HERA: Exploring the Power of Adaptive Scaffolding on Scientific Argumentation and Modelling Competencies in Online Learning Systems

Yigal Rosen, Meirav Arieli-Attali, Sue J. Ward, Jennifer Seery, Vanessa Simmering, Laurel Ozersky, Kristin Stoeffler, Kacy Webster, and Alina von Davier
yigal.rosen@act.org, meirav.attali@act.org, sue.j.ward.@act.org, jennifer.seery@act.org, vanessa.simmering@act.org, laurel.ozersky@act.org, kristin.stoeffler@act.org, kacy.webster@act.org, alina.vondavier@act.org
ACT, Inc.

Abstract: Scientific argumentation addresses how arguments are structured from evidence and is a central process in scientific enquiry and modeling. Although scientific argumentation is considered one of the critical competencies in the U.S. Next Generation Science Standards and internationally there is lack of effective and scalable adaptive online learning systems for students to practice the competencies, including mechanistic reasoning, across domains of scientific knowledge and contexts. In this paper we introduce an experimental framework aimed to facilitate the exploration of the effects of adaptive scaffolding in online learning systems on students' ability to develop mechanistic reasoning in scientific argumentation. We describe findings from a pilot study and discuss next steps in research and development of adaptive scaffolding in online learning systems.

Scientific argumentation and modelling

Scientific argumentation is an important competency for scientifically literate laypeople as well as scientists. This competency requires not only understanding how data or evidence could support a specific claim or conclusion, but also evaluating multiple possible interpretations to devise a compelling argument for one explanation over another. Engaging in scientific argumentation involves combining content, procedural, and epistemic knowledge, and supports both learning new scientific knowledge and participating in scientific debates (Osborne et al., 2016). Mechanistic reasoning is an important component of scientific argumentation (Odden & Russ, 2019), requiring thinking about the potential cause of a phenomenon. Rather than stating a description of the events that took place, mechanistic thinking involves following the causal chain of events that give rise to the observations.

Incorporating modelling into science education is an effective method to teach students about the nature of science (Berland et al., 2016). Students who understand scientific models know that models are a representation or an abstraction of a phenomenon that details the hypothesized reason for why a phenomenon occurs (Namdar & Shen, 2015). Models engage students' mechanistic reasoning by encouraging them to articulate explicitly their hypothesized causal chain that leads to the phenomenon (Krist et al., 2019). Additionally, by being simplified representations of the phenomenon at hand, students need to use systems thinking to identify the most relevant factors among a myriad of potential variables (Yoon et al., 2018). The ubiquity of scientific models in all fields of science also makes them an important cross-cutting skill, as it might help students integrate different topics in science (Park et al., 2017).

Principled approach for online learning system design

The integration and application of Evidence Centered Design (ECD) and Universal Design (UD) increases the likelihood that learning and assessment activities will be well aligned with the competency model and targeted knowledge areas, while also being accessible to all students, including students with disabilities. The advantage of following an integrated ECD+UD-principled design is particularly evident when the goal is to develop and assess complex competencies such as scientific argumentation and modelling by using complex performance tasks (e.g., multidimensional tasks such as simulations or games). It is important to explicitly identify how the relevant competencies and behaviors are connected because the complexity of the focal competencies and/or the rich data the tasks provide might pose difficulties in making inferences from behaviors to competencies.

ECD formulates the process of test development to ensure consideration and collection of validity evidence from the onset of the test design (Mislevy et al., 2006). ECD is built on the premise that a test is a measurement instrument, with specific claims about the test scores are associated, and that a good test is a good match of test items and test takers' skills. The ECD Framework defines several interconnected models, including competency or student model, evidence model, task model and assembly model. Using ECD as an organizing framework for learning and assessment can help address important design questions, namely: Which constructs or processes does each task reveal? Do the proposed scoring methods effectively recognize and interpret the evidence generated by students' responses and interactions? How is all of the evidence that is generated by students' choices synthesized across multiple tasks? Is all the evidence for a particular construct comparable when different students attempt different tasks? How do learning and assessment relate within an integrated system? To address these types of questions, Arieli-Attali et al. (2019) described an expanded framework (termed e-ECD) that included the specification of the relevant aspects of learning at each of the three core models in the ECD, as well as making room for specifying the relationship between learning and assessment within the system. The proposed framework does not assume a specific learning theory or particular learning goals, rather it allows for their inclusion within an assessment framework, such that they can be articulated by researchers or assessment developers that wish to focus on learning.

UD emphasizes the importance of addressing accessibility for the broadest range of potential users during the initial stages of learning and assessment design and throughout the development and implementation of the assessments. The use of UD principles creates flexible solutions because, from the start, designers consider the diverse ways in which individuals will interact with the learning and assessment activities. When sources of construct-irrelevant variance are identified using an ECD approach, the application of UD principles can guide the incorporation of appropriate options for how students interact within the system. In this way, ECD works synergistically with UD. By considering multiple means of perception, expression, cognition, language and symbol use, executive functioning, and engagement, the application of UD in the ECD process can accommodate individual differences in how students recognize, strategize, and engage in learning and assessment situations. This synergistic process minimizes the unintended negative influence that accessibility needs may have on student performance and maximizes the opportunities for students to show what they know and can do.

Developing HERA: Holistic Educational Resources and Assessment system

We used this combined ECD+UD approach to develop an adaptive learning system for scientific argumentation and modelling for middle school students, called HERA. In addition to the ECD+UD framework, HERA builds on ACT's holistic framework of cognitive skills (Camara et al., 2015) and NGSS aligned. The system combines computer-based science simulations with adaptive learning supports and performance tasks to elicit changes in knowledge, skills, and abilities through the course of each interactive lesson. The model also includes gamification features such as a "cost" for use of learning supports and points earned for correct responses, to promote learner effort and engagement. An initial pilot has been completed on Amazon Mechanical Turk as a proof of concept focused on functionality and usefulness of learning supports, with those supports presented in a variety of ways. We are now adapting the HERA system as a prototype for testing in middle-school classrooms on a customized OpenEdX experimentation platform.

At the beginning of each HERA lesson, the learner can experiment with a simulation to enact constructivist principles of learning by doing while building new or enhanced understanding of the scientific concept. This strategy also elicits intrinsic motivation, as the student explores the concept and learns autonomously (Ryan & Deci, 2000; van Roy & Zaman, 2017). The learner is then presented with a series of assessment items that include three levels of metacognitive learning supports which can be activated upon an incorrect response. As part of our development of the HERA system, we developed an item model that can be used to collect evidence for both assessment and learning, termed an Assessment and Learning Personalized Interactive item (AL-PI). This item looks like a regular assessment item, and only after an incorrect response, the learners are given "learning options" to choose from. We offer three types of learning supports: (1) Rephrase – rewording of the question; (2) Break-it-down – providing the first step out of the multi-steps required to answer the question; and (3) Teach-me – providing a text and/or visual explanation of the background of the question. Figure 1 presents a screenshot of an AL-PI item from a task about the force a door exerts on its hinges; targeting the skill of selecting data.

The item in Figure 1 also shows the iteration process between the competency and task model. During development, the goal was that each type of help would address the initial, proximal, and distal precursors in skill acquisition. In the final design, the first type of scaffold (rephrase) targets the skill itself, with the goal of reducing irrelevant variance if the student simply did not understand the item. The second type of scaffold (break it down) addresses the proximal precursor, and the third type (teach me) addresses initial, proximal, and distal together. Through these levels of scaffolding, students can receive the most appropriate level of support for their current level. Thus, HERA motivates all learners to explore and evaluate STEM-related phenomena with the help of structured supports. HERA lessons are adaptive to students' level of knowledge and support STEM learning success. Educators have flexibility to implement lessons as best fits the learning needs of their students and learning environment (e.g., whole class activity/demonstration, group or independent interactive lab, or self-paced independent enrichment).

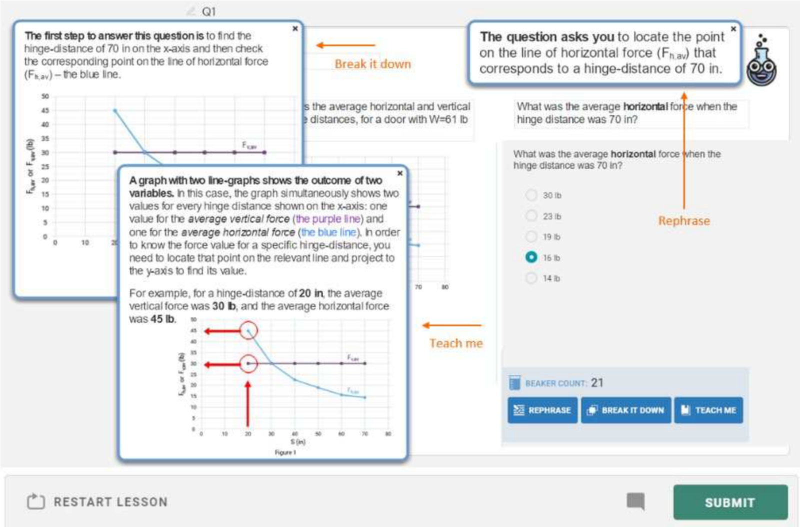

Figure 1. Sample Assessment and Learning Personalized Interactive item (AL-PI) showing the three types of support.

Pilot study

Following tests of system functionality, we conducted a large pilot test of HERA to examine how each of the types of learning supports would be used by real learners. Participants were 2,775 Amazon Mechanical Turk workers; each was assigned to one of five conditions and completed three lessons. Lesson topics were restitution, specific heat, Hooke's law, Beer's law, hinges, and self-pollination. Each lesson included ten items with learning supports as illustrated in Figure 1. Learners could choose between the three learning supports described above (rephrase, break-it-down, teach me); across conditions, these supports were either not offered (Condition 1), offered before the learner answered (Condition 2) or offered after the learner answered if the initial attempt was incorrect (Conditions 3-5). These latter conditions were differentiated by cost: no cost for the supports, differential costs (one, two, or three 'beakers' out of a bank of 30, for Rephrase, Break-it-down, and Teach Me, respectively), or equal costs (two beakers each out of a bank of 30) per support.

We first calculated reliability of each lesson in each condition across the 10 items for the first attempt, not including data from the simulation (as simulations were not scored). Across all lessons and conditions, reliability was Cronbach's $\alpha = 0.76$ on average (range = 0.68-0.82); this confirms that the first attempt on items with scaffolds is equivalent to a typical assessment data. We next compared preferences for type of support, which showed that Teach Me was the most preferred ($M = 45\%$ across conditions 2-5), followed by Break-It-Down ($M = 25\%$), then Rephrase ($M = 15\%$). When scaffolds were offered before answering, learners used more help ($M = 89\%$, compared to 83-87 in other conditions), especially Teach Me ($M = 49\%$, compared to 42-44%). Comparing Conditions 3-5, when there was a cost learners used less help (3-4% less than when no cost). When costs were differential, there was slightly less use of the 'expensive' supports relative to the condition with equal costs. These results show that learners will use the scaffolds differentially, and that timing of the offer seemed to matter more for use than the costs. It is promising to see that learners did not default to using the highest level of support and even chose no support for some items; this relieves potential concern of 'bottom-out hinting' in which learners will put in minimal effort and make maximum use of hints provided by a learning system (e.g., Razzaq & Heffernan, 2010).

Next steps

The success of HERA and similar learning systems in science education hinges on the AL-PIs and adaptive learning architecture that allows experimentation with scaffolding methods and optimization. In order to effectively and immediately engage learners, the quality and purpose of the initial simulations is critical. The included simulations are carefully designed to better support learning by focusing the simulation content to eliminate extraneous details. Additionally, simulations enable item designers to decompose the interaction into more deliberate and manageable pieces in order to not overwhelm the learner. Furthermore, the decomposition provides an opportunity to pair the assessment items and resulting data on a more granular level. Another benefit of using simulations is they are not limited by the physical constraints of a real-world system. For instance, students can explore a skatepark environment without friction.

HERA is currently being deployed on a custom Open edX instance running an Amazon Web Services backend to meet these needs. Within the instance, a component known as "flow control" enables item designers to specify how the system responds to various user inputs. Depending on the learner's metacognitive support choice, the program will redirect accordingly. This functionality manages the learner support options (Rephrase, Break-It-Down, or Teach Me). In future applications, flow control will manage which resources the learner sees and when. This model provides a learner-centric approach to building scientific reasoning skills in a personalized way. Adaptive learning architecture provides the foundation for future experimentation with different scaffolding strategies in HERA learning system and effective empirical examination of the effects on students' scientific argumentation and modelling competencies. Future studies will leverage the robust adaptive learning architecture and Open edX data pipeline to explore collaborative problem solving in Science (Rosen, Wolf, & Stoeffler, 2019) and the effects of various learning resources on transfer of knowledge and learning motivation in Science (Rosen, 2009). Customized dashboards quickly highlight the effectiveness of various interventions, track student engagement over time and across items, and identify assessment items needing refinement. The depth, specificity, and reporting options of data in the platform provides opportunities for future development of additional metacognitive features.

References

Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., & von Davier, A. A. (2019). The Expanded Evidence-Centered Design (e-ECD) for Learning and Assessment Systems: A Framework for Incorporating Learning Goals and Processes Within Assessment Design. *Frontiers in Psychology*, *10*.

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in Practice: Making Scientific Practices Meaningful for Students. *Journal of Research in Science Teaching*, *53*, 1082–1112.

Camara, W., O'Connor, R., Mattern, K., & Hanson, M. A. (2015). *Beyond Academics: A Holistic Framework for Enhancing Education and Workplace Success. ACT Research Report Series. 2015 (4).* ACT, Inc. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2015-4.pdf

Krist, C., Schwarz, C. V., & Reiser, B. J. (2019). Identifying Essential Epistemic Heuristics for Guiding Mechanistic Reasoning in Science Learning. *Journal of the Learning Sciences*, *28*(2), 160–205.

Mislevy, R. J., Steinberg, L. S., Almond, R. G., & Lukas, J. F. (2006). Concepts, Terminology, and Basic Models of Evidence-Centered Design. In D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Eds.), *Automated Scoring of Complex Tasks in Computer-Based Testing* (pp. 15–47). Lawrence Erlbaum Associates, Inc.

Namdar, B., & Shen, J. (2015). Modeling-Oriented Assessment in K-12 Science Education: A synthesis of research from 1980 to 2013 and new directions. *International Journal of Science Education*, *37*(7), 993–1023.

Odden, T. O. B., & Russ, R. S. (2019). Vexing Questions that Sustain Sensemaking. *International Journal of Science Education*, *41*(8), 1052–1070.

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S.-Y. (2016). The Development and Validation of a Learning Progression for Argumentation in Science. *Journal of Research in Science Teaching*, *53*(6), 821–846.

Park, M., Anderson, E., & Yoon, S. (2017). *Learning Biology Coherently Through Complex Systems, Scientific Practices, and Agent-Based Simulations*. Philadelphia, PA: International Society of the Learning Sciences.

Razzaq, L., & Heffernan, N. T. (2010). Hints: Is It Better to Give or Wait to Be Asked? In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems* (Vol. 6094, pp. 349–358). Springer Berlin Heidelberg.

Rosen, Y. (2009). The Effects of Animation-Based Online Learning Environment on Transfer of Knowledge and on Motivation for Science and Technology Learning. *Journal of Educational Computing Research, 40*(4), 439-455.

Rosen, Y., Wolf, I., & Stoeffler, K. (2019). Fostering Collaborative Problem Solving Skills in Science: The Animalia Project. *Computers in Human Behavior*, *104*.

Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Sevelopment, and Well-Being. *American Psychologist*, *55*(1), 68.

van Roy, R., & Zaman, B. (2017). Why gamification Fails in Education and How to Make it Successful: Introducing Nine Gamification Heuristics Based on Self-Determination Theory. In *Serious Games and edutainment applications* (pp. 485–509). Springer.

Yoon, S. A., Goh, S.-E., & Park, M. (2018). Teaching and Learning About Complex Systems in K–12 Science Education: A Review of Empirical Studies 1995–2015. *Review of Educational Research*, *88*, 285–325.

