AC 2010-998: ASSESSING ELEMENTARY STUDENTS' UNDERSTANDING OF ENGINEERING AND TECHNOLOGY CONCEPTS

Cathy Lachapelle, Museum of Science, Boston

Christine Cunningham, Museum of Science, Boston

Assessing Elementary Students' Understanding of Engineering and Technology Concepts

Abstract

Engineering is Elementary's newest large-scale assessments are much improved over early attempts, thanks to innovation and improvement in the development process. Because engineering is so sensitive to specifics of a situation, and because multiple solutions are nearly always possible, targeting engineering "knowledge" and "know-how" is often best done with rubrics in the classroom. For large-scale assessment, the development of assessment questions must mirror the development of learning objectives for the curriculum. It is vital to develop questions assessing each learning objective from a number of points of view, in a variety of ways, especially in early stages of assessment development when it is unclear how students will interpret questions. Cognitive interviews with students allow for testing of content validity of questions. The inclusion of questions of a variety of difficulties, with some repetition of content assessed, improves statistical normality and reliability. Avoiding "teaching" questions that give away answers to other questions on the test is a particular difficulty when assessing engineering. This paper addresses content questions; the issue of assessing skills and process knowledge is left for future research.

Introduction

Engineering is Elementary (EiE) is a research-based curriculum development project focused on creating curriculum units that cover topics in engineering and technology as a supplement to core science instruction. Each EiE curriculum unit is designed to build on and reinforce one science topic through the exploration and development of a related technology. EiE has been committed from the project's inception in 2003 to assessing students' knowledge about engineering and technology, and measuring the impact of EiE on student knowledge and attitudes. EiE is also committed to measuring the impact of EiE on student skills and procedural knowledge; however, this effort is only in the most nascent stages of development because of the challenges associated with such assessment.

This paper describes EiE's assessment development process, as well as lessons learned by EiE over six years of design and improvement of engineering curriculum assessments. EiE has been a pioneer in this area: few instruments designed for elementary school students exist which measure student understanding of engineering and technology concepts. Over six years, EiE has developed and refined a number of instruments for measuring statistically significant change in the engineering understanding of large samples of students. These instruments have been multiple measure assessments (pre-post). Some were designed to assess understanding of basic engineering and technology content (such as "What is Engineering?"), while others have focused on measuring STEM learning resulting from student interaction with specific EiE units.

Theoretical Framework

The main goal of the EiE curriculum is to help children to learn engineering and technology—not merely facts from and about these fields of study, but practices, habits of mind, and ways of looking at the world through disciplinary lenses. At the dawn of the twentieth century, John Dewey[1] advocated that progressive education should provide scaffolding for children to enter into the adult world of work and thought: curriculum should be designed to engage children with questions and activities that take into account their lived experience while making connections to the wider world. During the same time period, Piaget[2] laid the foundations for our understanding of how children construct their own understanding of the world, working from concrete to abstract, while Vygotsky[3] showed how children's skills and understanding grow through interaction with more competent others within the zone of proximal development. Today we know that, through their experience, students learn skills and apply concepts in context[4,5,6]. Children's learning is more profound when they engage in realistic disciplinary practices[7]. In particular, learning environments must be designed to engage children productively in the social and epistemic practices of a discipline, in addition to having them work with important concepts[8,9,10,11,12].

EiE's assessment development process: An overview

The assessment development process for an EiE unit begins with the first brainstorming about that unit. The curriculum development team and evaluation team work together to identify candidate learning objectives in the science and related engineering field for that unit. Based on these candidate learning objectives, we design a focus group interview protocol to probe students' understanding of target learning objectives. Students in grades 1-5 are interviewed in focus groups of 2-5 students using this protocol. Results of the focus groups are used to refine learning objectives, design lessons, and design rubrics and assessments.

We then draft forty or fifty questions that address four basic content learning objectives in the science and engineering domains identified. We conduct one-on-one cognitive interviews with elementary school students in order to validate the candidate questions, or identify problems with them. The best questions are used to create a pilot instrument, which is then given as a paper-and-pencil test to students in a variety of classrooms (some using the EiE unit, some not). The student answers are entered into a database, and the instrument is statistically tested for reliability. Results inform the design of a field test instrument that is used nationally. The development of field test instruments, designed to be used with children in grades 3 through 5 (8 to 11 years of age) will be the focus of this paper.

Mapping the construct space

Before beginning the construction of an assessment instrument for a new EiE unit, and before beginning the design of that unit, we start with a general idea of the topics we want to cover: the engineering design field, the (probable) technology, and the related science content. Our first task is to establish what children know about these topics. Much of this we do through a search of the literature; however, while this is often quite informative about the science topics related to the new unit, it rarely leads to much insight about the engineering we wish to address.

In order to learn more about what elementary school students know and don't know about our target topics, we design focus group and/or interview protocols. These protocols are designed to take about 15-20 minutes to administer. We have found that younger students, especially, have difficulty staying engaged for longer than this amount of time; also, since we conduct these interviews in schools during the school day, we try to avoid taking the students away from the classroom for too long.

The protocols generally involve a hands-on element (which helps to engage children and keep them focused), as well as open-ended questions. Most students find them enjoyable. An example of a focus group protocol is shown in Figure 1. The students are told before beginning that our purpose is to "learn what kids your age think about ___", and "there are no right or wrong answers here, we just want to know what you think". To help the interviewers complete the full protocol in the time allotted, times are provided to the right of each protocol heading that indicate how much time we expect will have elapsed before reaching that section of the protocol. (See Figure 1 for a sample interview protocol.) Interviewers are asked to stick to the initial question and suggested prompts at first, but they are free to explore further in promising directions, as long as there is time.

Figure 1. Excerpts from focus group protocol for Sinking & Floating / Ocean Engineering

---

Satellite images   (1:00)
Place one satellite image in the middle of the group (the one showing the Ocean red) and show students where the land is. Point out Massachusetts and Cape Cod and ask:
- o   What do you think this area is (point to the Ocean)?

….

Floating/Sinking/Density questions
Have you ever heard of the word "density"? What does it mean?   (10:00)
> *If students ask you what it means, after they have given their responses, you can say that "density" describes how much stuff is packed into a space.*

What kinds of things float in water? Why?   (13:00)
- What about metal?
- Why do you think big ships can float in the water?
    - o   If students have no idea, prompt: "you said metal can't float, but big ships are made of metal…"
- What if we tried to float something in salt water? Do you think it would make a difference? Do you think that something that sank in fresh water could float in salt water?

Show students objects (paperclip, soda straw, wooden pencil, plastic bead)   (16:00)
- (Allow students to predict and sort the objects into floaters and sinkers.) Which do you think will float in water? Which do you think will sink in water? Why?

We have found that most students are quite willing to contribute their ideas and speculate about phenomena. As long as the groups are not too large, students interact productively, prompting each other with new ideas. For example, when asked "does light move?" these two students each contributed very different ideas, though they interacted almost synergistically:

> Interviewer: Does light move?
> Girl: It can, it goes through the wires and comes out of the bulb as light.
> Boy: Yeah, light can move because the Sun moves.
> Girl: Everything has to move.
> Interviewer: If the Sun is far away, how do we have light here?
> Girl: From outer space and from electricity and light bulbs.
> Boy: From stars, the Sun is the closest star.
> Interviewer: Does the light move to get here?
> Boy: Its rays do
> Interviewer: What are rays?
> Girl: Like strips of fire that come out of the Sun
> Boy: But they're not really fire, they come out of fire.

The idea of the focus groups is to get a sense of what kids understand and how they think and reason about the target topics, as well as their naïve conceptions. The goal is to collect enough of a picture of the range and depth of student thinking in order to inform the design of curriculum and assessments that address students' naïve conceptions and build on their basic understanding. Therefore we are not concerned with getting a perfect "snapshot" of any one child's understanding at one point in time.

Development of key questions and learning objectives

Development of key questions and learning objectives is an important step in the development of both the unit and assessments. Once we have a clearer picture of what children in grades 1 through 5 understand about the topics of interest for a new unit, we decide as a team on a range of content the unit and assessments will address. This content includes core science and engineering ideas drawn from state and national curriculum standards. For example, in the EiE unit *Evaluating a Landscape*, two core content areas that are addressed are (1) the geological structure of the Earth and (2) the importance of models and modeling in science and engineering; both are widely found in state and national curriculum standards.

In addition to core content, the team also decides on more specific engineering content (and non-engineering content) that the unit and assessments will address. The core content provides a foundation for this more specific engineering content. The more specific content has a lower "essential" value—it is probably not found in most curriculum standards. Its value lies in the opportunities it provides for students to apply core content in a context that will capture their imagination and provide for a unique view on the field of engineering. In the *Evaluating a Landscape* unit, for example, the specific engineering content includes: (1) the broad outline of the work of geotechnical engineers; (2) how the structure of the Earth—as well as events like earthquakes and phenomena like erosion—can affect structures, (3) different foundation types, (4) criteria for choosing sites for structures, and (5) techniques for adapting sites to be more

suitable for structures. This unit is set in rural Nepal, in a realistic (though fictional) setting, where a village must decide where and how to site a gondola-type "bridge" called a TarPul.

The context and specific content, together with core content, provide the basis for the unit to address the "Six Facets of Understanding" described by Wiggins & McTighe[13]: students should have the opportunity to explain, interpret, apply, have perspective, empathize, and have self-knowledge. The Guiding Questions that students must be able to answer to demonstrate understanding of the chosen concepts (Figure 2) are developed along with the outline of the unit and the learning objectives to be addressed (Figure 3Figure 1). This is a key step in the design process: as with any engineering design endeavor, the success of the curriculum and assessments hinges upon the quality of the learning objectives, in addition to how well the learning objectives are addressed by the design of curriculum and assessments. It corresponds to Wiggin & McTighe's Step One of "Backwards Design": Identify desired outcomes and results.

Figure 2. Guiding questions from the *Evaluating a Landscape* EiE Unit

Guiding Questions:
- Lesson 2: *How do geotechnical engineers use their knowledge of the layers of the earth to prevent damage to structures during earthquakes?*
- Lesson 3: *What are important factors to consider when choosing a site to build a TarPul along a riverbank? How can we improve a site for building a TarPul?*
- Lesson 4: *How can we use geotechnical engineering knowledge and the Engineering Design Process to select and improve a site for building a TarPul across a river in Nepal?*

In the curriculum unit, engineering learning objectives are limited to four per lesson, preferably with considerable overlap between lessons. For the unit as a whole, there are generally about 10-16 interrelated engineering learning objectives, with at least four of those dedicated to the engineering design process, and several more to experimental processes in engineering (for example: making predictions, extrapolating from data). In addition to the engineering learning objectives, approximately 5-8 science learning objectives are identified for the unit. For these, the goal of the curriculum is to reinforce the target science concepts, not (as a rule) to introduce them. This is generally done by creating activities, experiments, and a design challenge that require or at least afford an understanding of target science concepts for successful completion.

From the learning objectives for the unit, four key learning objectives are identified for assessment: one or two in science, and two or three in engineering. These are usually more global statements than the learning objectives stated in the unit—summarizing the learning goals of the unit. See Figure 3: the curriculum unit learning objectives and science tie-in concepts are subheadings marked by lower-case letters, while the summary learning objectives for the assessment correspond to the numbered headings—objective 3 is for students to understand "The use of models in engineering and science".

Figure 3. Learning objectives for the *Evaluating a Landscape* EiE Unit

1. Role of a geotechnical engineer / nature of geotechnical engineering
    a. (Lesson 1) Students will be able to identify and explain the role of geotechnical engineers in evaluating landscapes as sites for structures.
    b. (Lesson 2) Students will be able to identify and explain the role of geotechnical engineers in evaluating landscapes to decide how and where to build foundations for structures.
2. Design challenge: choosing sites and preparing foundations
    a. (Lesson 1) Students will be able to identify factors affecting whether a site is a good choice for building a structure and explain why these factors are important.
    b. (Lesson 1) Students will be able to identify and explain ways that a structure is affected by its site.
    c. (Lesson 2) Students will be able to explain the role of pier foundations in anchoring a building, and how their effectiveness depends both upon their depth and the kind of soil in which they are anchored.
    d. (Lesson 3) Students will be able to use models and controlled experiments to test, observe, and record the performance of TarPul foundations in different soil types and at different levels of compaction.
    e. (Lesson 3) Students will be able to analyze experimental results to identify how soil type and compaction affect TarPul foundation strength.
    f. (Lesson 3) Students will be able to identify relevant criteria for choosing a TarPul site and explain the properties that make a good TarPul site.
3. The use of models in engineering and science
    a. (Lesson 2) Students will be able to identify and explain how the models at hand are similar to and different from the "real thing" that they are modeling.
    b. (science tie-in) Models can be used to study the characteristics of objects, systems, or theories.
4. Landforms
    a. (Lesson 3) Students will be able to identify the parts of a river that will erode most quickly and explain why using maps.
    b. (science tie-in) A landform is a naturally formed feature on the surface of the earth.
    c. (science tie-in) Landforms change over time.
    d. (science tie-in) The course of a river can change over time and is affected by many factors.
    e. (science tie-in) Erosion is the wearing away of earth materials by water, wind, or ice.
    f. (science tie-in) The crust of the Earth is part of the layer known as the lithosphere.
    g. (science tie-in) The layers of soil found on the surface of the earth (including soil, sand, clay, and bedrock) can vary in thickness.
    h. (science tie-in) The layers of soil found on the surface of the earth may be found at different depths, depending on the specific site.

Curriculum unit learning objectives are written in such a way that teachers will be able to use student products and discussion to evaluate student progress on the objectives against a rubric. Preferred verbs for the objectives are ones that imply observable student actions: students will be able to identify, explain, predict, use, describe, analyze, etc. Some of these objectives lend themselves more easily than others to assessment via a paper-and-pencil instrument.

Creating assessment questions

Creating assessment items begins with clarification of the content to be assessed, as described in the prior section. As Taylor and Smith advise in their paper summarizing the findings of a project studying assessment development (funded by the National Science Foundation)[14], "drawing clear boundaries around the intended content ensures that items focus on what we care about…". Learning objectives are organized into summary statements and detailed objectives (Figure 3). Each assessment item is written to address one and only one of the detailed objectives, if at all possible, so that the item can clearly assess that particular idea, again as suggested by Taylor and Smith[14]. Where, possible, we draw on publicly available sources for possible items to use in our assessments, such as the NAEP database[15] or the MCAS published assessments[16].

For each detailed objective, we try to find or write at least 3-4 assessment items, more if possible, at various levels of difficulty. There are multiple reasons for this. One has to do with the assessment development process: it is extremely likely that some (or even many) items will be rejected before the development process is complete. The rest have to do with the structure of the final assessment: in order to be reliable and to have a good balance of questions for creating normally distributed scaled scores, it's necessary to have multiple questions assessing the same idea, preferably at multiple levels of difficulty. The multiple questions "triangulate" on the idea, improving reliability. And the multiple levels of difficulty allow for better discrimination between students with differing levels of understanding.

Our goal is to create multiple-choice questions each with the same number of effective response options, for uniformity—so that students aren't confused when the number or type of responses is changed. The optimal number of response options is still a current topic of research; four response items (three distractors) is the most common format, but it has been argued that three response options is more optimal,[17] since (1) it is difficult to write three plausible distractors, and if students don't tend to choose one of the distractors (because of its poor quality), the item effectively becomes a three-response item[18]; (2) a smaller number of distractors reduces the cognitive load on the children taking the assessment[19], and (3) the use of three response items instead of two apparently does not significantly increase the quality of the assessment for the purposes of conducting statistical analysis[17]. We are currently considering whether to switch to a 3-option format from a 4-option format.

Before settling on a final format for any one item, however, we frequently make up as many distracters and possible answer choices as possible, with the intent that we will interview students about what they think about each response item during cognitive interviews (described below). In this way, we can test hypotheses about what will make a good distracter, or how the response options should be worded, as we will explain below.

Figure 4. Sample questions from the _Evaluating a Landscape_ field test assessment

Some people are designing a building in an area that has a lot of earthquakes.

9. **What is LEAST important for them to consider when designing the building?**

   Ⓐ the type of soil in the area

   Ⓑ how packed down the soil is

   Ⓒ how old the soil is in the area

   Ⓓ how thick the soil layers are in the area
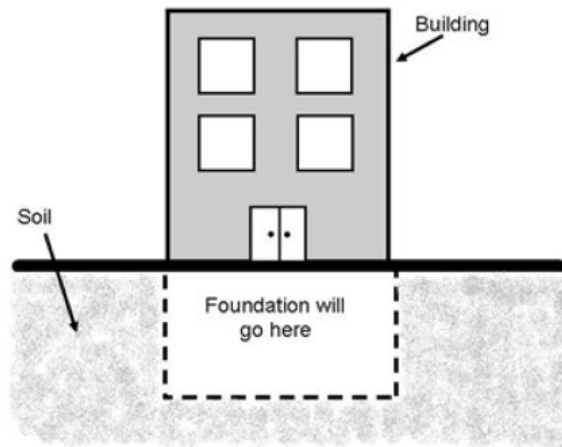
10. **How could they design the foundation to help the building stay standing in an earthquake? Choose the BEST answer.**

    Ⓐ Make the foundation as deep as possible.

    Ⓑ Make the foundation as shallow as possible.

    Ⓒ Make the foundation as thick as possible.

    Ⓓ The building should not have a foundation.

11. **A fifty-year-old bridge crosses a river in the woods. Every five years people come to check on the bridge to make sure it's in good shape. How could the river damage the bridge?**

    Ⓐ The river could flood and damage the bridge.

    Ⓑ The river could wear away the supports holding up the bridge.

    Ⓒ The river could change course so it flows around the bridge.

    Ⓓ All of the above.
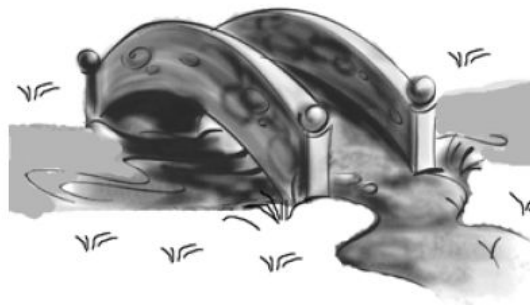
Over a series of weeks and meetings, the EiE staff brainstorms as many items as possible. These are inspected for wording—we put significant effort into rephrasing for easy vocabulary, since our goal is to test elementary school students, and we want our assessments to test science and engineering understanding, not knowledge of vocabulary and complex sentence structures. The candidate questions are also inspected to ensure they meet other criteria enumerated by Taylor and Smith[14]: (1) necessity: in order to answer the question correctly, students need to know the target content; (2) sufficiency: answering the question correctly does not require that students have other knowledge not likely given the demographics and grade range of our student

population; (3) no one item divulges or teaches the answers to other items; and (4) the structure of any one item does not give away its answer.

We have generally found that writing engineering questions is more difficult than writing science questions. Some engineering questions work best with a context: what should you do given a certain situation? Therefore we have developed a number of techniques for creating assessment items to address our more context-dependent engineering learning objectives: (1) creating a context and writing multiple questions asking about it; (2) incorporating data tables for students to look at; and (3) using diagrams and pictures to illustrate the engineering context and help students with difficult vocabulary. For an example, see Figure 4.

Conducting cognitive interviews

Validity testing of candidate assessment items is accomplished using cognitive interviews. Once a number of questions have been brainstormed for a particular assessment, the questions are printed (one to a page) and tested with students. Each interviewer takes a number of the questions (no more than can be tested in 20 minutes with a student) and sits one-on-one with a student to test the questions.

During a cognitive interview, the interviewer first reassures the student that the purpose of the interview is to see if the questions are good ones, not to test the student. Whether or not the student knows the answer to a question, we ask that s/he let us know what s/he thinks the question is asking, and whether any words or phrases are particularly confusing. The interviewer shows the student one question at a time, at first showing the stem of the question only—she hides the response options. The student is asked to give a free-response answer to the question. Then the student is shown the response options, and asked to read each one and explain what he or she thinks about it. Next the student is asked to choose an answer, and justify it. Finally, he or she is asked to comment on any diagrams or tables—whether they helped with answering the question—and also if they notice anything particularly confusing or problematic with the question.

Based on the results of cognitive interviews, some questions are revised or discarded. For example, one set of cognitive interviews revealed that when students were asked to mark the function of the plant structure "leaves", some chose the answer "making food" not because they realized that plants make their own food (as intended), but because they were thinking that plants make food (i.e. salad) for people and animals! This question was discarded because of the high risk that students would answer it correctly without having an understanding of the intended learning objective (that plants can make their own food via leaves).

Some questions in the cognitive interviews are intentionally supplied with a larger number of response options than we intend to keep. In these cases, the intention is to learn which of the distracters are plausible and appropriate for students in the population we are sampling. Other times, we find through students' free responses to the questions before they see the response options that students have unintentionally supplied us with excellent plausible distracters that we can use in the final assessment.

Once cognitive interviews are complete, the assessment items are chosen, discarded, or revised based on what we have learned as a team through group discussion. New versions of the items are ready for further validity testing through further cognitive interviews. Once the team is satisfied with the range and quality of items available, the pilot version of the assessment instrument is compiled.

Testing and revising the pilot assessment instrument

For the next step of assessment development, the questions are formatted onto a scannable assessment and distributed to approximately 12 classrooms in grades 3 to 5 (the target grades), for a target N of approximately 250. Some of these classrooms have covered the target content in their classrooms during the year (EiE pilot test classrooms), and some have not. The assessments are completed by students and returned to EiE, where they are scanned. The data are then transferred to SPSS, and subjected to statistical analysis. Internal reliability analysis is used to determine which questions are not contributing to the overall reliability of the instrument, and need to be dropped (or other questions of similar type added). Principal components analysis is used to check whether factors correspond to general learning objectives as we defined them earlier in the development process. We are currently planning to add item response analysis to our list of tools for analyzing the suitability of assessment items[20].

If all goes well, approximately half of the candidate items will be dropped, leaving four reliable scales with Cronbach's alpha >.8—one scale for each of the general learning objectives. If not, further development is required, or else a decision to reduce the number of learning objectives that will be assessed.

Further directions for exploration

Addressing students' learning of engineering, technology, and science concepts is only part of the goals of our curriculum. We also expect that students will learn how to use the engineering design process, analyze and results of prototype testing and other experiments, apply findings to their designs, and compare trade-offs. We plan to develop methods to assess students' learning of these practices in future.

Discussion

EiE is still in the process of revising our curriculum development process; over the past five years we have made many adjustments and improvements. Producing an assessment which is valid and reliable for elementary school students is a challenge itself; the fact that there are so few models for assessment of engineering content (and so few models for what is appropriate engineering content for this age group) has doubled the challenge.

We have provided an overview of assessment development methods that as yet rely upon questions that have not received much attention in the literature: What engineering content is appropriate for elementary school students? What assessment methods are appropriate? Given that our funders and participating districts demand quantitative measurements of success across

wide populations of students, we have done our best to apply the most current methodologies for unit and test development to this new field.

Our attempts to create engineering questions are an early foray into a new frontier. We have found that assessing engineering knowledge with a paper-and-pencil assessment often requires that young students be provided with a context and background information. As the vocabulary is often unknown to them, we provide it when needed with diagrams and illustrations. It is particularly difficult to avoid "teaching": spoiling the usefulness of other questions in the assessment with too much information in one item.

It is also difficult, with young children, to construct assessment items which require engineering know-how to answer, not just common sense—engineering relies so much upon knowledge and experience with the world, where do you draw the line between them? The comparison of test samples to control samples contributes to illuminating this dimension; but we would like to posit that just as some science questions can be answered by an observant and uninstructed student, so too can some engineering questions. The challenge for our time is to decide as engineering educators, and as a society, how much young children need exposure to new engineering skills, situations, materials, and experiences—and what those should be. Are the best engineers and most engineering-savvy citizens the ones with the most experience with the world both natural and human-made, objects and processes and systems? If so, how do we assess how well we've educated our youngest students?

[1] Dewey, J. (1938, 1998). *Experience and education*. Indianapolis: Kappa Delta Pi.

[2] Gruber, H. & Vonèche, J.-J. (1977). *The Essential Piaget: An Interpretative Reference and Guide*. New York: Basic Books.

[3] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychologial processes*. Cambridge, MA: Harvard University Press.

[4] Bruner, J. (2004). The psychology of learning: A short history. *Daedalus*, Winter, 13-20;

[5] Lave, J. and Wenger, E. (1991). *Situated learning : Legitimate peripheral participation*. New York: Cambridge University Press.

[6] Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in a social context*. New York: Oxford University Press.

[7] Sawyer, R. K. (Ed.). (2006). *The Cambridge handbook of the learning sciences*. New York: Cambridge University Press.

[8] Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education, 32,* 268-291.

[9] Duschl, R. A. & Grandy, R. (Eds.). (2007). *Teaching scientific inquiry: Recommendations for research and implementation.* Rotterdam, Netherlands: Sense.

[10] Engle, R. A. and Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a Community of Learners classroom. *Cognition and Instruction*, *20*(4), 399-483.

[11] Hmelo-Silver, C., Duncan, R. G., and Chinn, C. A. (2007). Scaffolding and achievement in Problem-Based and Inquiry Learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107.

[12] Rosebery, A. S., Warren, B., and Conant, F. R. (1992). Appropriating scientific discourse: Findings from language minority classrooms. *Journal of the Learning Sciences*, *2*, 61-94.

[13] Wiggins, G. and McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

[14] Taylor, M. and Smith, S. (2009). How do you know if they're getting it? Writing assessment items that reveal student understanding. *Science Scope*, *32*(5), 60-64.

[15] National Center for Education Statistics, *NAEP Questions Tool*, 8 Jan 2010, http://nces.ed.gov/nationsreportcard/itmrlsx/landing.aspx.

[16] Massachusetts Department of Elementary and Secondary Education, *Massachusetts Comprehensive Assessment System: Test Questions*, 8 Jan 2010, http://www.doe.mass.edu/mcas/testitems.html.

[17] Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24*(2), 3-13.

[18] Haladyna, T. M., Downing, S. M., and Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309-333.

[19] Kettler, R. J., Rodriguez, M. R., Bolt, D. M., Elliot, S. N., Beddow, P. A., & Kurz, A. (2008). *Modified multiple-choice items for alternate assessments: Reliability, difficulty, and the interaction paradigm*. Manuscript submitted for publication.

[20] Baker, Frank B. (2001). *The Basics of Item Response Theory.* USA: ERIC Clearinghouse on Assessment and Evaluation.