Apt Epistemic Practices in a High School Science Classroom: A Contrastive Case Study

Huma Hussain-Abidi, Clark A. Chinn
hh429@rutgers.edu, clark.chinn@gse.rutgers.edu
Rutgers University, New Brunswick, NJ

Susan A. Yoon, Amanda M. Cottone, Noora F. Noushad, Thomas Richman
yoonsa@upenn.edu, amandaco@upenn.edu, noora@gse.upenn.edu, trichman@upenn.edu
University of Pennsylvania, PA

Abstract: In the current “post-truth” era, there is a growing need to promote apt epistemic practices in science education. In this study, we investigated two high-school biology students’ epistemic practices during a modeling unit and appraised them for aptness using the Apt-AIR framework. Additionally, we analyzed their responses to a post-implementation focus group interview, designed to elicit their metacognition regarding epistemic practices, as they answered probing questions about practices within the curriculum and reflected on video clips of other students engaging with the units. We document the epistemic practices that students engage in during a modeling unit and evaluate the extent to which they are apt. Findings suggest a disassociation between students’ cognitive engagement in modeling practices and their metacognitive understanding.

Introduction and Theoretical Background

In the current “post truth” era, science denialism and the spread of misinformation are prevalent, and science education has failed to meet these challenges (Chinn et al., 2020). The widespread rejection of well-justified scientific consensus on matters such as COVID- 19 risks, vaccination safety, and climate change demonstrates the urgent need to develop and implement instructional strategies that enable students to accurately appraise scientific information (Barzilai & Chinn, 2018; Chinn et al., 2020; Duncan et al., 2018; Kavanagh & Rich, 2018). To address these challenges, we examined an existing high-school science learning environment that seeks to promote apt epistemic practices in science classrooms. Our approach builds on Barzilai and Chinn’s (2018) analysis of apt epistemic performance, defined as successful performance (e.g., reaching an accurate conclusion about vaccination safety) that is produced through competence (e.g., one’s ability to evaluate the evidence, identify scientific consensus, and integrate the relevant information). Apt epistemic performances are further unpacked in terms of:

- *Aims* (e.g., aiming to reach an accurate conclusion)
- *Ideals* or standards for evaluating whether an aim has been achieved (e.g., fit with relevant evidence and the consensus of experts as ideals for determining whether an accurate conclusion has been reached)
- *Reliable processes* (RPs) that are used to achieve the aims (e.g., evaluating multiple sources of information, evaluating the expertise of sources, and so on)

Apt epistemic practices (AEPs) are clusters of interrelated apt epistemic performances that are governed by shared community norms. Examples of such clusters of AEPs are constructing scientific models, running and interpreting simulations, and engaging in argumentation. The various aims, ideals, and RPs used in these practices are shared by a community; for example, a community can share the aim of reaching well-justified conclusions, adopt ideals such as fit with ample evidence, and employ processes such as gathering and weighing large numbers of studies to ensure ample evidence. Our goal is to promote growth in AEPs that can be used inside classrooms in inquiry and that can be adapted for use outside classrooms to evaluate scientific claims encountered in the contemporary information society.

The first objective of this study is to examine the interrelations between cognitive and metacognitive aspects of epistemic practices. AEPs encompass both cognitive uses of relevant aims, ideals, and processes (e.g., members of a community regularly conduct multiple experimental trials to improve the reliability of data) and a metacognitive understanding of these aims, ideals, and processes (e.g., understanding as a community that multiple trials are needed to avoid being misled by outliers) (Barzilai & Chinn, 2018). Both cognitive and metacognitive aspects are important to epistemic competence (Ford, 2008; Hofer, 2005; Barzilai and Zohar, 2012, 2014). Although some research suggests close connections between metacognitive and cognitive aspects of performance (e.g., Barzilai & Zohar, 2014), other research suggests that associations between cognitive performance and metacognitive understanding is low (e.g., Van der Stel & Veenman, 2010). Yet there are reasons to think that metacognitive understanding might help support improved cognitive performance. For example, if students metacognitively grasp that multiple trials are needed in experimentation to establish stable, average effects, then the students may be more likely to use such practices themselves and to notice if others do not use them. Accordingly, this study uses a comparative case study approach to examine the relationships between students' metacognitive understanding of AEPs and their cognitive use of these practices in their own inquiry.

A second goal of this study is to contribute to the field's understanding of what growth in epistemic practices looks like during inquiry. There is a need to develop more comprehensive accounts of what epistemic growth involves as well as how to promote it (Barzilai & Chinn, 2018). In this paper, we seek to apply the Apt-AIR framework developed by Barzilai and Chinn (2018) to understand students' epistemic practices, and thus explore the affordances of this framework in identifying epistemic practices that students engage in, and how these epistemic practices might be improved. This approach is line with other approaches to define and promote epistemic growth (Berland et al., 2015; O'Neill & Polman, 2004; Osborne et al., 2003). Thus, we employ the Apt-AIR framework as a lens to characterize epistemic growth in terms of aims, ideals, and RPs, and we ultimately intend to promote growth in aims, ideals, and RPs in science classes.

The focus of this study was to examine the epistemic practices of students using the Apt-AIR lens, and to examine the extent to which students' metacognitive understanding of epistemic practices was similar to the epistemic practices they demonstrated in their own inquiry. To address these goals, we used a comparative case study method: We investigated two students' epistemic practices while they participated in a widely accessible, agent-based modeling curriculum titled BioGraph, in which students experiment with and learn about complex biological systems (Yoon et al., 2017; Yoon at al. 2018). Additionally, we analyzed post-implementation focus-group interviews in which students metacognitively identified ideals and RPs. We address the following research questions: (1) What ideals and RPs did the students discuss in the focus-group interviews? (2) What ideals and RPs did students demonstrate in the classroom as they participated in the BioGraph unit? (3) What was the relationship between the cognitive classroom performance of these students and what they metacognitively identified in their post-implementation focus group interview?

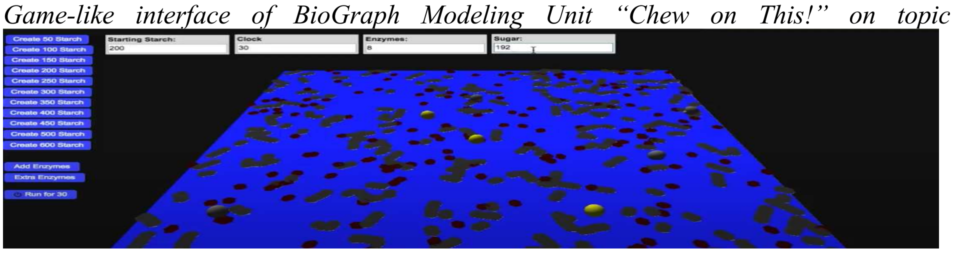
Methods

Context

In 2013, BioGraph researchers conducted an exploratory study in which students worked in groups of 2 to 4 with two to three units of the BioGraph curriculum, which offers five biology units on the topics of Genetics, Evolution, Ecology, Sugar Transport, and Enzyme Kinetics. In the units, students worked with agent-based simulations that combine graphical blocks-based StarLogo programming with a 3-D game-like interface (see Figure 1). The curricular materials for each unit took 2–3 days to complete and included popular and academic literature about complex systems as well as short movies, PowerPoint presentations, and teacher and student activity guides (Yoon et al., 2016; Yoon et al., 2017).

In the present study, we used two sources of data: students' classroom interactions while actively participating in the BioGraph unit *"Chew on This!"* from the 2020-2021 school year. In this unit, the learning objectives were: (1) Recognize that enzymes speed up chemical reactions that take place in cells, (2) Explain the shapes of the reaction rate curves both with and without enzymes, (3) Understand that enzymes and substrates interact with one another due to random (not directed) motion of molecules. Additionally, some students volunteered to participate in a focus group interview, post implementation. According to teachers, the students in the focus groups represented a broad range of students. In the focus group interview, we used for some questions previously collected video footage from the BioGraph research project from 2013. The clips included students interacting with the *"Chew on This!"* unit and the *"Gene Regulation"* unit. There were minimal changes made from the 2013 modeling curriculum to 2020-2021; thus, the clip content was familiar to the students.

Figure 1
Game-like interface of BioGraph Modeling Unit "Chew on This!" on topic of Enzyme Kinetics



Participants

In this comparative case study, we focus on two students—pseudonyms Nate and Hannah—who were ninth grade, accelerated biology students from a public high school in northeastern USA. Both students were in the same biology class and were in the same group of four students while they participated in the BioGraph unit. Nate was a male student, and Hannah was a female student. Both identified as White, non-Hispanic. Nate was a vocal member of the group while he engaged with his groupmates as they worked through the BioGraph unit as well as when researchers asked questions in the focus group. Hannah made fewer verbal contributions than Nate in both formats yet participated relatively frequently in the focus group. Nate and Hannah were selected for this case because of their participation within the same group both in class and in the focus group.

Data Sources and Analysis

We analyzed classroom data from February 2021, when the class was virtually conducted over Zoom due to the COVID-19 pandemic. Nate and Hannah were in the same student group participating in an Enzyme Kinetics unit titled "Chew on This!" with two other students. We video recorded the students as they engaged in inquiry with the modeling unit for one class period, or roughly 30-minutes. Additionally, we conducted a 60-minute post-implementation focus group interview in May 2021. In the interview, we prompted students to examine and reflect metacognitively on scientific practices. The questions were designed to elicit comments about the ideals they endorsed for good models (whether normative or not) and the processes they viewed as reliable for engaging with simulations on complex systems (see Table 1). In the focus groups, students were also shown clips from previously collected data of students participating in BioGraph units, and they evaluated and reflected on the epistemic practices demonstrated in the clip.

Table 1
Examples of questions asked by research team and the AIR components (Ideals and RPs)

Questions and goals of questions
What are the most important criteria or characteristics of good scientific models and/or good explanations? Which is most important and why? [This question invites students to express what they view as ideals for good models such as: Models should fit the evidence. Models are understandable. Models have mechanisms that show unseen entities and step-by-step processes.]
When working with BioGraph models, what steps did you or might you take to make your prediction? In other words, how would you justify that your prediction was reasonable when answering a question (i.e., what sources or knowledge would you use to make a prediction)? Do you think it is useful to make a prediction before conducting an experiment? Why or why not? [In response to this question, students might mention ideals such as fit with prior knowledge or the use of good models to make accurate predictions. They might also mention processes such as checking predictions against outcomes, basing their prediction on their last experiment, discussing ideas with their team, etc.]
Students watch a video clip of two students working on a unit on enzymes. In this unit, the students enacted some less RPs , including considering only one prediction, cherry-picking data to match their prediction, using data points as distinct, instead of computing averages Then they were asked: **How would you describe what the students are doing to collect data?** [With this question researchers probed students' ability to metacognitively critique the videorecorded students' use of ideals and RPs. The question provides an opportunity to identify the less RPs evinced in the clip and to suggest what ideals and processes that are better, such as checking for random variation, evaluating evidence consistently, aligning beliefs with evidence, etc.]
Researchers showed some example graphs from the student worksheets to jog memories about how they may have used them. Then asked: **Why do you think this is useful to create graphs when learning about science?** [This question elicits the ideal of emergence—that micro interactions produce macro level patterns which can be depicted through graphs. Students might also suggest processes such as comparing the qualitative pattern on the screen with the graph, which can aid in describing trends and synthesizing information.]
Averaging data and creating graphs are two scientific skills. Can you think of any other scientific skills that you practiced when working with the BioGraph lessons in general? [In response to this question, students might consider RPs such as using evidence to justify a claim, collaborating with peers in inquiry, collecting and recording data, etc.]

Students' ideals and RPs were analyzed based on student discourse in both the classroom inquiry and in the focus group data. (Students seldom mentioned aims explicitly, so we will not analyze them here.) Ideals and

RPs were coded using both top-down and bottom-up methods. As a starting point, previous research on ideals and processes on science tasks provided a guide to possible ideals and RPs that might emerge (Barzilai & Eilam, 2018; Schwarz, et al., 2009; Pluta et al., 2011). Yet the data were coded with the expectation that new ideals and processes would emerge that had not been previously found. The team developed a comprehensive code book through a constant comparative method with multiple iterations, views, and negotiations of the data sources as well as interactional analysis (Jordan & Henderson, 1995). From this code book, Table 2 presents those codes that are most relevant to the data presented in this paper. Note that students sometimes used or advocated processes that could be considered less reliable. For example, for some students, many trial runs of the simulation were needed to establish stable conclusions (more reliable), but for others, two or three trials, or even just one trial (less reliable), was sufficient.

Table 2
Coding List of Ideals and RPs with their Explanations

Ideals	***Explanation***
Fit with ample evidence	Good models fit not only some evidence but a sufficient amount of evidence (e.g., fit with multiple simulation runs rather than just one run).
Fit with systematic evidence	Good models fit not only evidence from single trials, but from careful, controlled experimentation that tests many combinations of variables.
Understandability	Good models are clear and understandable.
Incorporation of underlying biological processes	Good models incorporate underlying, step-by-step biological processes hypothesized to produce the phenomenon.
Fit with prior knowledge	Good models fit with prior well-established knowledge.
Randomness	Good systems models include random variation in their processes.
Successful prediction	Good models can be used to make accurate predictions.
Causal complexity	Good models are complex with many causes and many effects.
Processes viewed or treated as reliable	***Explanation***
Develop arguments with evidence-based reasons	Giving reasons grounded in data/evidence to support ideas.
Consider and test multiple hypotheses	Making predictions based on two or more alternative hypotheses and testing them together.
Consider and test single hypotheses (less reliable)	Making and testing a single prediction.
Run and gather data from multiple trials	Running simulation multiple times to establish average phenomena under the same starting conditions.
Run and gather data from single trials (less reliable)	Running a simulation only once and using the results only from that single run.
Conduct careful experimentation	Planning/running systematic series of controlled experiments carefully contrasting different starting conditions.
Evaluate evidence consistently	Treating information that supports your ideas consistently with information that contradicts your ideas when gathering data (e.g., don't dismiss a single simulation trial that goes against your ideas while accepting a single trial that contradicts your ideas).
Cherry-pick data (less reliable)	Selecting data points in biased ways to confirm prior prediction/beliefs.
Align beliefs with evidence	Revising explanatory models when there are large or even small discrepancies with evidence.
Evaluate evidence based mainly on prior beliefs (less reliable)	Deciding to believe evidence is "correct" to determine whether to keep or discard a data point.
Average data from trials	Calculating the average of data across multiple trials.
Identifying trends	Identifying trends in the data output.
Determine biological processes	Examining simulations carefully to determine underlying biological processes.

Table 2 (continued)

Using prior scientific knowledge	Using scientific knowledge developed or learned previously to evaluate and understand simulation phenomena.
Checking whether agents are behaving randomly	Observing agents' behaviors in the simulations carefully to judge whether the behaviors are partly random.
Explaining using multiple interacting agents	Explaining that multiple interacting agents produce system level outcomes rather than a single leading agent.
Compare simulation phenomena with graphs	Comparing the qualitative pattern on the screen with the graph.
Collaborate with peers	Collaborating with peers and working together to develop a consensus of ideas.
Apply ideas out of school	Applying what is learned in class inquiries to outside-of-school situations.

Results

We present the results in three parts. First, we analyze Nate and Hannah's responses relating to ideals and processes in the focus groups, including their reflections on the presented video clips. Second, we present the ideals and processes that Nate and Hannah used while they participated in the modeling unit. Lastly, we integrate and investigate the results from both contexts to bring forth the relationship between students' cognitive and metacognitive grasp of ideals and RPs.

Ideal and Processes Present in the Focus Group Interview

Analysis of the focus group responses indicate that Nate and Hannah explicitly and implicitly invoked an impressively wide range of RPs and ideals (See Table 3). Nate made multiple references to the ideals *Fit with ample evidence* and *Fit with systematic evidence,* as well as to the RPs *Evaluate evidence consistently* and *Align beliefs with evidence.* He explicitly identified seven of the eight ideals captured in our coding scheme and articulated a wide range of different RPs, as shown in Table 3, which presents several of his comments. Hannah articulated three of the eight ideals: *Randomness, Successful Prediction, and Understandability.* She also identified several RPs. She specifically mentioned the following RPs at least twice: *Run and gather data from multiple trials, Conduct careful experimentation*, *Utilize graphs for insight, Align beliefs with evidence,* and *Evaluate evidence consistently.* Several of her contributions are also highlighted in Table 3. Hence, the two students successfully identified an impressive range of ideals and RPs when they were metacognitively asked to reflect on probing questions and video clips from other students engaged in BioGraph units.

Epistemic practices present during classroom inquiry

In the classroom data, there were two instances in which Nate explicitly engaged with appropriate ideals and RPs. He employed the RP of *Running and gathering data from multiple trials*, as he said, "I clicked for 30 multiple times…" Additionally, he implicitly demonstrated the ideal *Incorporation of underlying biological processes* when considering models to explain the phenomena of enzymes breaking up the starch molecules. He slowed the simulation to make sense of this phenomena before he articulated the underlying processes. Hannah demonstrated use of only one ideal and no RPs in the classroom data. Like Nate, she implicitly demonstrated the ideal of "*Incorporation of underlying biological processes"* by saying that she would slow down the simulation to grasp the biological phenomena to express how enzymes were breaking up the starch molecules.

Nate and Hannah also employed less RPs and less appropriate ideals in the classroom inquiry. Indeed, their inquiry practices included some of the less RPs that they had criticized in the focus group. Both students exhibited *cherry picking, considering and testing a single hypothesis, and evaluating evidence based on prior beliefs*. These practices directly conflict with the normative ideal *Fit with systematic evidence* and the RPs *Evaluate evidence consistently, Consider and test multiple hypothesis* and *Align beliefs with evidence*. For example, Nate endorsed the group's practice of cherry-picking data in comments about trials such as "Don't use the 17" and "I think 54 is very off" (because these were not close to his prediction of 30). Instead of considering all data as evidence, it was clear that Nate and his groupmates (including Hannah) selected data points that justified their favored group hypothesis. Although Hannah did not actively contribute to her group's evidence selection, she also did not push back on the groups' practices, and she reported the same answers as her group in her student activity guide. Thus, both students used less RPs more frequently they used more RPs, and their judgments were guided more by the ideal of fit with prior expectations than by the ideal of fit with evidence.

Table 3
Examples from Transcript of Focus Group with researchers coding of the excerpt

Examples of students' responses to the focus group questions	Ideals/RPs metacognitively discussed
Nate's responses	
Nate: A lot of the different things require ATP for energy, ATP is going to be created using other things. Everything requires certain factors in order to work, and then to create those factors it requires other factors. So all these processes rely on other processes.	*Ideals:* Causal complexity; Incorporation of underlying biological processes *RPs:* Explaining using multiple interacting agents
Nate: The graph there in the Gene Regulation, that's useful. So it's not just that I think there's more of these over time and then they decrease, you can see the graph, you can see it's actually happening. Also, I think the randomness.... In all of these, if you run the simulation multiple times, it's not just the same exact thing. The factors are working off each other with a bit of randomness. You can tell that whatever is happening is actually happening. In the real world, it's not going to be the same every single time. It's more realistic.	*Ideal:* Randomness *RPs:* Utilizing graphs for insight; Explaining using multiple interacting agents.
Nate: And also, one very important thing in science is going to be building off of other scientific research that's already been done. By doing these CER questions, we're able to do this research ourselves and learn about how to use other people's studies that they've conducted to create our own.	*Ideal:* Fit with prior knowledge *RP:* Using prior scientific knowledge
Nate: If you're looking for a specific result, you can make the data match that. For example, as soon as they saw the result that there's 49, they went with that instead of looking at the overall average which disproved their hypothesis.	*Ideal:* Fit with ample evidence *RPs:* Evaluate evidence consistently; Align beliefs with evidence; Average data from trials
Nate: Also, to help you stop conforming your data to your hypothesis, you need to know how many trials you're going to do ahead of time. Otherwise, you might just keep saying, "One more trial" until you get a result that matches what you want to see.	*Ideal:* Fit with systematic evidence *RPs:* Run and gather data from multiple trials; Evaluate evidence consistently; Align beliefs with evidence.
Hannah's responses	
Hannah: It's different every time, so that way when you see it again you don't necessarily get the same data over and over again. That can give you the impression that there is going to be the same data over and over again when of course there wouldn't be. Interviewer 2: Can you explain that a little bit, by "of course there wouldn't be the same data every time"? Hannah: In the case of lactose, for example, if you consume more lactose, there'll be more lactose in your body, as a base point. But if it always came up with 20 lactose, for example, at the beginning, that could give you the impression that there's always going to be 20 lactose at the beginning instead of, you eat more of a dairy product and now you have more lactose in you.	*Ideal:* Fit with ample evidence *RP:* Run and gather data from multiple trials [to deal with the randomness that can affect results from a single trial)
Hannah: We were seeing the same thing, but we had completely different ideas. Interviewer 1: Do you remember what you did to work out the conflicting ideas that you had? Hannah: We just looked at everything else we had said and tried to piece together a step-by-step process that would make sense. That's how we decided which one was correct.	*Ideal:* Incorporation of underlying biological processes *RPs:* Gather and use conclusive, systematic evidence through careful experimentation; Consider and test multiple hypotheses

Table 3 (continued)

Examples of students' responses to the focus group questions	Ideals/RPs metacognitively discussed
Hannah's responses (continued)	
Hannah I was going to say the same thing. It's bad if someone's stuck to one idea even when all the evidence is going against that. Interviewer 1: That's an important point. Interviewer 2: What do you do when that person is stuck on an idea, even if all the evidence is going against that, what would you do to try to convince the other person? Hannah: I would try to explain what the evidence is actually supporting."	*Ideal:* Fit with ample evidence *RPs:* Align beliefs with evidence; Develop arguments with evidence-based reasons
Hannah: If there's a graph provided, I would definitely try to use the quantitative data and identify the trend.	*RPs:* Identify trends, Utilizing graphs for insight
Hannah: I would say the ability to make a prediction is very important in everyday life because there are a lot of times you just need to guess based on what you know. Interviewer 1: … do you mind, if you have one in your head, can you give an example of when you might make a prediction in everyday life? Hannah: If there's rainclouds out, you'd be able to make a prediction that it's going to rain. Then you might have to change your plans…."	*Ideal:* Successful prediction *RP:* Apply ideas out of school

Relationship between the Cognitive Performance and Metacognitive Identification

The findings revealed a disassociation between students' metacognitive grasp of ideals and RPs. Students identified and elaborated on significantly more ideals and RPs when asked to metacognitively reflect on probing questions and critique other students participating in a BioGraph unit in video clips. Even more tellingly, Nate and Hannah both identified and suggested the use of RPs such as *Evaluate evidence consistently, Consider and test multiple hypothesis* and *Align beliefs with evidence*, whereas they violated these very RPs in their own inquiry learning as they engaged in less RPs such as *Cherry-pick data, Consider and test a single hypothesis, and Evaluate evidence based on prior beliefs*. Thus, these students' metacognitive understanding on ideals and RPs surpassed their use of these same ideals and RPs in their own inquiry.

Discussion and Conclusion

In this contrastive case study, two students excelled at metacognitively identifying appropriate ideals and RPs, in contrast to their cognitive engagement with these ideals and RPs while working with the BioGraph unit. Classroom data suggested that students engage in less RPs, whereas in the focus groups, the students expressed impressive metacognitive knowledge of what ideals and RPs are apt. Interestingly, students engaged in the same less RPs in the classroom that they critiqued in the focus group video clips. Thus, they expressed a good metacognitive grasp of AEPs that they did not implement in their own inquiry.

This study expands on Van der Stel & Veenman's (2010) work by showing the dissociation between the cognitive performance and metacognitive understanding in an inquiry task. Further, it provides valuable insights for researchers on what epistemic practices are present in science classrooms by using the Apt-AIR framework (Barzilai & Chinn 2018) to promote educational efforts to encourage better epistemic practices (e.g. Berland et al., 2015; O'Neill & Polman, 2004; Osborne et al., 2003) We acknowledge the limitations arising from our analysis to date of only two students; research with larger samples is needed to see if these initial findings are borne out in analyses of larger samples.

An important implication of this study for the design of inquiry environments is that at least some high school students have a metacognitive grasp of AEPs that designers and teachers can build on in classroom inquiry. Designers and teachers might incorporate metacognitive discussions within inquiry units to help students make the connections between their metacognitive grasp of AEPs and the practices that they are actually using in the classroom. These metacognitive discussions might serve to help students develop publicly endorsed norms for engaging in their own inquiry (cf. Barzilai & Chinn, 2018).

References

Barzilai, S., & Chinn, C. A. (2018). On the goals of epistemic education: Promoting apt epistemic performance. *Journal of the Learning Sciences*, *27*(3), 353–389.

Barzilai, S., & Eilam, B. (2018). Learners' epistemic criteria and strategies for evaluating scientific visual representations. *Learning and Instruction*, *58*, 137-147.

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction, 30*(1), 39–85.

Barzilai, S., & Zohar, A. (2014). Reconsidering personal epistemology as metacognition: A multifaceted approach to the analysis of epistemic thinking. *Educational Psychologist, 49*(1), 13–35.

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2015). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching, 53*(7), 1082–1112.

Chinn C. A., Barzilai, S., & Duncan, R. G. (2020). Education for a "post-truth" world: New directions for research and practice. *Educational Researcher, 52*(1), 51-60.

Chinn, C. A., Barzilai, S., & Duncan, R. G. (2020). Disagreeing about how to know: The instructional value of explorations into knowing. *Educational Psychologist*, *55*(3), 167-180.

Duncan, R. G., Chinn, C. A., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding The Next Generation Science Standards' conceptualization of evidence. *Journal of Research in Science Teaching*, *55*(7), 907-937.

Ford. (2008). Disciplinary authority and accountability in scientific practice and learning. Science Education (Salem, Mass.), 92(3), 404–423.

Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. *Learning and Instruction, 39*(5), 11–22.

Hofer, B. K. (2005). The legacy and the challenges: Paul Pintrich's contributions to personal epistemology research. *Educational Psychologist, 40*(2), 95–105.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*(1), 39 – 103

Kavanagh, J., & Rich, M. D. (2018). *Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life*. Rand Corporation.

O'Neill, D. K., & Polman, J. L. (2004). Why educate "little scientists"? Examining the potential of practice-based scientific literacy. *Journal of Research in Science Teaching, 41*(3), 234–266.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "ideas-about-science" should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching, 40*(7), 692–720.

Pluta, W. J., Chinn, C. A., & Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching, 48*(5), 486–511.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Schwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*, 632-654.

Van der Stel, & Veenman, M. V. (2010). Development of metacognitive skillfulness: A longitudinal study. *Learning and Individual Differences, 20*(3), 220–224.

Yoon, S., Klopfer, E., Anderson, E., Koehler-Yom, J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., Oztok, M., Evans, C., & Goh, S. (2016). Designing computer-supported complex systems curricula for the Next Generation Science Standards in high school science classrooms. *Systems*, *4*(38), 1–18.

Yoon, S., Goh, S., & 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*(2), 285–325.

Yoon, S., Anderson, E., Koehler-Yom, Evans, C., Park, M., J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., & Klopfer, E. (2017). Teaching about complex systems is no simple matter: Building effective professional development for computer-supported complex systems instruction. *Instructional Science, 45*(1), 99–121

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants No. 2009803 and 1019228. 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.

