The Pedagogy of Science Teaching Test

William W. Cobern,* David Schuster, Betty Adams, Brandy Skjold
The Mallinson Institute for Science Education, Western Michigan University

Ebru Muğaloğlu
Bögaziçi University, Istanbul, Turkey

Amy Bentz, Kelly Sparks
The Mallinson Institute for Science Education, Western Michigan University

Abstract

Planning and implementing successful classroom science instruction for concept learning is a demanding task for teachers, requiring several kinds of knowledge: content knowledge, pedagogy knowledge, and knowledge of inquiry. Together this knowledge is the Pedagogical Content Knowledge of Science Instruction. Science teacher education programs routinely include science content courses where knowledge of science is assessed. Much less attention has been given to the assessment of knowledge pertaining to the pedagogies of science content instruction as typically taught in science teaching methods courses. Our assessment items are for this purpose. Each item begins with a classroom teaching vignette followed by a question asking either for an evaluation of what was done in the vignette or preference for what should be done. Although the items were designed with formative assessment in mind, sets of items can also be compiled for summative or research purposes, with versions for different science subjects and grade levels. The instruments can be used to identify science teaching orientations and pedagogical content knowledge of science instruction. This paper describes the development and testing of the items, concluding with comments on applications for instruction and future research. Example items of different types are provided and illustrative results discussed.

Keywords: STEM, Conference Proceedings, Problem Solving, 21st Century Skills, Educational Quality, Primary and Secondary education, Science, Teaching Quality, Critical Thinking

*Corresponding author: bill.cobern@wmich.edu

Introduction

An important goal if not the most important goal, for science teacher education, if not *the* most important goal, is to acquire knowledge of science teaching pedagogy. Toward this end most any K-8 teacher education program will include a science methods course. With regard to learning the pedagogy of science teaching, such courses commonly feature readings, observations of science teachers (live or by film), micro-teaching, and the practice writing of science lesson plans. It follows that the assessment of pre-service teachers' acquisition of pedagogical knowledge of science teaching is commonly done by evaluating the science lessons they have constructed, where they have the opportunity to demonstrate their knowledge of science teaching pedagogies. Given the broad range of topics typically found in K-8 curricula (e.g., the *National Science Education Standards* and *A Framework for K-12 Science Education*) and the time limitations of a college course on science methods, observing films of practicing teachers or constructing science lesson plans can touch on only a fraction of the science topics potentially taught across the K-8 grades. Take for example the idea of teaching science by inquiry. In any science methods course, at best students will see inquiry instruction applied to only several science topics. One has to wonder how far these teachers will be able to transfer their knowledge of inquiry instruction based on such limited exposure. We know from cognitive studies the transfer of knowledge best occurs when the learner sees knowledge applied in various situations (Donovan & Bransford, 2005). Might not that apply to the learning of science teaching pedagogy? Is it possible that transfer of knowledge about inquiry instruction is related to the number of different teaching situations in which inquiry is observed or considered? We think so. Our research and development work began with the recognition that the time limitations on typical science methods courses and professional development programs limit the range of exposure to science teaching pedagogies. Only a few films of teachers at practice can be viewed. Only a few lessons can be constructed by the student. We therefore sought an alternative teaching strategy that would present science teaching pedagogy across a much broader range of implementation. An efficient method for doing this could have significant instructional impact as well as utility for research. This article reports on our preliminary efforts to develop and test a collection of multiple-choice items that provide a practical means for the presentation of multiple science teaching events, where the items can be used for instructional purposes (as in Assessment for Learning) or possibly collected into tests to be used for summative assessment in science teacher education. Collectively the items are dubbed the POSTT (Pedagogy of Science Teaching Test) and are freely available at: http://www.wmich.edu/science/inquiry-items/.

Theoretical Background

The POSTT draws on ideas from "worked problems" in the physical sciences and mathematics, assessment for learning, and problem-based learning. In the physical sciences and mathematics there is a long history of assigning problem sets to students. The idea is simply that to gain skill at solving problems students must practice problem solving. In order to transfer knowledge of problem solving, students must practice solving a diverse range of problems. Recent work on the efficacy of retrieval assignments suggests why practice problems are effective (Karpicka & Blunt, 2011). Worked problems or practice problems can also be used for formative assessment, or what is called assessment for learning (Black et. al., 2003), which has been shown to improve student learning.

We suggest that the preparation of science teachers can be improved by giving students experience with worked problems, that is, to present students with classroom situations (the problem) where they have to either decide what to do next or to evaluate what has already been done. Our items are essentially 'problems' involving alternative pedagogical approaches to a given teaching situation. 'Working through' such problems with students operates as a scaffold for novices' current lack of schemas, and promotes effective instruction based on active engagement with example cases. Certain studies also suggest that students can learn effectively from suitable worked 'teaching' examples, rather than simply attempting lots of problems on their own. We proposed to create "problem sets" by drawing upon problem-based learning (PBL). PBL is an approach widely used in medical education (Albanese & Mitchell 1993) and more recently adopted in science teacher education (Dean 1999). PBL presents pre-service and in-service teachers with a practical teaching problem, often in the form of a realistic scenario or vignette. Our model, borrowing from PBL, uses realistic K-8 science teaching situations. Each item begins with a brief classroom vignette followed by a question and set of response choices.

The responses might be possible evaluations of the teacher's actions so far, or alternative suggestions for what the teacher should do next in the science lesson. Basing a pedagogy-of-science assessment on PBL has several advantages. First, the assessment is realistic and hence more authentic, given that items are based on actual classroom occurrences. The scenario-based approach also complements texts that build on case studies for teaching methods. Secondly, it is an assessment that does not lapse into measurement of rote memory, nor of generalities about inquiry. Each item specifically requires either application or evaluation, in terms of Bloom's Taxonomy (Anderson & Krathwohl, 2001) involving specific cases. Successful application and evaluation require that one understand science pedagogy and its use in science content areas at the appropriate grades. Thirdly, because the assessment involves pedagogical approaches, items are easily adapted for instructional use. A reciprocal relation can be developed between PBL as instruction and PBL as assessment. Problem-solving application of knowledge is needed before students fully grasp an area. Once a set of problem-based items on science pedagogy is available, it can be used to help students develop a usable understanding of the general principles they are learning.

Science Teaching Orientation Spectrum

As clearly documented by the National Research Council's *National Science Education Standards* (1996), through to the recently released *Next Generation of Science Education* standards (2013), the inquiry teaching of science across the K-12 grade levels is a goal to which American education aspires. However, as especially noted by the *Framework* and *Ready, Set, Science!* (Michaels et al., 2007), there is not just one type of learning objective in science education for science pedagogy. *Ready, Set, Science!* suggests four interconnected objectives: Understanding Scientific Explanations, Generating Scientific Evidence, Reflecting on Scientific Knowledge, and Participating Productively in Science. Conceivably an assessment system for teachers could address each of these objectives and could additionally address them in an integrated fashion; but the task of constructing such a system would be substantial. The POSTT primarily addresses "Understanding Scientific Explanations" for three reasons. First, feasibility of research and development requires limitations on scope. Second, our chosen limitation is based on the perspective that understanding scientific explanations is fundamental to the success of learning the other objectives. Third, teachers are often inclined toward non-inquiry instruction for the teaching of scientific explanations, more so, for example, than for teaching to the objective of "Generating Scientific Evidence." Hence, there is more need for an assessment program with respect to teaching the "Understanding Scientific Explanations."

Of further importance is that inquiry instruction for the understanding of scientific explanations is not of a single format. As noted by the National Research Council (2000), inquiry instruction can vary from instruction that is guided by the teacher to very open student-centered discovery instruction. On the other hand, direct instruction is typically interpreted as passive, teaching-by-telling. We have eschewed any simplistic contrasts between direct and inquiry instruction. Rather, our approach to the instructional strategies used in our items is framed by David Ausubel's theory of learning (Ausubel et al., 1986). Ausubel developed a theory of learning relating to type of instruction that remains cogent and relevant today for science education and for our POSTT items. Ausubel's ideas form a theoretical/conceptual framework for our development work because they elegantly integrate the foundational issues of learning and instruction critical to science education, with a focus on what is arguably most important, meaningful learning**.** Meaningful learning is defined as the result of a learner's cognitive engagement such that new knowledge becomes integrated within the learner's conceptual schemata. Ausubel provides a two-dimensional diagram representing nature of learning along a horizontal axis and type of instruction along a vertical axis (Figure 1). Learning can range from Rote to Meaningful and instructional type from Reception to Discovery. Note that Ausubel's terms Reception and Discovery are best reflected today by 'direct' and 'inquiry.'

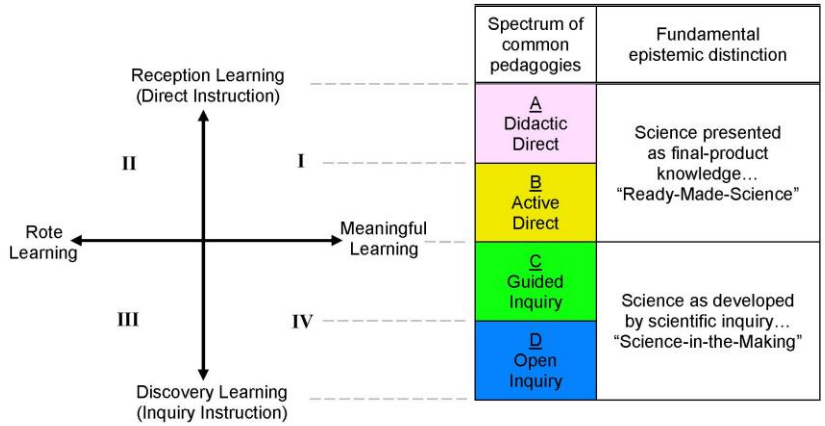


Figure 1: Ausubel's axes and the pedagogical foci for the POSTT

The four quadrants, I, II, III and IV, represent various possible combinations of learning outcomes and instructional type. Ausubel's framework indicates that *both* reception and discovery learning can be either meaningful or rote. Clearly one aims at meaningful learning via either direct or inquiry routes, implying a focus on quadrants I and IV in both instruction and research. The table in Figure 1 further explicates the vertical instructional axis by overlaying four common categories in the spectrum of science instructional types: didactic direct, active direct, guided inquiry, and open inquiry. These are not to be seen as rigid compartments, but as a useful way of characterizing instructional approaches found in practice. In the 1960s and 1970s, Ausubel (1961, 1963) and Novak (1976, 1979) argued that the important learning goal was 'meaningful learning' as opposed to 'rote learning,' *whatever the type of instruction*. They believed that reception learning could be meaningful with appropriate instructional design; Novak referred to this as 'direct facilitation of concept learning,' and they developed tools such as advance organizers and concept mapping for fostering meaningful reception learning (Mayer, 1979; Novak, 1976). Strategies from this research have shown success in shifting learning with such instruction from rote (Quadrant II) to meaningful (Quadrant I).

Ausubel and Novak understood the value of hands-on activities for learning science, and viewed this partly as a method for encouraging cognitive activity (i.e., 'minds-on'). The idea of 'active learning' is quite Ausubelian. Teaching for active learning means using instructional techniques that encourage cognitive engagement with the subject matter rather than passive listening. Ausubel also would caution that hands-on activity without cognitive engagement would not lead to meaningful learning. What Ausubel called 'discovery' learning, as advocated by Bruner (1961), subsequently developed into today's commonly known inquiry instruction. This approach advocates that learners engage with the practices of science during concept learning, that is, in activities that reflect the investigative nature of science. Unfortunately, Ausubel's two-dimensional framework of orthogonal constructs tended to be collapsed to one dimension, with direct implicitly identified with rote and inquiry with meaningful. Although reception learning research in various forms continues today (e.g., Clark et al., 2011; Klahr, 2002; Sweller, 2009), by the late 1980s the rote/meaningful learning dimension tended to be forgotten as the inquiry/direct instructional dichotomy became the focus. The widely referenced book *How People Learn* (see Donovan & Bransford, 2005) specifically advocates active learning and inquiry instruction with no mention of Ausubel, Novak, or reception learning.

Assessment items that contrasted rote instruction with instruction for meaningful learning, or poor instruction with good instruction, would be of little educational value. For this reason, the construction of POSTT items was guided by Ausubel's idea that instruction for meaningful learning can range from direct to discovery modes. Our goal was to construct items using a fixed set of instructional types so that responses could be quantified and useful for assessment (see Figure 1), broadly characterized in Table 1 below in terms of instruction for science concept development. There may be some variation in how people feel each instructional type should be characterized, but the brief descriptions give the basic nature of each and make the distinctions between them clear. Note that the instructional method details will also depend on the particular aspect of instruction involved in any item. Thus, it may be best to look at a particular item and its options (in the examples given later) to see what the nature of each of the alternative methods is for that case, consistent with the four basic categories. Note that these four types are not to be seen as rigid compartments but as regions along a spectrum, and as useful characterizations of four common pedagogies found in practice.

Table 1: Instructional Types

Didactic Direct	Teacher presents the science concept or principle directly and explains it. Illustrates with an example or demonstration. No student activities, but teacher takes student questions and answers them or clarifies.
Active Direct	Same as the direct exposition above initially, but this is followed by a student activity based on the presented science, e.g. hands-on practical verification of a law.
Guided Inquiry	Topics are approached by student exploration of a phenomenon or idea, with the teacher guiding them toward the desired science concept or principle arising from the activity. Teacher may explain further and gives examples to consolidate. Questions are dealt with by discussion.
Open Inquiry	Minimally guided by the teacher, students are free to explore a phenomenon or idea in any way they wish, and devise ways of doing so. Teacher facilitates but does not prescribe. The process is generally considered the most important thing and students present what they found.

Item Creation

The item writing team was comprised of four science education faculty, five experienced k-8 teachers, and three doctoral research associates with school teaching experience. It was important to have teachers and teacher educators involved in item creation because they have direct knowledge of situations that teachers encounter and of their own and colleagues' ways of dealing with them. The team created a wide range of draft items, working both individually and together, and then refined them after discussion. Ideas for the vignettes came from the teaching experiences of those on the item writing team, or were created with the instrument in mind for topics they were familiar with teaching, or arose from thoughts and insights about various pedagogies in context, as well as from the literature, the *National Standards on Science Education* and the Harvard-Smithsonian *Case Studies in Science Education* (2003). All items were cast in a standard format. Each item begins with a short titled vignette representing a real instructional situation for a particular topic. The vignettes specify instructional aim and grade level. Although topics are not tightly tied to grade level, stating a level within a grade band appropriate to the topic adds a sense of realism. Giving the teacher a name serves the same purpose. After the vignette comes a lead-in sentence posed in terms of an instructional method decision; for example, “Thinking about how you would teach this lesson, of the following…” The purpose is to encourage the responder to envision their self in the particular teaching situation and respond accordingly. “Of the following” is intended to keep attention on the four responses rather than other things that might come to mind. The lead-in sentence concludes with a question such as: “which one is most similar to what you would do?” “How would you evaluate Mr Goodchild’s lesson? “How would you advise Ms. Katinka to structure her lesson?”

The question is then followed by four responses representing didactic direct, active direct, guided inquiry, and open discovery. This overall item format is consistent across items although the order of the response options varies from item to item. Promising draft items, cast in this way, were modified and improved by the researchers and a sample was dissected and discussed in detail by focus groups. Focus groups comprised from 3 to 15 pre-service students or practicing teachers meeting with a project investigator. Each of four focus groups was asked to respond to a small set of draft POSTT items, and then to discuss as a group their thoughts on the items. The response was enthusiastic and as a result there was usually only time for each focus group to discuss two or three items at a time. We were particularly interested in whether the focus groups found the items realistic, in portraying topic teaching issues and decisions that might be encountered in the real world of classrooms. Indeed, it was the clear opinion of the focus groups that the items were realistic; moreover, the amount of discussion precipitated by the items boded well for the formative instructional use of the items in both pre- and in-service settings. Finally, several science education professors at other universities were asked to comment on a sample of items. They found the items understandable and appropriate, and the vignettes and responses consistent with their experiences with respect to the breadth of instructional practice.

As a result of our item development work, we have 100 multiple-choice items distributed over the three science areas (Earth Science, Life Science, and Physical Science) and three grade groupings (K-2, 3-5, 6-8). Table 2 below shows the item distribution. The numbers within each cell are the actual item numbers. One of our goals was to have items spread across science areas and grade bands; otherwise the items would not represent the broad band of instructional situations that we wanted. However, a notable feature of the collection of 100 items is that the cells in Table 2 are not even. We found it difficult to write good items featuring direct instruction at the K-2 band; and we found it easier to write items in the physical sciences across all four teaching modes. Nevertheless, distributing 100 items adequately populates each cell.[1]

[1] In the process of writing items we ended up with a few that could not be categorized by science discipline. We chose to leave them in the bank of items.

Table 2: Distribution of POSTT Items

Grades	Sciences				
	Earth	Life	Physical	General	#
K-2	1, 8, 9, 53, 98	6, 16, 37, 83, 85, 87, 90, 92	2, 7, 17, 27, 28, 34, 35, 43, 46, 56, 59, 68, 70, 71	11, 52	29
3-5	12, 49, 54, 94, 95, 96, 97, 99	15, 19, 44, 48, 67, 75, 93	10, 13, 20, 60, 61, 62, 63, 64, 65, 66, 69, 74, 80, 88	23	30
6-8	14, 18, 24, 25, 26, 32, 39, 40, 41, 42, 47, 72, 73, 100	4, 5, 29, 31, 33, 36, 50, 51, 82, 89	3, 30, 38, 45, 55, 57, 58, 76, 77, 78, 79, 81, 86, 91	21, 22	41
#	27	25	43	5	100

Piloting of Selected Items

The items were constructed to avoid responses that would have little appeal to pre-service or in-service teachers representing poor teaching generally; we wanted responses that would be seen as reasonable or reflect common practices thus attracting some responders, and engendering reflection and debate. The focus group findings were encouraging, but until a larger number of the items were actually piloted, we could not know if most subjects would focus mainly on a single response or whether we would get ranges of responses. Ultimately, the effectiveness of the items for instructional purposes can only be assessed by classroom discussions of the responses to items. However, a quicker assessment of potential item effectiveness can be achieved through a quantitative approach. Items that produce little response variation are unlikely to be of much value for promoting classroom discussion of teaching approaches. For a quantitative evaluation of the items, two POSTT forms of 16 items each were compiled. The decision to go with 16 items was based on an estimate of 3 minutes for the minimum time it would take a person to respond thoughtfully to an item. In light of the time constraints in a typical science methods course, the POSTT should take under 60 minutes. A 16 item instrument would achieve this and still contain a range of three science subjects and three grade levels. We began with two versions of POSTT to increase the number of items piloted, but also had four common items to both versions to serve as an item reliability check.[2] Hence both tests together have 28 unique items out of the bank of 100 items. Copies of both versions were randomly mixed and then administered mid-semester in three sections of a pre-service elementary science methods course at a Midwestern university. By the time of taking the POSTT, all of the students had had an introduction to science teaching strategies, including inquiry. Twenty eight subjects took POSST-1 and thirty two took POSTT-2.

Findings

Because our primary purpose in developing the POSTT was for formative assessment during teacher preparation or workshops, we are especially interested in the response spread in the pilot data. Does an item precipitate a breadth of response choices across the spectrum of four teaching options, or is the response rather narrow? A narrow spread would indicate that an item failed to capture differences amongst the subjects and thus classroom discussion arising from the item would be limited. Table 3 shows distributions across students and item responses. No student chose the same instructional strategy across all items. Most students selected three or all four of

[2] The response patterns to the four common items did not vary between POSTT version 1 and version 2. In both contexts, the responses to the four common items were about the same.

the possible strategies at least one time. For example, 32 of 60 students used all four responses at least once. Similarly, every item precipitated a range of responses; no item precipitated only a single response. For example, subjects responded to 24 items (out of 32 across POSTT-1 and POSTT-2) using all four responses at least once.

Table 3: Student/Item Response Variation

Student Response Variation		Item Response Variation	
# of different choices (1-4)	# of students	# of different choices (1-4)	# of items
1	0	1	0
2	4/60	2	2/32
3	24/60	3	6/32
4	32/60	4	24/32

We exemplify items and responses by presenting four POSTT assessment items, giving the complete item including vignette, question and response options, and the corresponding histogram of student responses, and the histogram showing how the POSTT developers themselves responded to the item. For purposes of illustration and discussion in this paper the responses are all ordered by instructional mode and labeled DD, AD, GI, an OI (didactic direct, active direct, guided inquiry, open inquiry). However when administered to participants the item responses were in random order, and labeled A to D. The three items below are those that show the strongest inquiry orientation overall, the narrowest response distribution, and the most even response distribution, respectively.

ITEM 1: Strongest Inquiry Instruction Response Orientation Overall

Magnetic attraction

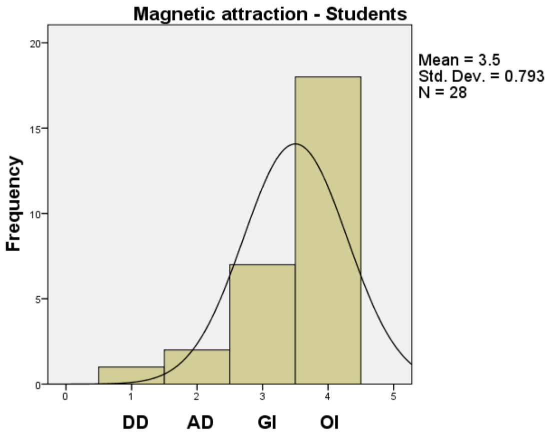
Mr. Golden is beginning a unit on Magnetism with his 1st grade students, and his objective is for them to learn about magnetic attraction. He gives each student group a bar magnet and a tray that contains a paper clip, a coin, an iron nail, school scissors, a pencil, some keys, a marble, a crayon, aluminum foil, some sand, and students can add a few objects of their own. Mr. Golden introduces the term "magnetic attraction," and demonstrates how to test a couple of objects with a magnet. Student groups are then asked to sort the objects in their trays according to whether they are attracted by the magnet or not. Thinking about how you would teach, of the following, how would you evaluate Mr. Golden's lesson?

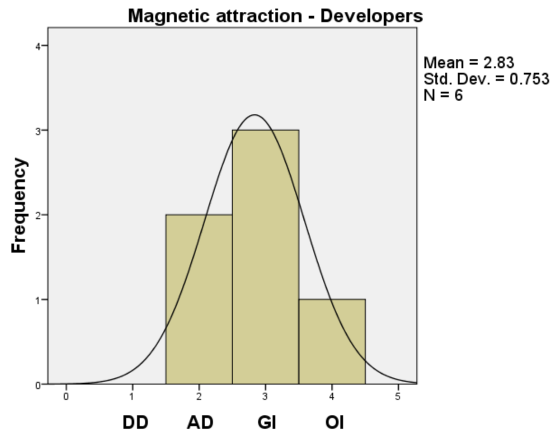
DD) This is a good lesson because Mr. Golden introduces the important terminology right at the start. However, having demonstrated how to test an object using a magnet, he might as well have demonstrated what happens with all the objects, sorting as he goes.

AD) This is a good lesson because Mr. Golden introduces the important terminology right at the start, and follows up with the students doing a hands-on activity, testing and sorting the objects themselves.

GI) Instead of beginning with terminology, Mr. Golden should have had the students first test the various objects themselves and discuss their ideas about it. In wrapping up the session, Mr. Golden could introduce the term magnetic attraction, and how it applies to what they observed.

OI) Mr. Golden should have allowed the students to explore freely with magnet and objects, without bringing up terminology. He could then let them discuss any ideas they might have about it and share these with the class. The only contribution he needs to make is to present the term magnetic attraction at the end.





The student N for this item was 28. The item clearly drew a strong inquiry response with few students opting for either form of direct instruction. What is arresting about this item is that the developers were less inquiry oriented than the students, suggesting that different methods instructors could have different views on science pedagogy, depending on the case.

ITEM 2: Narrowest Response Distribution

Earth materials

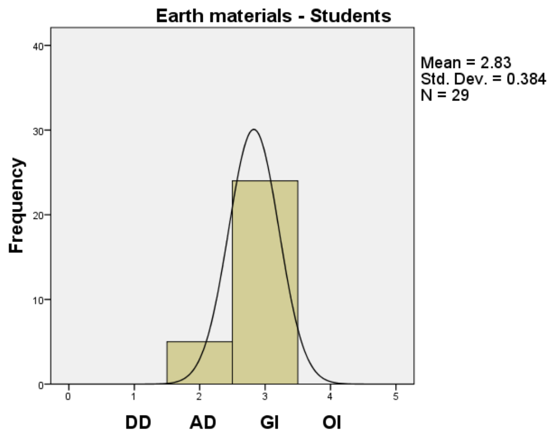
Mr. Sanchez wants his 3rd grade students to be able to recognize and describe different types of earth materials, namely rock, mineral, clay, gravel, sand, and soil samples, which he has available for use in the lesson. Mr. Sanchez is considering four different approaches to the lesson. Thinking about how you would teach this lesson, of the following, which one is most similar to what you would do?

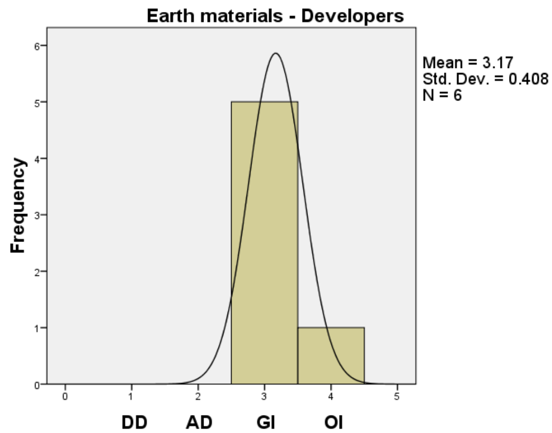
DD) I would write the different types of earth materials on the board and define them for my students. Then I would individually describe the unique characteristics of each type of material to the students, and pass the samples around.

AD) I would write the different types of earth materials on the board and define them for my students. Based on the descriptions on the board, I would then ask the students to sort the earth materials, and describe why they sorted the materials the way they did.

GI) I would have the students sort and describe the various earth materials displayed on their tables, according to their unique characteristics. I would then guide a class discussion about these different types of earth materials.

OI) I would ask the students to think about what types of materials the earth is made up of. The students would be free to explore this question with different earth materials in the classroom, and then report back on their conclusions.





The N for this item was 29. The item is one of two that drew only two responses by the students. The student mode is guided inquiry with some students opting for active direct. The developers' mode is also guided inquiry, but with one developer opting for open inquiry. Assuming that the developers are fairly typical science teacher educators, it would be interesting to compare student reasoning for their choices with the reasoning of the science teacher educators.

ITEM 3:. Most Even Response Distribution

Sink or float

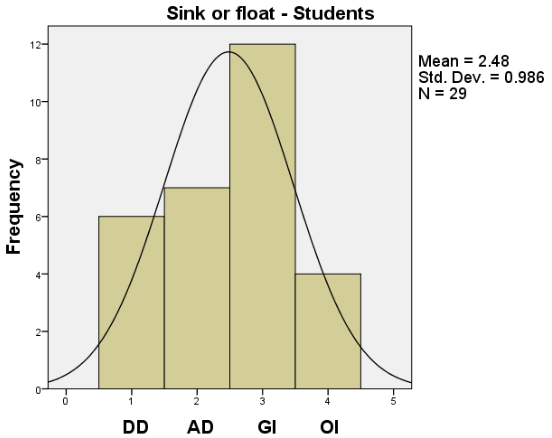
Ms. Hoo has her Kindergarten students gather around a small pool of water. She has a set of objects of different sizes and different materials; some will sink and some will float. Ms. Hoo's goal is for her students to first distinguish the objects by whether they sink or float, and then realize that this does not depend on the size of the object but on what it is made of (e.g., the stones will all sink no matter how big or small they are, and the wooden blocks will all float). Thinking of how you would teach this lesson, of the following, what would you most likely do?

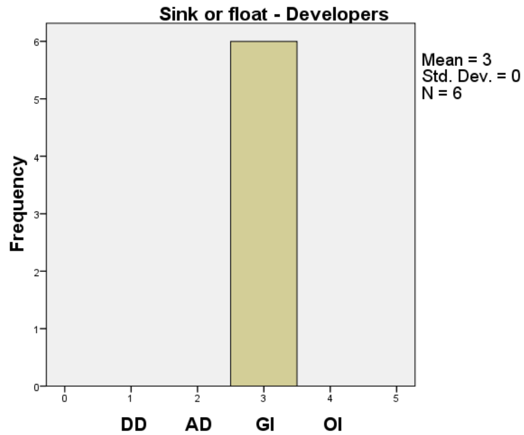
DD) Drop objects one by one into the water, and have the children notice that some sink and some float. Point out that all the stones sank, no matter how big or small, and all the wooden blocks floated, etc. Conclude by stating the lesson objective, that it is not size that matters but the material the object is made of.

AD) Have students come one by one and drop an object into the water, with everyone calling out whether it sank or floated. Point out that all the stones sank, no matter how big or small, and all the wooden blocks floated, etc. Conclude with the lesson objective, that it is not size that matters but the material the object is made of.

GI) Have students come by one by one and drop an object into the water, with everyone calling out whether it sank or floated. Ask them to suggest what this depended on; when some suggest size and others what it is made of, have them test these ideas by dropping more objects. Then have them agree on a conclusion.

OI) Have all the students dropping various objects in the water and seeing what happens. Then have them talk among themselves about this and ask volunteers to give their ideas about it, with others saying if they agreed or not.





The N for this item was 29. The item response is very interesting. All six developers took the guided inquiry position. And while the mode of the distribution for the students was also guided inquiry, the students in marked contrast to the developers were attracted by all options including didactic direct. Moreover, the students taking the POSTT were upper division teacher education students who had already taken several inquiry-oriented science content courses and were more than halfway through their science methods course, making this an unexpected response set. We suggest, therefore, that the value of the POSTT as a formative assessment tool is particularly evident in the responses to this item.

Discussion

The responses to the four items described above are typical with respect to the balance of the 32 items used for POSTT-1 and POSTT-2. All items precipitated a range of responses and most students used a range of responses, suggesting that classroom discussions based on these items could be quite telling of how the students understand and value different approaches to science pedagogies. Hearing such discussions should help the methods course science educator gauge progress toward course goals on the learning of science pedagogies. The response spread for the various items also raises interesting questions. Does grade band make a difference in how students respond to items? Does area of science or topic in an item make a difference? Does aspect or phase of lesson make a difference? These questions are relevant to instruction and hence also relevant to research. Of great interest, of course, are the reasons that students give for their choices. Student reasoning reflects instructional effectiveness and thus should be of interest to both teacher educators and researchers. The POSTT can be expected to have applications in teacher education programs in many countries. One of our members has culturally adapted POSTT-1 for use at an English-language Turkish university and has begun to collect data. Many of the response patterns are similar to ours, but there are also intriguing differences to be studied. In South Africa an earlier version of POSTT specifically for physical science was recently used to assess and compare the pedagogical orientations of in-service physical science teachers practicing in township (disadvantaged) schools and suburban (advantaged) schools, and results so far indicate remarkable differences between the preferred teaching practices of the two groups of teachers in their particular circumstances.[3] With respect to research, there is likely to be

[3] For information regarding the Turkish English language POSTT, contact Dr. Ebru Muğaloğlu <akturkeb@boun.edu.tr>. For information regarding a Turkish language POSTT, contact Dr. Hulya Guvenc

interest in using POSTT items in an instrument format for summative assessment purposes. Here we urge caution given the response spread shown in our pilot study. Nevertheless, we conducted preliminary studies of the test characteristics of the items. We deliberately placed four common items in both POSTT-1 and POSTT-2 so that we could estimate reliability. Sixty students responded to these four items across three sections of a science methods course. There were no significant differences for any of the four items between groups ($p<0.05$). The rest of the items were taken by 28 or 29 students; and, of these 24 items, 21 items showed no significant differences between sections responding to the POSTT items ($p<0.05$). The N sizes are not large, but the findings nonetheless suggest that 25 items perform rather reliably;[4] and eventually, we will have data on all 100 items.[5] At that point, the POSTT website will have histograms for every item, all inter-item correlations, and the results from factor analyses.

Potential for summative assessment use: Whether the items represent a single construct, as one would expect for summative assessment purposes, is another question. The very characteristic that makes the items useful for formative assessment (i.e., response spread), is problematic for summative assessment. The response spread shows itself in the weak inter-item correlations that we calculated. And, an exploratory factor analysis showed small clusters of items loading on separate factors. While it is possible that as data is garnered on each of the 100 items, subsets of highly correlated items may be located which then could be used for summative assessment, we suggest two alternative approaches. We are posting the developers' histogram for each item. Anyone who teaches k-8 science methods or works in k-8 science teacher development will have their own perspective on how the various item responses correspond with their instructional goals for pre-service and in-service teachers. Hence, one approach to a summative assessment is for items to be selected and scored consistent with instructional goals, making it a criterion referenced assessment. For example, if one were teaching open discovery then one could select a set of items that best fit that model. Test versions can also be compiled using items with particular science content or classroom grade level. An instrument could also be used so as to have both qualitative and quantitative aspects. In such an assessment, participants would be asked to give reasons for their instructional choices and why they did not choose the other options. Responses and reasons could be evaluated on criteria having to do with instructional decision-making. In this case, the assessment would be about both teaching orientations and the validity of rationales given for specific decisions.

Conclusion

As noted, our work is motivated by a concern that pre-service k-8 teachers typically are not able to see very many examples of science teaching pedagogies and learn from them. Moreover, with the amount of time given to science instruction at the elementary level falling (Petrinjak, 2011), the prospects are not good for seeing a wide range of science instruction as student teachers. Our project contributes a formative assessment tool, composed of case-based, problem-based objective items. Cognitive science findings and the practice of worked-problems both suggest that people need multiple exposures to instances of a practice over a wide range of situations in

<guvenchulya@hotmail.com>, for a Korean language POSTT, contact Dr. Young-Shin Park <parkys@Chosun.ac.kr>. For use in South African, contact Dr. Umesh Ramnarain < uramnarain@uj.ac.za >.

[4] The correlation matrix is available at http://www.wmich.edu/science/inquiry-items/.

[5] As noted, all 100 items are available at the POSTT website. As we gather data on each item, we are posting the student response histogram and the developers' histogram with each item.

order to develop competence and adaptable expertise. To that end, POSTT items can be used to provide novice science teachers with multiple exposures to science teaching pedagogies. The response spread in our pilot studies is a promising outcome, indicating that the items are likely to prompt lively discussion in a teacher education or professional development situation about ways to teach science. Having that discussion encourages teachers to think through instructional options and the reasons for making one choice rather than another, and the contextual factors that might come into play. It also provides teacher educators and professional development leaders with an assessment of teacher understanding of science pedagogical decision making, and thus provides formative feedback to them too about how to shape their instruction and program. POSTT items are particularly well suited for use with clicker technology, for in-class formative assessment with immediate feedback and discussion. Judicially selected individual items can be popped up for students to view and consider anytime during a methods lesson or in professional development settings. The clickers provide a way to get a quick assessment on student thinking about science pedagogy as they respond to different POSTT items. The items precipitate discussion leading to further learning as students reflect on their ideas; while the instructor is able to use formative assessment-based data to adjust a current lesson on the spot and to shape future lessons. With 100 items publically available, the POSTT constitutes a valuable resource for the improvement of science teacher education and professional development; however, we are well aware of the differences that can exist between the way pre-service teachers respond in a methods course (and specifically to the POSTT) and what they eventually do in the classroom. Hence, we are interested in the extent to which POSTT responses may be a predictor of actual teaching practices and the pedagogy choices they make in their classrooms, keeping in mind that there may be a variety of other influences constraining their preferred practices.

Acknowledgement

Funded by the National Science Foundation’s "Course, Curriculum and Laboratory Improvement" program (NSF 04-565), Award #0512596. Any opinions, findings, conclusions or recommendations in this paper are those of the authors and do not necessarily reflect the views of the NSF.

References

Albanese, M., & Mitchell, S. (1993). Problem based learning: A review of literature on its outcomes and implementation issues. *Acad Med*, *68*(1), 52-81.

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Ausubel, D. P. (1961). In defense of verbal learning. *Educational Theory, 11*(1), 15-25.

Ausubel, D. P. (1963). The *Psychology of Meaningful Verbal Learning*. NY: Grune & Stratton.

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1986). *Educational Psychology: A Cognitive View* (2d ed.). New York: Werbel & Peck.

Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning*. Berkshire, UK: Open University Press.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*(1), 21-32.

Clark, N., Eyler, G., Rivas, A., & Wagner, T. (2011) Direct Instruction vs. Scientific Inquiry: Evaluating Student Outcomes [Web Page]. URL emurillo.org/Classes/Class2/DirectvsScientific.docx [2012, June 25].

Dean, C. D. (1999, April). Problem-Based Learning in Teacher Education. Paper presented at the Annual Meeting of the American Educational Research Association. Montreal, Quebec, Canada.

Donovan, M. S., & Bransford, J. D. (2005). *How Students Learn: Science in the Classroom*. Washington, DC: Committee on How People Learn: A Targeted Report for Teachers, National Research Council, The National Academies Press.

Klahr, D. (2002) *Paths of Learning and Their Consequences: Discovery Learning Versus Direct Instruction In Elementary School Science Teaching* [Web Page]. URL http://www.lrdc.pitt.edu/supergroup/ [2009, April 10].

Harvard-Smithsonian Center for Astrophysics. (2003) *Learning Science through Inquiry* [Web Page]. URL http://www.learner.org/resources/series21.html [2011, July 20].

Mayer, R. E. (1979). Twenty years of research on advance organizers: Assimilation theory is still the best predictor of results. *Instructional Science, 8*(2), 133-167.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*. Washington, DC: National Academy Press.

National Research Council. (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.

Novak, J. D. (1976). Understanding the learning process and effectiveness of teaching methods in the classroom, laboratory, and field. *Science Education*, *60*(4), 493-512.

Petrinjak, L. (2011) *Elementary Teachers Getting Less Time for Science* [Web Page]. URL http://www.nsta.org/publications/news/story.aspx?id=58727 [2011, July 21].

Sweller, J. (2009). What human cognitive architecture tells us about constructivism. S. Tobias, & T. M. Duffy (editors), Constructivist Instruction: Success or Failure? (pp. 127-143). NY: Routledge.

Authors Information

William W. Cobern, Director of the Mallinson Institute for Science Education, holds a joint appointment in the Department of Biological Sciences and the Mallinson Institute for Science Education, Western Michigan University. He began his science education career as a high school biology and chemistry teacher. At the Institute he teaches a variety of courses for science teacher development, including an online course titled "Evolution for Teachers," and science education research. He has active research programs funded by the National Science Foundation regarding factors the influence effective science teaching and learning. Dr. Cobern is an elected 1996, elected Fellow, American Scientific Affiliation (1996), an elected Fellow, American Association for the Advancement of Science (2005), named Fellow, American Educational Research Association (2008), named Distinguished University Professor, Western Michigan University (2008), and named Fulbright Fellow to Turkey (2011).

David Schuster holds a joint appointment in the Physics Department and the Mallinson Institute for Science Education at Western Michigan University. He was previously at the University of Natal in South Africa, where he also served as national moderator for all matriculation physics examinations, and thereafter Chief Examiner for Physics for the

International Baccalaureate. His research and development interests in science education include cognition, assessment, instructional design, inquiry, conceptual understanding and problem solving.

Betty Adams, is the science lab coordinator for the Mallinson Institute for Science Education at WMU, managing technology, experimental procedures/equipment/supplies, and occasionally teaching. She has an engineering degree and experience working in the aerospace industry, but has more recently earned master's degrees in physics and science education. Betty expects to complete her science education doctorate in 2014. Her research interests include the creative use of materials and technology in teaching/learning science, the relevance of language in science instruction, scientific and mathematical modeling at all learning levels, and the role of scientific thinking in human history, philosophy, and culture.

Brandy Skjold, is a Faculty Specialist in the Mallinson Institute for Science Education at Western Michigan University. She has 15 years of college science teaching experience in various settings from undergraduate biology laboratories to graduate science education course lectures. Her research interests focus on science teachers and faculty and how their language use in the classroom effects student performance, conceptual understanding and movement through communities of practice. She is also interested in the areas of Nature of Science and Nature of Scientific Inquiry, specifically the use of models and modeling in science classrooms.

Ebru Muğaloğlu is assistant professor at the Department of Primary Education at Bögaziçi University, Istanbul, Turkey. She completed her PhD in Science Education Department at Marmara University in 2006. Before her PhD studies, she pursued MA in Philosophy program at the Institute of Social Sciences between 1997–2001 and BS in Chemistry Teaching program at the Department of Science Education at Bögaziçi University between 1991-1996. Her main research interests are nature of science, teacher training and socio-scientific issues.

Amy Bentz is a doctoral student at the Mallinson Institute for Science Education at Western Michigan University, scheduled to defend her dissertation summer 2013. Amy earned a Bachelor’s of Science in Environmental Geosciences from Michigan State University and a Masters of Arts in Teaching from Wayne State University. Ms Bentz is a former high school science teacher and an expert in Assessment-for-Learning.

Kelly Sparks is a doctoral student at the Mallinson Institute for Science Education at Western Michigan University, scheduled to defend her dissertation summer 2013. Ms Sparks holds an MA degree in physical geography (Western Michigan University, 2009) and an M.Ed. degree in Educational Technology (Grand Valley State University, 2002). Ms Sparks was for 15 years secondary science teacher. Fall 2013 she joins the faculty at Southern Indiana University.