated and authentic datasets provided. Students' scientific arguments are scored in real-time using c-rater-ML. Based on their score, students are provided instant feedback on how to revise and improve their written responses. In the Water Module, additional scaffolding is presented via model-based feedback. Use of the models is analyzed with a new decision-tree algorithm and immediate, targeted feedback directs students to specific features in the model that they should explore or to actions necessary to generate meaningful data.

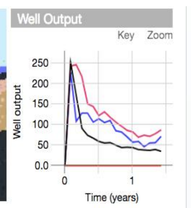
	Climate Change Module	Water Module
Interactive models	✓	✓
Model-based feedback		✓
Scientific argumentation	✓	✓
Automated scoring	✓	✓
Argument feedback	✓	✓
Teacher dashboard	✓	✓

Example Models from Curriculum

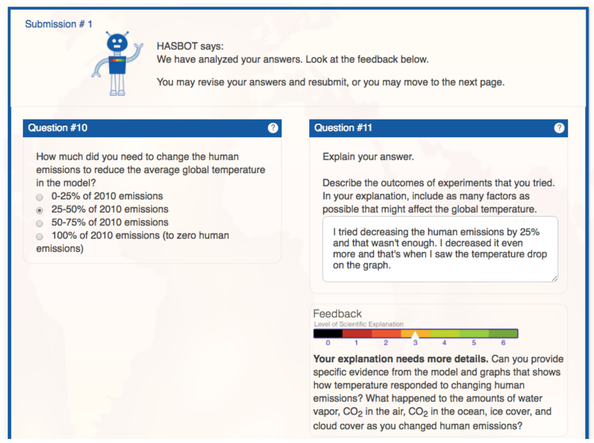




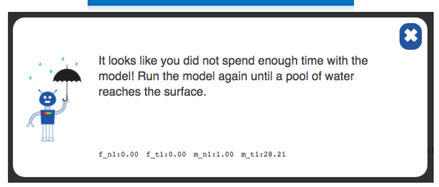




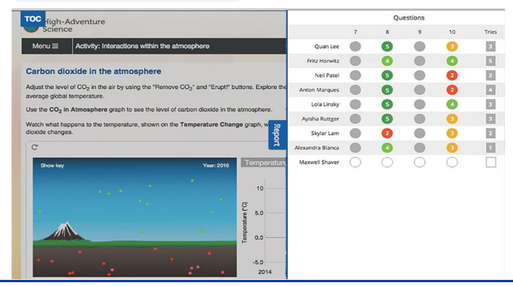
Argument Question & Feedback



Model-based Feedback



Teacher Dashboard



Left: A teacher moves through the classroom with the dashboard running on a tablet. She clicks to see student scores, written responses and revision attempts.

Study 3: Students who used HASbot feedback improved their ability to write scientific arguments with uncertainty.

- ❖ 343 middle and high school students taught by 9 teachers across 7 states used the Water Module with the contextualized feedback from 2016 to 2017.
- ❖ Students made statistically significant gains from pretest to posttest on all four argument elements.

Argument element	Pretest		Posttest		Effect size Cohen's d
	Mean	SD	Mean	SD	
Claim	1.78	0.73	2.13	0.70	0.49***
Explanation	5.29	1.55	6.73	1.77	0.87***
Uncertainty rating	7.39	2.36	9.79	2.09	1.08***
Uncertainty attribution	4.31	2.17	7.35	2.18	1.39***
Total	18.79	4.47	26.01	4.76	1.52***

Note: *** $p<.001$. p-values for effect size were obtained from paired t-tests results

- ❖ The significant predictors of post-test argumentation score include pre-test argumentation score, full completion of argumentation tasks, and number of revisions after HASbot feedback.

Predictor variables	Unstandardized coefficient (B)	Standard error (SE)	t
Constant	15.02***	1.67	8.98***
Pre-test argumentation score	0.42***	0.05	7.68***
Female	0.83	0.51	1.65
English as second language	0.28	0.73	0.38
Computer use for science learning	0.17	0.53	0.31
Full completion of argumentation tasks	2.12*	1.09	1.94*
Number of revisions after HASbot feedback	0.26*	0.11	2.45*

Note: *** $p<.001$, * $p<.05$

From video analysis of 14 student groups working on the Water Module, we identified how HASbot supports and hinders student learning of scientific argument writing.

HASbot helped students to:
- ❖ determine what information to include and how to revise argument responses
- ❖ be motivated to revise with feedback from a friendly, non-judgmental robot
- ❖ frame how to talk about uncertainty as part of argumentation
- ❖ engage more deeply with the content and the data

HASbot constrained students because:
- ❖ false positive machine scores hindered students' revision efforts
- ❖ some students had difficulty interpreting the feedback statements
- ❖ repetitive feedback statements irritated some students when their revisions did not yield different score

Broader Impact

Since the project was awarded in Fall of 2014, the project has achieved:

Benefits for 1187 students and 14 teachers

Two publications in journals:
- ❖ Mao, L., Liu, O. L., Roohr, K., Belur, V., Mulholland, M., Lee, H. –S., & Pallant, A. (2018). Validation of automated scoring for formative assessment of students' scientific argumentation in climate change. *Educational Assessment, 23*(2), 121-138.
- ❖ Zhu, M., Lee, H. –S. Wang, T., Liu, O. L., Belur, V., & Pallant, A. (2017). Investigating the impact of automated feedback on students' scientific argumentation. *International Journal in Science Teaching (NARST)*, *39*(12), 1648-1668.

Four peer-reviewed publications in conference proceedings:
- ❖ 2 Computer-Supported Collaborative Learning conference (CSCL)
- ❖ 1 IEEE Integrated STEM Education Conference (ISEC)
- ❖ 1 IEEE International Conference on Data Mining (ICDM)

Ten conference presentations:
- ❖ 2 American Educational Research Association (AERA)
- ❖ 1 National Association for Research in Science Teaching (NARST)
- ❖ 2 National Council on Measurement in Education (NCME)
- ❖ 4 Northeastern Association for Research Association (NERA)
- ❖ 1 Society for Text & Discourse

Three papers are currently in progress for journal submission.

Lydia Liu, Ph.D., Educational Testing Service, lliu@ets.org Amy Pallant, The Concord Consortium, apallant@concord.org Hee-Sun Lee, Ph.D., The Concord Consortium, hlee@concord.org
Grant number 1418019
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation