THE DEVELOPMENT OF AN ASSESSMENT OF SCIENTIFIC ARGUMENTATION ABOUT ECOLOGY

Anna MacPherson
Stanford University

Introduction

Early in the item development phase of this project, I took a set of four argumentation tasks to several ecologists so that they could evaluate the accuracy of the scientific content and offer their opinions of the validity of the tasks for measuring students' argumentation. The tasks were subjected to an enthusiastic round of "shredding" by the scientists and emerged with less over-simplification and more precise language, though they were certainly far from perfect. One ecologist, an assistant professor at Yale, remarked:

> These kids are getting better science education than I was in high school. I don't think anybody ever talked to me about claims and evidence and how to build an argument in a science class ever. I mean, we just had to memorize chemistry equations, right? I don't think I could have done a very thorough job of [building an argument] in high school.

Though I was encouraged that a university professor of science thought that the materials being developed for my dissertation were "better" than what she remembered being a part of her own education, I had to let her in on the truth: kids do not really hear much about claims and evidence in science class; it is still mostly about memorizing chemistry equations. Practices (such as argumentation) that are of great interest to researchers in education do not necessarily make their way seamlessly into classrooms. Just because my dissertation is about the development of materials to support argumentation does not mean that argumentation is a topic that receives any attention in classrooms around the country. And, perhaps more to the point, there is little evidence that science education is getting any "better" at all.

Assessments may be one way in which to shift science teaching toward practices that align with the new Framework for K-12 Science Education (NRC, 2012) and the Next Generation Science Standards (Achieve, 2013). Teachers use assessments to read the values of the science curriculum (Sunal & Wright, 2006) and they rate state assessments as an important influence on how they choose the content to teach (Au, 2007; Hannaway & Hamilton, 2008; Weiss, Pasley, Smith, Banilower & Heck, 2003). Scholars in the fields of science education and educational policy agree that implementing better assessments, specifically ones that focus on problem-solving and critical thinking, could improve the quality of science education. Shavelson and Baxter (1991) argue that changing the tests may have a direct impact on how science is taught. According to them, "Teaching to tests that reward diverse problem solving . . . may directly influence teachers' instructional decisions" (p. 81). A decade later, in a paper entitled "Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking," Yeh (2001) states the argument even more explicitly:

> If state-mandated tests focused on critical thinking, rather than rote recall of factual knowledge, teachers who feel pressured to teach to the test could focus on teaching critical thinking rather than the universe of items that students might otherwise be asked to recall. (p. 12)

However, being able to use assessments as levers to improve instruction requires the *existence* of such assessments; specifically, assessments that measure the types of complex competencies we want to develop in students.

This dissertation describes the development of an assessment of scientific argumentation, one of eight practices identified by the new Framework for K-12 Science Education (NRC, 2012) and the Next Generation Science Standards (Achieve, 2013). A significant program of work undertaken within the science education research community supports the idea that argumentation should be a key feature of the learning and teaching of science (Driver, Newton, & Osborne, 2000; Erduran & Jimenez-Aleixandre, 2008; Khine, 2012; Kuhn, 1993). The authors of the new Framework defend the importance of the process of argumentation by pointing to its centrality to the nature of the scientific enterprise. Argumentation is at the core of how scientific theories gain acceptance—scientists are constantly engaged in the defense of their ideas, a process which happens informally during collaboration, lab meetings and symposia and, formally, as part of the process of peer review (Latour & Woolgar, 1986). As Crombie (1996) states, “The history of science is the history of argument” (p. 12.). Furthermore, argumentation is seen as a useful practice for citizens engaged in everyday negotiation of scientific claims. As the authors of the new Framework explain, “Becoming a critical consumer of science is fostered by opportunities to use critique and evaluation to judge the merits of any scientifically based argument” (NRC, 2012, p. 71). In short, scientific argumentation is seen as an essential practice for scientists and citizens alike.

In this dissertation, I argue that the way we *assess* argumentation will ultimately have the greatest impact on the way that it is taught and learned. The process of developing innovative assessments of complex competencies such as argumentation is fraught with challenges (Gorin & Mislevy, 2013) and the field currently lacks high-quality exemplars (Pellegrino, 2012; Songer & Ruiz-Primo, 2012). The work proposed here will address these challenges in the particular content domain of *ecology*. Ecology is the scientific study of interactions among organisms and their environment. The NGSS defines the science curriculum in terms of "disciplinary core ideas", eight scientific practices, and a set of "cross-cutting concepts", and thus does not include a particular section devoted to a field called ecology; however, the concepts that underlie the field of ecology are all part of the new standards. These concepts include the idea that energy flows and nutrients cycle through an ecosystem, interactions between species can maintain the equilibrium in the system, and human activities can disrupt the balance of the system.

Why develop assessment materials for argumentation in ecology? One of the most compelling reasons to conduct research on the assessment of student understanding in the domain of ecology is that *all* students will learn about ecology during their schooling. Unlike chemistry or physics, which in some schools are considered "advanced" sciences, reserved only for the most able students, practically every high school student takes biology. Very nearly all (98%) of high schools offer a biology course, biology is available to the most number of students, and lower ability students are more likely to take biology than chemistry or physics (Banilower et al., 2013). Thus, an understanding of the practice of argumentation within the field of ecology will inform the development of materials that would impact the largest number of American high school science students.

Article I sought to describe the arguments that ecologists engage in as part of their professional work. The purpose of the investigation was to inform the development of items to assess argumentation in the content domain of ecology. Seven ecologists, working in empirical or theoretical research, in sub-fields ranging from individual/behavioral ecology to global ecology, participated in semi-structured interviews. Interview transcripts were coded using a framework for analyzing arguments (Erduran, Simon & Osborne, 2004; Toulmin, 1958) as well as open coding to identify emergent themes (Strauss & Corbin, 1998). Ecologists' descriptions of their arguments were compared with a set of published argumentation tasks designed for middle and high school students. Ecologists' descriptions differed from the school argumentation tasks mainly in terms of the types of scientific questions and claims offered. Whereas the research ecologists all mentioned constructing and critiquing *causal claims* (i.e. claims that attempt to offer an underlying cause for natural phenomena), school argumentation tasks situated in the domain of ecology asked students to construct and critique *descriptive claims* (i.e. claims that describe natural phenomena) and *prescriptive claims* (i.e. claims about human-decision-making for the environment) only. The lack of materials addressing causal claims raises questions: why are these types of claims absent from school science, and how can we design argumentation tasks that ask students to construct and, importantly, *critique* such claims? An explanation for the mismatch, as well as proposed material for argumentation about causal claims in ecology is offered.

Article II describes the development of an assessment designed to measure students' competence with argumentation in the particular biological field of ecology. Assessments have been shown to be powerful influences on what is taught and how it is taught in science classrooms (Au, 2007). Thus, the motivation for this work is that by developing a set of tasks that "operationalize" argumentation, teachers and students will have a powerful model for what argumentation should look like. The Construct Modeling approach (Wilson, 2013) was utilized to develop a set of four tasks to assess students' competence with argumentation. An existing construct map for scientific argumentation (Osborne, et al., in review) guided item development. Ecological scenarios in which to situate argumentation were developed through interviews with scientists (discussed in Article I). Think-aloud interviews were conducted with twenty high school students to identify areas of construct-irrelevant variance and to gather basic validity evidence. A scoring guide was developed and moderated based on pilot assessment data. The final set of four argumentation tasks (called "Moose," "Invasive," "Leaves," and "Grass") were administered to a sample of 203 biology students in New York City public high schools. Assessment data were analyzed for evidence of reliability and validity via traditional test statistics (Cronbach's coefficient alpha for internal reliability, Cohen's kappa for inter-rater reliability, classical difficulty and discrimination) and Rasch analysis. Findings indicated acceptable levels of reliability and as well as basic construct validity evidence. A multi-dimensional "testlet" Rasch model (Wang & Wilson, 2005) did not fit the data significantly better than a unidimensional, partial credit Rasch model (Masters, 1982) indicating that the argumentation dimension is perhaps more salient than the four separate content dimensions. Think aloud interviews showed that students who were more successful on the tasks drew on task-relevant evidence and engaged in cognitive weighing of alternatives. Though the argumentation tasks are far from perfect, they have several notable strengths: they are based on an explicit map of cognition via the construct map for scientific argumentation and they demonstrated internal and inter-rater reliability. Future work should focus on the effects of surface features (e.g. type and format of scientific evidence) on difficulty estimates of items.

Article III investigated the relationship between high school biology students' knowledge of ecology and the quality of their arguments. Researchers in science education have examined the role that specific content knowledge plays in students' ability to construct and critique arguments (e.g. Koslowski, 1996; Kuhn, 1991; Lawson, 2003; Sadler, 2004), and have concluded that students need content knowledge to be able to engage in argumentation. This paper takes up the question of whether knowledge of ecological concepts is necessary, *and sufficient*, to engage in argumentation at a high level within this domain? Findings are presented from think-aloud interviews conducted as part of the process of assessment development described in Article II. Participants, all high school students, differed in their level of biology preparation. Transcripts were analyzed to determine the types of content knowledge (conceptual, procedural, and/or epistemic) students explicitly referenced while completing a set of tasks designed to measure argumentation in ecology. Transcripts were also coded to identify characteristic cognitive processes in which students engaged while solving the argumentation tasks. Findings indicate that neither students' prior biology preparation nor number of declarative knowledge statements uttered during interviews consistently predicted performance on the argumentation tasks. Rather, students needed to have epistemological understanding that went beyond declarative knowledge to be able to successfully argue in this domain. Implications for the teaching and learning of science and argumentation are discussed.

References

Achieve (2013). Next Generation Science Standards. http://www.nextgenerationscience.org

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258–267. doi:10.3102/0013189X07306523

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 National Survey of Science and Mathematics Education*. Horizon Research, Incorporated.

Crombie, A. C. (1996). *Science, Art and Nature in Medieval and Modern Thought*. London ; Rio Grande, Ohio: Bloomsbury Academic.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Erduran, S., & Jiménez-Aleixandre, M. P. (2008). *Argumentation in Science Education: Perspectives from Classroom-Based Research*. Florida State University-USA: Springer.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, *88*(6), 915–933. http://doi.org/10.1002/sce.20012

Gorin, J., & Mislevy, R. (2013). Inherent measurement challenges in the Next Generation Science Standards for both formative and summative and summative assessment. K-12 Center at ETS.

Koslowski, B. (1996). *Theory and Evidence: The Development of Scientific Reasoning*. MIT Press.

Khine, M. S. (2012). *Perspectives on Scientific Argumentation*. Springer.

Kuhn, D. (1991). *The Skills of Argument*. Cambridge University Press.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319–337. http://doi.org/10.1002/sce.3730770306

Latour, B., & Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts (2nd ed.)*. Princeton, NJ: Princeton University Press.

Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education, 25*(11), 1387–1408. doi:10.1080/0950069032000052117

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. http://doi.org/10.1007/BF02296272

National Research Council. (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC.: Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education.

Osborne, J., Henderson, J. B., MacPherson, A., Wild, A., Szu, E., & Yao, S-Y (in review). The development and validation of a learning progression for argumentation in science.

Pellegrino, J. W. (2012). Assessment of science learning: Living in interesting times. *Journal of Research in Science Teaching, 49*(6), 831–841. doi:10.1002/tea.21032

Sadler, T. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *41*(5), 513–536.

Shavelson, R.J. & Baxter, G.P. (1991). Linking assessment with instruction. In Oser, F.K, Dick, A. & Patry, J. (Eds.), *Effective and Responsible Teaching* (80-90), San Francisco: Jossey Base Publishers.

Songer, N. B., & Ruiz-Primo, M. A. (2012). Assessment and science education: Our essential new priority? *Journal of Research in Science Teaching, 49*(6), 683–690. doi:10.1002/tea.21033

Strauss, A., & Corbin, J. M. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications.

Sunal, D. W., & Wright, E. (2006). *The Impact of State and National Standards on K-12 Science Teaching*. IAP.

Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Wang, W.-C., & Wilson, M. (2005). The Rasch Testlet Model. *Applied Psychological Measurement*, *29*(2), 126–149. http://doi.org/10.1177/0146621604271053

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States*. Horizon Research, Incorporated.

Wilson, M. (2013). *Constructing Measures: An Item Response Modeling Approach*. Routledge.

Yeh, S.S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher*, *30*(9), 12 -17.