What Do Science Teachers Consider When Selecting Formative Assessment Tasks?

Debra Tomanek, Vicente Talanquer, Ingrid Novodvorsky

College of Science Teacher Preparation Program, P.O. Box 210106, University of Arizona, Tucson, Arizona 85721-0106

Received 26 July 2006; Accepted 21 October 2007

Abstract: The purpose of this qualitative exploratory study was to identify factors that influenced prospective and experienced secondary level science teachers’ reasoning as they evaluated or selected tasks to formatively assess their students’ understanding of scientific concepts. The analysis of the coded written responses revealed two categories of factors that influenced the teachers’ reasoning: (1) characteristics of the task and (2) characteristics of students or the curriculum. Characteristics of the task related to qualities of the task regardless of the learning environment in which it would be used, such as the level of student thinking demanded by a task. Characteristics of the students and the curriculum related to the learning environment in which an assessment task would be implemented, such as students’ abilities to complete the task. Both prospective and experienced teachers’ task evaluations were influenced by the same factors related to the characteristics of the task, although their interpretations of the meaning of each factor varied. In addition, experienced teachers’ task evaluations were more likely than prospective teachers to be influenced by factors related to characteristics of students and the curriculum. The findings are discussed as a conceptual framework that presents the identified factors along three different dimensions: (1) the influence of task, student, and curriculum characteristics, (2) the influence of expectations for success, and (3) the influence of teaching experience. © 2008 Wiley Periodicals, Inc. J Res Sci Teach 45: 1113–1130, 2008

Keywords: formative assessment; teacher cognition; teacher education

A goal of science teacher education is to prepare practitioners who are knowledgeable about and competent in creating conditions that result in meaningful student understanding of science. The goal of teaching for student understanding has become a standard for evaluating science pedagogy reform. Teaching for understanding has been described as coherent instruction involving a negotiated style of interaction and analytic or diagnostic in its pedagogical approach (Prawat, 1989). Teaching for understanding has also been described as being organized around generative topics, focused on clear learning goals, built upon active learner engagement, and practiced with ongoing assessment of student performance (Wiske, 1998). Although teaching for understanding is broadly recognized as a worthy educational goal, it is far less easy to implement in real practice.

Achieving the goal of teaching for understanding will require more than simply adopting new teaching strategies (Darling-Hammond & Snyder, 2000). With regard to teacher education, achieving the goal will require a shift in the educational focus away from teaching teachers how to present the curriculum and toward teaching teachers how to think about and practice assessment of student understanding. Darling-Hammond and Snyder argue that educating teachers to successfully develop teaching for understanding practices will require experiences that connect prospective teachers’ thinking about curriculum presentation (e.g., lesson planning, selection of teaching strategies) directly to their thinking about evidence of student understanding. Since teacher education programs often provide little preparation to assist prospective teachers in developing...
their thinking about and skills in assessing student understanding (Plake, Impara, & Wise, 1997), the shift suggested by Darling-Hammond and Snyder will require a major commitment by the teacher education community.

Prospective teacher education is not the only point in the professional development continuum in need of greater attention to teachers’ assessment thinking. Assessment practices of practicing teachers are also in need of alignment with the goal of teaching for understanding (Brookhart, 2004; Crooks, 1988). Teachers’ beliefs about assessment are limited and result in too much emphasis on grading and evaluation with too little attention to assessment as a way to inform students and teachers about students’ learning (Radnor, 1994; Shepard, 2000). Shepard argued that correcting the emphasis imbalance will require changing both what is assessed by teachers and what assessment means to teachers. What is assessed must shift away from knowledge that is sequentially and hierarchically learned and toward knowledge that is integrated, principled, and useful in transfer to novel tasks. What assessment means must shift away from grading and evaluation and toward evidence useful in informing the teaching/learning process.

Empirical research findings on practicing teachers’ assessment practices support Shepard’s argument. Teachers base their instructional decisions on many factors unrelated to evidence of student understanding (Duschl & Gitomer, 1991; Morrison & Lederman, 2003). For example, teachers’ thinking about assessment is often constrained by beliefs about fairness to students (Yung, 2001), pressures of accountability (Brickhouse & Bodner, 1992; Duschl & Wright, 1989), and potential for contributing to supportive relationships with students (Fieman-Nemser & Floden, 1986). Many of the barriers to teachers’ implementation of assessments compatible with teaching for understanding appear to be psychological and social, rather than technical or psychometric (Dwyer, 1998).

Changes in teachers’ assessment practices are also driven by factors unrelated to evidence of student understanding. For example, mathematics teachers were found to implement new forms of assessment aligned with mathematics education reform efforts, but for purposes aligned with their prior beliefs about the function of assessments rather than the potential that the assessments offer for revealing students’ mathematical understanding. Similarly, mathematics teachers were found to revise and tweak tried-and-tested assessment forms, rather than adopting assessments created and aligned with reform curricula, to meet the “press” for reforming the purposes of classroom assessment (Jones, 1997; Saxe, Gearhart, Franke, Howard, & Crockett, 1999; Wood, Cobb, & Yackel, 1991). We suspect that change in science teachers’ assessment practices might also involve similar patterns of reconstruction and revision to accommodate personal values and beliefs as well as expectations by other stakeholders. Understanding how teachers’ understandings and beliefs change with training and experience is a key to improving classroom assessment (Otero, 2006).

In sum, there is little research evidence suggesting that teachers commonly build or revise instruction and assessment based on evidence of student learning, a major part of a teaching for understanding practice. Even less is known about the reasoning used by teachers who do select or create assessments that effectively target student understanding. Rather, research in this area has focused on teachers’ general knowledge and beliefs about assessment (Graham, 2005; Maclellan, 2004), the forms of assessment tools commonly used by teachers (Campbell & Evans, 2001), and the levels of understanding that various assessment tools are designed to target (Black & Wiliam, 1998; Brookhart, 2004; Crooks, 1988).

In Knowing What Students Know: The Science and Design of Educational Assessment (National Research Council, 2001), the authors describe a model for teachers’ reasoning about assessment that is referred to as an “assessment triangle.” The triangle represents essential components of the design and implementation of an assessment by a teacher. The three essential elements are: (1) Cognition—the teacher’s theories about what students know, how they learn, and what is important to understand in the subject domain, (2) Observation—the teacher’s assumptions about which tasks or situations will elicit and reveal students’ understanding, and (3) Interpretation—the teacher’s set of premises and interpretive methods for drawing inferences from collected evidence of student learning (see Figure 1).

Collectively, the essential elements of this model predict how teachers’ decisions about assessment derive from a set of complex interacting factors that shape and are shaped by the learning environment in which the reasoning and decision-making occurs. From our perspective, the model can be a useful construct for inspecting teacher thinking and decision-making with regard to assessment. In particular, we focused our
attention in this work on the Observation corner of the triangle and explored the different factors that science teachers consider when making decisions about tasks designed to formatively assess student understanding. We were particularly interested in the characterization of the features that teachers use to make assessment decisions and the analysis of how these factors differ among science teachers with different levels of preparation and experience.

Why This Study?

Our first reason for developing this study was to learn about science teachers’ reasoning associated with their selection or evaluation of “tools” that could be used as planned formative assessments. Classroom assessment refers to all activities undertaken by teachers and students “that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative when the evidence is used to adapt the teaching to meet student needs” (Black & Wiliam, 1998, p. 140). The feedback provided by formative assessment is critical in a teaching-for-understanding practice. On-going recognition of the level and quality of students’ understanding enables a teacher’s decision making about “next steps” in instruction. Formal or planned formative assessments can be defined as those that are aligned with a specific curricular framework or goal, are planned in advance by the teacher as part of a lesson or unit, and involve activities or tasks that usually result in a structured performance (Bell, 2000; Cowie & Bell, 1999; Rowntree, 1987). Planned formative assessments are different from other formative assessment situations that occur spontaneously in lessons, arise out of a learning activity, and are not specifically stipulated in the teachers’ plans (informal or interactive formative assessments). During planning, a teacher who practices teaching-for-understanding carefully considers what an assessment task has the potential to reveal about her students’ knowledge or understanding.

A second reason for the development of this study, where data were collected outside of classroom contexts, was to reveal teachers’ reasoning, to the extent possible, without influences associated with the implementation and outcomes of teachers’ decisions in real classroom practice. Of course, we are ultimately interested in understanding science teachers’ reasoning about formative assessment at all stages of the instructional process (e.g., planning, implementation, assessment of outcomes, revisions in plans). However, since so little is known about science teachers’ assessment reasoning, we felt it important to begin our exploration by focusing on the identification of teachers’ knowledge and beliefs that likely ground their assessment decisions at the early planning stages of instruction. Therefore, the following research questions were developed for this study:

- What factors influence science teachers’ reasoning when evaluating and selecting formative assessment tasks?
- Do the factors influencing science teachers’ selection/evaluation differ depending upon their level of preparation or experience? If so, how do the factors differ?

We defined the term factors used above in the research questions as knowledge, beliefs, interpretations, and assumptions that support or constrain teachers’ selection or evaluation of formative assessment tasks.
Methods

Context and Participants

The study was conducted at a public Research I University in the southwestern United States. Study participants included prospective and practicing teachers associated with the University’s College of Science based secondary science teacher preparation program (STPP). The STPP is designed for undergraduate majors in the College of Science pursuing careers in secondary level science teaching and offers all pedagogical and content courses within the College of Science. The STPP philosophy is grounded in teaching for understanding (Prawat, 1989; Wiske, 1998) and involves a major commitment to integrated field experiences throughout the four-semester program. The STPP requires a minimum of 30 semester credits of general education courses, 60 credits of science and mathematics courses, and 30 credits of science education courses. The STPP pedagogy courses are grouped into three distinct levels: (1) Level I Introduction and Exploration Courses, (2) Level II Experiential Courses, and (3) Level III Application through Internship Courses. Students enrolled in Level I courses have extended observations in secondary science classrooms. Students enrolled in Level II courses have extended experiences assisting classroom teachers with instruction and planning. However, major responsibilities for teaching are not experienced by prospective teachers until enrollment in the Level III courses.

Given our interest in exploring differences in assessment thinking for teachers with different levels of preparation and experience, participants in our study included prospective science teachers at two different stages in our teacher preparation program as well as experienced science teachers. In particular, study participants were 24 students enrolled in the first Level I course called Teaching Science, 27 students enrolled in the Level II course called Curriculum Decisions and Assessments in Science, and 41 experienced science teachers who regularly collaborate with STPP faculty members as teacher mentors for the prospective teachers and advisors to the program. The experienced teachers represented the professional demographics of the local science teacher population: they exhibited a range of 3 to 30+ years of teaching experience, a variety of science discipline teaching areas, a variety of science and non-science undergraduate majors, a mix of bachelors and masters level degrees, and a broad range of school diversity with regard to socioeconomic and cultural groups of students.

Data Collection

Two written probes were initially designed as tools to allow the exploration of what the prospective science teachers considered when selecting planned formative assessment tasks. Probe 1 (Figure 2) was on the topic of density and Probe 2 (Figure 3) was on the topic of the human body’s circulation system. These topics are representative of topics covered in physical science and life science courses at the secondary level of science instruction and likely understood by prospective teachers in the study.

We designed Probes 1 and 2 based upon our experiences with prospective teachers in the STPP and our intention to present tasks at various levels of Bloom’s Taxonomy of Cognitive Objectives (Bloom & Krathwohl, 1956). Our experiences with prospective teachers suggested that their thinking about assessment often revealed a preoccupation with hands-on activities as a highly appropriate way to assess students’ understanding. Depending on how hands-on activities are structured, they can represent either comprehension or application levels in Bloom’s Taxonomy. We had also noticed how prospective teachers often associated explaining a scientific concept with understanding the idea that it represents; this is at the comprehension level of Bloom’s Taxonomy. Finally, the prospective teachers in the STPP had also revealed through various class activities their belief in the importance of allowing for students’ creativity to be demonstrated when designing or selecting activities for their lessons. Activities that allow for the demonstration of creativity might be appropriately designated as the synthesis level of Bloom’s Taxonomy. Additionally, we included a sample task (Task 3 in Probe 1) that asked students to justify a prediction, a cognitive activity at the evaluation level. Since the study was exploratory, the new probes were not subjected to rigorous tests of instrument reliability and validity. Rather, the probes served as our first-effort tools for identifying some of the factors used by prospective science teachers in their consideration and selection of planned formative assessment tasks.
Probes 1 and 2 were administered during class to prospective teachers who had consented to be study participants in the Level I and II courses. The probes were introduced to prospective teachers as tools that would allow us to "see" their reasoning associated with selection or evaluation of tasks that they would consider to formatively assess student understanding. Level I prospective teachers had not yet been introduced to the concept of formative assessment. Therefore, formative assessment was defined for them and contrasted with summative assessment before the probes were distributed. Then, experienced science teachers collaborating with the STPP were also asked to participate in the study during one of our regular courses.

**Probe 1**

Three high school science teachers are discussing the best way to assess whether their students understand the concept of density. They propose three different approaches:

*T1:* I think the best way to assess student understanding is to ask them to design and carry out a procedure to measure the density of a solid object with an irregular shape.

*T2:* I would prefer to ask them to state in their own words what density is and to explain the difference between mass, volume, and density.

*T3:* I think we could better assess their understanding if we ask them to predict what the density, mass, and volume of a solid object would be when divided in half, and then justify their reasoning.

Which of these three tasks will best assess student understanding of the concept of density?

Select one task as your top choice, indicating its advantages over the other two possibilities.

**Probe 2**

A teacher proposes to assess her students' understanding of the circulatory system by asking them to write a creative story of a blood droplet traveling around the body. The story should include a series of keywords selected by the teacher; students are not allowed to check their notes or a textbook while writing the story.

Evaluate the potential of this task to assess student understanding.

Probes 1 and 2 were administered during class to prospective teachers who had consented to be study participants in the Level I and II courses. The probes were introduced to prospective teachers as tools that would allow us to "see" their reasoning associated with selection or evaluation of tasks that they would consider to formatively assess student understanding. Level I prospective teachers had not yet been introduced to the concept of formative assessment. Therefore, formative assessment was defined for them and contrasted with summative assessment before the probes were distributed. Then, experienced science teachers collaborating with the STPP were also asked to participate in the study during one of our regular courses.
meetings. Study participants individually read, considered, and responded in writing to one of the two probes that had been completed by the prospective teachers in the Levels I and II courses. All participants were asked to complete the probes with formative, and not summative, assessment task selection in mind.

A third probe was created later in the study because of concerns raised from our coding of the Probe 1 and 2 responses. We found that the Probe 1 and 2 responses did not result in the variety of factors that we expected teachers to apply in their assessment task selections. In retrospect, we were not surprised since we had administered the probes to about 50 prospective teachers and only 16 experienced teachers. Prospective teachers do not have the rich storehouses of experience from which a wide variety of factors involved in assessment thinking are likely to develop. We wanted another opportunity to identify factors associated with the experienced teachers’ selection of assessment tasks. Therefore, Probe 3 was developed and administered only to experienced teachers during another meeting with the STPP faculty the following semester. Probe 3 was similar to Probe 1, asking the teachers to first select, among a choice of three, the best task for formatively assessing student understanding. The task choices reflected the same basic differences in task features demonstrated in the Probe 1 task about density.

However, Probe 3 was different from Probe 1 in two ways. First, the topic of Probe 3 was changed to how cells work. The decision was made to change the topic because the topic of density in the earlier Probe 1 administration had resulted in few factors in addition to what the prospective teachers had also identified. We felt that perhaps this had something to do with the subject areas taught by the experienced teachers. We decided to use a biology topic in Probe 3 since most of the experienced teachers either taught or had taught biology. Second, a creative writing task, rather than an experimental design task, was incorporated as a choice in Probe 3. This change was made because Probe 2 (creative story about the droplet of blood) had elicited a wider variety of factors coded in the responses from the earlier administration of the probe to the experienced teachers (see Figure 4).

Data Analysis

Data were first analyzed using a method of probe response coding at the descriptive level to identify factors associated with teachers’ thinking as they selected or evaluated an assessment task (Miles & Huberman, 1994). For each of the three probe response data sets, the three investigators independently read the responses written by study participants and coded the responses as factors used by teachers in their reasoning about the selection of a preferred task (Probes 1 and 3) or the evaluation of a given task (Probe 2). Each factor was identified as a phrase. For example, one factor resulting from our coding was that the task “requires a perceived lower/higher level of thinking.” This factor was identified if the response included use of the specific words, “lower or high level thinking,” or if it indicated that the general quality of thinking required by the task was evaluated by the teacher (e.g., challenging or easy). Another factor was that the task “can provide evidence of understanding/misunderstanding.” This factor was coded from responses that included a reference to the task’s potential for revealing student knowledge, thinking, understanding, or misconceptions related to the target concept. The factor, “requires the application of student knowledge,” was coded from a response that included use of that specific phrase or if it indicated that prior knowledge was used in a new situation or in the solving of a novel problem. The factor, “focuses on assessing the target concepts without distracters,” was coded from responses that described the task as providing an opportunity for a student to focus on something other than the intended conceptual understanding. For example, several teachers’ responses indicated that a creative writing task might cause a student to exert more effort on his/her writing than on displaying understanding of the circulatory system. In this example, some teachers perceived the process of creative writing as a possible distracter.

Separate coding was conducted on probe responses from Level I prospective, Level II prospective, and experienced teachers. The numbers of participants in each of these three categories whose responses coded for the factor were tabulated. The tabulation allowed us to identify commonly applied factors in the teachers’ consideration of the tasks. Then, the investigators discussed their three individual sets of tabulated factors, negotiated consensus on the final interpretation of the factor data (i.e., descriptive coding), and discussed and came to final agreement on the labeling of the descriptive codes. The descriptive codes were then discussed by the investigators and consensus was reached about grouping the codes into two general categories of factors.

Journal of Research in Science Teaching
A second round of coding was then conducted on the factors identified in the first coding round. This coding procedure was more interpretive than descriptive. Interpretive codes result from a higher level of inference and interpretation than is the case with descriptive coding (Miles & Huberman, 1994). The assertions described in the next section, corresponding to the exploratory study questions, represent the product of the interpretive coding process.

Findings and Discussion

Teachers’ reasoning about task selection or evaluation was influenced generally by two broad categories of concerns: (1) characteristics of the task and (2) characteristics of students or the curriculum. For example, many prospective and experienced teachers’ selections were influenced by what they perceived as the level of student thinking demanded by a task. Selected tasks were often perceived as requiring a high level of student thinking. An example of a factor categorized as characteristics of students or the curriculum was the concern raised by many experienced teachers about how well prepared students were to successfully complete the task. Tables 1–3 below show the variety of factors that were associated with teachers’ reasoning about assessment task selection/evaluation. Most teachers’ responses included more than one factor influencing their task selections/evaluations. Factors used by fewer than 10% of the teachers were not included in the tables.

The remainder of this section presents a discussion of four assertions that were developed from our second level of analysis. Each discussion also includes our speculation on the basis of the assertion, and,

\begin{figure}[h]
\centering
\includegraphics{probe3.png}
\caption{Probe 3.}
\end{figure}
where appropriate, discussion of differences in reasoning demonstrated by prospective and experienced teachers.

**Factors Related to Characteristics of the Task**

**Assertion #1:** Four factors related to the characteristics of the task consistently influenced teachers’ assessment reasoning with regard to task selection or evaluation. The task: (1) requires a perceived lower/higher level of student thinking, (2) can provide evidence of understanding/misunderstanding, (3) requires the application of students’ knowledge, and (4) focuses on assessing the target concepts without distracters. However, teachers appeared to have variable interpretations of the meanings of each factor.

*Journal of Research in Science Teaching*
The most frequently identified factor related to characteristics of the task among the responses to Probes 1, 2, and 3 was that the task “requires a perceived lower/higher level of thinking.” In general, teachers with more experience or STPP training revealed the highest levels of attention to this factor in their reasoning. In Probe 1, only 33% of Level I prospective teachers compared to 63% of Level II prospective teachers and 67% of experienced teachers used the factor in their reasoning. For Probe 2 responses, 39% of Level I prospective teachers compared to 74% of Level II prospective teachers and 57% of experienced teachers used this factor. Teachers’ attention to this factor did not necessarily increase linearly with training and experience. Nonetheless, it is worth noting that beginning prospective teachers paid far less attention to the level of student thinking required by a task than was the case with Level II prospective or experienced teachers.

We feel that there are two important points to be made about this factor. First, although many of the teachers were clearly attentive to the level of student thinking demanded by a task, the meaning attributed to high level thinking was variable among the teachers. Second, there does not appear to be a relationship

<table>
<thead>
<tr>
<th>Characteristic of the Task Factors: The task . . .</th>
<th>% of Teachers Who Used the Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>requires a lower/higher level of student thinking</td>
<td>Level I (n = 24)</td>
</tr>
<tr>
<td>provides evidence of understanding /misunderstanding (makes knowledge /understanding/thinking visible)</td>
<td>39</td>
</tr>
<tr>
<td>includes structure (e.g., key words) to guide /constrain students in constructing a complete or appropriate product</td>
<td>26</td>
</tr>
<tr>
<td>focuses on assessing the target concept without distracters (e.g., creative writing)</td>
<td>57</td>
</tr>
<tr>
<td>allows students to express/demonstrate their creativity</td>
<td>9</td>
</tr>
<tr>
<td>requires students to explain their knowledge in their own words</td>
<td>26</td>
</tr>
<tr>
<td>involves application of students’ knowledge is “fun,” likely to be enjoyed by students, or likely to capture students’ attention would be easy/difficult to evaluate</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic of Students or the Curriculum Factors: The task . . .</th>
<th>% of Teachers Who Used the Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>raises questions about alignment between the task, the curriculum, teaching, objectives, and assessment (e.g., a rubric would be necessary with this task)</td>
<td>94</td>
</tr>
<tr>
<td>raises concerns about variation in students’ creative writing skills (i.e., good skills may mask low levels of understanding while poor skills may mask the demonstration of sufficient understanding)</td>
<td>0</td>
</tr>
<tr>
<td>raises concerns about variation in students’ abilities to remember terminology (e.g., poor recall is a disadvantage in completing the task even if the student understands the concept)</td>
<td>13</td>
</tr>
<tr>
<td>is relevant to the student’s life</td>
<td>0</td>
</tr>
</tbody>
</table>
between teachers’ abilities to judge the level of thinking demanded by a task and the teachers’ levels of training and experience. These points are discussed in the following paragraphs.

Although many of the study participants were clearly attentive to the level of student thinking demanded by a task, the meaning that teachers attributed to “high level thinking” was variable. This can be illustrated by reviewing the teachers’ task selections for Probe 1 in comparison to the factors that influenced their reasoning. Of the three task choices, Task 3, the task that required students to “predict what the density, mass, and volume of a solid object would be when divided in half, and then justify their reasoning” required the highest level of cognitive processing, that of evaluation (Bloom & Krathwohl, 1956). However, teachers’ evaluation of the cognitive demand of this task varied widely as revealed by the following excerpts:

- “Task 3 only shows if they can do a math problem and false logic may lead to correct answers.” (Level I prospective teacher, Probe 1)
- “Although task three is better than task two . . . it only requires a prediction from students as to the density of an object.” (Level II prospective teacher, Probe 1)
- “Having the ability to make a prediction shows that the students are thinking at a high level about the differences between density, mass, and volume.” (Level II prospective teacher, Probe 1)

A pattern of high level attention to a factor, but variable interpretations of the meaning of the factor, was also revealed by inspecting the data for the three other factors related to characteristics of the task: the task “can provide evidence of understanding/misunderstanding,” “requires the application of students’ knowledge,” and “focuses on assessing the target concepts without distractors.” For example, the excerpts below illustrate two teachers’ interpretations of the potential of the creative story writing task to assess student understanding of the circulatory system (Probe 2).

### Table 3

**Probe 3 task selection and percent teachers who used each factor for selecting the task**

<table>
<thead>
<tr>
<th>Respondents’ Professional Level</th>
<th>Probe 3 Task Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Task 1</td>
</tr>
<tr>
<td>Experienced teachers (n = 25)</td>
<td></td>
</tr>
<tr>
<td>Characteristic of the Task Factors: <strong>The task...</strong></td>
<td></td>
</tr>
<tr>
<td>requires a lower/higher level of student thinking</td>
<td>76</td>
</tr>
<tr>
<td>provides evidence of understanding/misunderstanding (e.g., makes knowledge/understanding/thinking visible)</td>
<td>48</td>
</tr>
<tr>
<td>requires an appropriate or inappropriate level of thinking (e.g., the task was not challenging enough for high school level)</td>
<td>40</td>
</tr>
<tr>
<td>focuses on assessing the target concept without distractors</td>
<td>36</td>
</tr>
<tr>
<td>allows students to demonstrate their creativity</td>
<td>28</td>
</tr>
<tr>
<td>requires transfer or application of student knowledge</td>
<td>20</td>
</tr>
<tr>
<td>Characteristic of Students or the Curriculum Factors: <strong>The task...</strong></td>
<td></td>
</tr>
<tr>
<td>raises concerns about the variation in students’ writing/reading abilities (e.g., ESL students or students with low confidence in their writing would have difficulty writing a creative story, but may know the content)</td>
<td>16</td>
</tr>
<tr>
<td>raises concerns about the prior knowledge that students can be expected to have (e.g., They may not know what happens in a bread factory)</td>
<td>12</td>
</tr>
<tr>
<td>raises possibilities for use in some other capacity, such as “building” toward another curriculum activity (e.g., Model building may be good as an on-going review of information)</td>
<td>12</td>
</tr>
</tbody>
</table>
“This task targets three levels of understanding: knowledge, comprehension, and application. The students need to recall the definitions of the terms that must be in the story. They have to be able to explain how the circulatory system works. Finally, they have to be able to apply the definitions and the explanations together into a story about a blood drop. If a student understands all of the key components, they should have no trouble applying the definitions and the explanation of the system into a story.” (Level II prospective teacher, Probe 2)

“This task has little potential to assess student understanding of the circulatory system. The students would only have to memorize the path that blood would take throughout the body. They would not have to know how all the organs in the body are connected together; they would not have to describe how oxygen is exchanged in the lungs for CO₂ and such. His assessment could be accomplished by a student who knows the major parts of the circulatory system and nothing more. Therefore, a student could write that his/her blood drop started in the heart, went to the lungs, then on to the arteries, through the capillaries, then to the veins, and finally back to the heart again. This tells nothing about how the circulatory system feeds the organs with O₂ and glucose, or how it carries away waste, or anything else.” (Level II prospective teacher, Probe 2)

We speculate that the variability of meanings that teachers assigned to the task’s potential for requiring a “high level student thinking” or providing “evidence of understanding” is a critically important part of this finding. It is the variability of meanings revealed in this study, more than the identification of reasoning factors regularly applied by teachers, that suggests to us a major challenge for efforts to improve teachers’ thinking about assessment (Darling-Hammond & Snyder, 2000; Shepard, 2000). The challenge involves teachers’ own knowledge and understanding of subject matter. But, it also involves teachers’ recognition of what it means to understand subject matter. The data in this study suggest that the prospective teachers as well as the experienced teachers would benefit from opportunities to consider and reflect upon their beliefs and knowledge about what it means to understand various science concepts.

The second point of discussion about the level of student thinking factor is that there appears to be no relationship between teachers’ abilities to judge the level of thinking demanded by a task and their levels of training and experience. At first glance, we thought such a relationship might be supported by the data. But, this was not the case. With Probe 1, Task 3 was selected by 17% of the Level I prospective teachers, 30% of the Level II prospective teachers, and 44% of the experienced teachers. However, when we searched for corroborating evidence in Probe 2 data, the only other probe completed by all three levels of teachers, the data did not support the idea that teaching experience was related to the teachers’ perceptions of the level of thinking demanded by a task. High percentages of teachers at all three levels, relative to their use of other factors, used the “requires a perceived lower/higher level of thinking” in reasoning about Probe 2. Our analysis of the raw response data indicates that roughly half of the teachers at each level thought that the Probe 2 creative writing task required high levels of student thinking and half judged it to be low level.

Assertion #2: Some of the factors identified as relating to the characteristics of the task appear to selectively influence science teachers’ reasoning and only in response to assessment tasks with particular characteristics.

Four factors were identified, relating to characteristics of the task, which consistently influenced the reasoning of all levels of teachers in all probes. However, there were also some factors that were less consistent in their appearances, but nonetheless important, as influences on teachers’ assessment reasoning. These factors appeared to be applied more selectively and only in response to tasks that exhibited specific features. Three examples of these factors were that the task: “requires performance through the creation/design/completion of an experiment” in Probe 1, “allows students to express/demonstrate their creativity” in Probes 1 and 3, and “requires students to explain their knowledge” in Probes 1 and 2.

In Probe 1, Task 1 involved students designing and completing their own experiments. Task 1 was selected by 51% of all prospective teachers and 56% of the experienced teachers as the task that would best assess students’ understanding of density. Twenty-one percent of Level I prospective teachers, 48% of Level II prospective teachers, and 33% of experienced teachers applied the factor that the task “requires performance through creation/design/completion of an experiment” in their reasoning about the Probe 1
task selection. This was a factor that influenced the assessment reasoning of many teachers in the study. In particular, it was the performance quality, rather than the hands-on quality, of the task that elicited the high level of teacher attention. However, this particular factor did not appear in the reasoning associated with Probe 2 or 3. For example, no teachers applied the “requires performance through creation/design/completion of an experiment” factor in evaluating the creative writing task in Probe 2. If this factor had been broadly applied in assessment reasoning, we would have expected some responses to indicate that a “performance” or at least a “hands-on” component was missing in the task under consideration. We interpret this to mean that once a minimum set of factors related to the characteristics of the task have been considered (identified in Assertion #1), teachers apply additional factors in their reasoning about the quality and potential of assessment tasks. The hierarchy of value assigned to the factors is not clear from these data. In other words, an influential factor in teachers’ reasoning about a task does not clearly predict the selection of a particular task. However, it is clear that some factors were applied more generally resulting in consistent influence on reasoning while other factors were applied more selectively.

The factor that the task “allows students to express/demonstrate their creativity” was also an important influence on teachers’ reasoning when a task was perceived as having the potential for students to use “creativity.” Teachers perceived Probe 2 (creative story about how the circulatory system works) as allowing for students’ demonstration of creativity. In Probe 3, experienced teachers also perceived the tasks involving story writing about the plant cell and bread factory analogy and the building of a model plant cell using various materials as opportunities for student creativity.

“The assessment of the circulatory system is original. It will give students the opportunity to be creative in their explanations of their knowledge of the circulatory system and may appeal to those students more interested in art and writing than science.” (Level II prospective teacher, Probe 2)

“The [creative story that describes how a plant cell is like and unlike a large factory that produces bread] would be my first choice. It gives a student a chance to be creative while at the same time showing his/her knowledge.” (Experienced teacher, Probe 3)

We speculate that teachers interpret creative writing and model building as demonstrations of student creativity because the activities result in new products (e.g., a story, a 3-D model). However, tasks involving activities such as predicting the outcome on density of a solid substance that is divided in half or designing an experiment are not perceived as tasks that “allow students to express/demonstrate their creativity.” Perhaps the teachers interpreted creativity, with regard to factors related to characteristics of the task, as a quality associated with tangible, rather than cognitive, products.

An additional example is useful for considering how a factor was not only selectively applied, but applied differently by prospective and experienced teachers. The factor is that the task “is relevant to real life,” found only in Probe 2. This factor was present in the reasoning of 29% of the experienced teachers, but none of the prospective teachers. Also, the factor influenced the reasoning of no teachers in Probe 1 or 3. Thus, the factor was uniquely applied only by experienced teachers and only in responses to Probe 2. In our frequent interactions with experienced science teachers, we are struck by the number of teachers who regularly comment on “relevance to real life” as an important element of their thinking about motivating their students to engage in the curriculum. But, it appears that this factor is important only when relevance to life is perceived as being exhibited in the task. In the case of Probe 2, the creative story involving the human body appeared to elicit the “relevance” factor. In Probes 1 and 3, we believe that the teachers did not find the given task choices as “relevant to real life” and therefore did not apply the factor in their reasoning about the probes. Yet, we wonder why no experienced teachers identified the lack of relevance to real life as a factor influencing their selections in Probes 1 and 3.

One final example from the data illustrates the influence of a factor used consistently, but only by prospective teachers in this study. In Probes 1 and 2, tasks were included that “required students to explain their knowledge,” a factor that influenced the assessment reasoning of Levels I and II prospective teachers. In Probe 1, Task 2 required students to use their own words to explain density and the differences between mass, volume, and density. Task 2 was selected by 37% of Level I prospective teachers, 15% of Level II prospective teachers, and none of the experienced teachers. Seventeen percent and 22% of Level I and II prospective
teachers, respectively, applied the factor that the task “requires students to explain this knowledge” in their reasoning, but none of the experienced teachers used this factor in their reasoning about Probe 1. Similarly, Probe 2 data revealed the use of this factor by prospective teachers, but not experienced teachers in the study. One way to interpret this finding is that prospective teachers in the study appear to see performances such as “student explanations” as demonstrations of student understanding and experienced teachers do not. However, another equally plausible interpretation of this finding is that experienced teachers do not interpret “student explanations” in the same way as prospective teachers in the study. For example, we wonder whether experienced teachers might have invoked the “requires students to explain their knowledge” factor into their reasoning if the choices in Probe 1 (concept of density) had been expanded to include: “I think the best way to assess student understanding is to have the students explain why an ice cube floats in a glass of water.” Much like the multiple interpretations of “high level thinking,” perhaps teachers’ interpretations of “explanations” are similarly variable.

Assertion #3: The type of factors related to characteristics of the task did not vary with respect to the teachers’ level in the STPP or teaching experience. However, the number of factors identified by individual teachers peaked with Level II prospective teachers.

When we began data collection in this investigation, we anticipated seeing differences between the reasoning applied by prospective teachers who were at different points in the STPP. We expected to see a greater number and variety of factors demonstrated by the prospective science teachers at the more advanced level in the program. We expected this because the Level I prospective teachers completed Probes 1 and 2 during the first 2 weeks of their first STPP course when little instruction and no classroom observations had yet occurred. Level II prospective teachers had received considerably more instruction on science teaching, student learning, and assessment of student understanding when they completed Probes 1 and 2. Additionally, the Level II prospective teachers had completed many hours of observation and school internship experiences where they assisted teachers in planning and teaching lessons to real students. However, our expectations were not supported. A variety of factors related to characteristics of the task were coded from the Level I prospective teachers’ responses for Probes 1 and 2. Level II prospective teachers’ responses for Probes 1 and 2 revealed essentially the same factors used by Level I prospective teachers.

Yet, a trend was noticeable among Level I, II, and experienced teachers’ responses in the number of factors identified by single individuals within each group. In general, the number of factors used by a teacher to select/evaluate a formative assessment task peaked with the Level II teachers (see Table 4).

We found that the Level I and experienced teacher groups each used, on average, fewer factors to select/evaluate an assessment task than Level II prospective teachers. The excerpts below illustrate the difference in the number of factors used in reasoning by the three groups of teachers as they described their reasons for selecting a particular task in Probe 1.

“I think Task 1 will be the most effective because in order to design and carry out the procedure, the students will first have to examine what they know already and what they really think density is. Task 2 asks students to explain too much without having an opportunity to learn it. Task 3 may get at an idea, but I think it may confuse students and draw them away from the main ideas.” (Level I prospective teacher, Probe 1)

“I believe that Task 1 is the best way of measuring understanding. There are a few reasons for this. First, Task 1 cannot be successfully performed without the students knowing the science behind the concept of density (Task 2 could be performed by rearranging a definition, and Task 3 by rearranging a mathematical law). To successfully accomplish what is asked of them in Task 1, students must not only understand the definitions and math, but be able to apply these in a hands-on situation and create something of their own, which also means that Task 1 requires higher level thinking skills than Task 2 or Task 3. Finally, Task 1 is, if not completely authentic, at least more engaging than the other two questions, and students would be more likely to remember the activity as a result, which is the goal in the end—not to teach something for a month, but for a lifetime.” (Level II prospective teacher, Probe 1)
“Task 1—designing and carrying out the procedure clearly indicates understanding (or not) of a concept. Just having to state it demonstrates memorization—not comprehension. Predicting and justifying without testing also do not necessarily demonstrate understanding.” (Experienced teacher, Probe 1)

We speculate that the increased number of factors used by Level II teachers could have been related to their greater familiarity with the reasons for and the types of assessment than is the case with Level I prospective teachers. The drop in number of factors used by experienced teachers, when compared to Level II prospective teachers, could be associated with at least two possibilities: (1) experienced teachers may be more likely than Level II prospective teachers to interconnect factors that influence their reasoning, or (2) experienced teachers’ reasoning may downplay factors deemed less important than others and therefore may not be worth mentioning when rationalizing the selection of a task.

**Factors Related to Characteristics of Students or the Curriculum**

Assertion #4: Experienced teachers are more likely than prospective teachers to use factors unrelated to the task’s characteristics, but rather related to the abilities of their students or their curriculum, when selecting/evaluating planned formative assessment tasks.

Prospective teachers’ reasoning was largely influenced by factors related to the characteristics of the task. In contrast, experienced teachers’ reasoning was influenced by both characteristics of the task and characteristics of students or the curriculum. A striking example of this was provided by one of the highly experienced teachers who, after completing his response to Probe 1, stated that he felt Task 3 was definitely the best task to assess students’ understanding of density. But, he would not select it for use with his students because the vast majority of the students would not be able to “handle it.” The task would be too difficult for the students. Additionally, 43% of the experienced teachers used “variation in students’ abilities to remember terminology” as a factor in evaluating the creative writing task in Probe 2. And 40% of the experienced teachers used “thinking abilities of high school students” in reasoning about their selection of the best of three tasks for assessing student understanding of how cells work in Probe 3. Both of these percentages are relatively high when compared to responses from prospective teachers. Level I prospective teachers demonstrated the use of factors related to the curriculum or students in their reasoning only once, in Probe 2 requiring the creative writing exercise. Level II prospective teachers used only four factors related to the curriculum or students in their responses to Probes 1 and 2. However, 30% of Level II prospective teachers used the factor relating to the task’s “alignment with the curriculum, teaching, objectives, and assessment” in their reasoning in Probe 2. This higher level of attention to curriculum issues, as compared to the Level I prospective teachers, could be related to the topics of instruction in Level II courses in the STPP where lesson and curriculum planning are emphasized. Of the three study probes, Probe 3, administered only to experienced teachers, elicited the greatest variety of factors related to the curriculum or students.

Finally, it is important to note in this section that the factors influencing reasoning that are related to students and the curriculum will likely challenge efforts to reform assessment practices. These factors included characteristics such as “alignment between the task and previous lesson activities” and the “potential it holds for ‘building’ toward some other curriculum activity.” These factors are related to an experientially derived knowledge of where the curriculum has been, where the teacher intends the curriculum to go, and where it is likely
Building a Framework for Describing Teachers’ Assessment Reasoning

The findings from this exploratory study allowed us to begin building a set of factors that account for science teachers’ selection or evaluation of planned formative assessment tasks. We have interpreted these factors as possibly existing across three different dimensions: (1) characteristics of the task vs. characteristics of students or the curriculum, (2) expectations of the learner vs. expectations of the task, and (3) the influence of teaching experience. This section describes the three dimensions included in the framework that is presented visually in Figure 5. The bold lines around boxes represent factors that were particularly salient in the teachers’ reasoning.

![Figure 5](image-url)
Dimension One—Characteristics of the Task Versus Characteristics of Students or the Curriculum

One way to think about the findings of this study is to consider the identified factors with respect to teachers’ knowledge, beliefs, and assumptions about relevant characteristics of the task or the context in which it will be used. This view is represented by the left and right sides of the diagram (separated by a vertical line through the center). On the left, factors identified as characteristics of the task are related to teachers’ interpretations of the inherent qualities of the task separate from the particular students or curriculum with which it will be used. To the right, factors identified as characteristics of the context are related to teachers’ ideas about the learning environment in which the task will be used, rather than qualities inherent to the task.

Dimension Two—Expectations of the Learner Versus Expectations of the Task

A second way to think about the findings of this study is to consider the identified factors in light of teachers’ knowledge, beliefs, and assumptions associated with the expectations of learners versus those of the task. We represent this view of the framework by comparing the factor categories on the top of the diagram with the factor categories on the bottom (separated by the horizontal line through the center). At the top, factors identified as expectations of the learner are related to teachers’ interpretations of the cognitive and performance demands that the task places on learners and their beliefs about learners’ abilities to successfully complete it. At the bottom, factors identified as expectations of the task are related to the demands that teachers require a task to fulfill: clarity, focus, assessment of understanding, and fit within the curriculum.

Dimension Three—Influence of Teaching Experience

A third way to think about the findings of this study is to consider the identified factors in relation to the influence of teaching experience. This interpretation of the framework is again represented by comparing the left side with the right (separated by the vertical line through the center). The identified factors on the left side of the figure were used equally by prospective and experienced teachers in their selection or evaluation of planned formative assessment tasks. Therefore, the teachers’ use of factors on the left side of the diagram was not highly influenced by teaching experience. However, the factors on the right side of the figure were used much more commonly by experienced, as compared to prospective, teachers. This indicates that teaching experience did influence the use of factors related to the characteristics of students or the curriculum.

Conclusions

In this study we explored teachers’ assessment reasoning by identifying factors that influenced prospective and experienced science teachers’ selection or evaluation of planned formative assessment tasks. We learned that the teachers in our study demonstrated some general tendencies with regard to the selection/evaluation of tasks to formatively assess student understanding. These tendencies included: (1) consistent use, but variable interpretation, of some factors that were related to the cognitive demands of the task and the task’s efficacy as an assessment tool, (2) selective use of some factors in response to particular qualities of the task, (3) use of similar varieties of factors related to characteristics of the task by prospective and experienced teachers, and (4) greater use of factors related to students or the curriculum by experienced teachers. Furthermore, we learned that there were two general categories of factors that characterized teachers’ reasoning: (1) factors related to characteristics of the task, and (2) factors related to characteristics of students or the curriculum. However, the identified factors could also be considered with respect to expectations of the learners versus the expectations of the task and the influence of teaching experience.

Implications

Understanding teachers’ reasoning while selecting/evaluating planned formative assessments is important because teachers who teach for understanding are competent and strategic assessors of student understanding. As such, we believe there are three implications from this exploratory study. First, teachers apply factors in their reasoning that sometimes work against the selection of tasks that have great potential to assess students’ understanding of science concepts. Factors such as perceptions of students’ abilities to
successfully complete a task weighed heavily on experienced teachers’ task selections, even when the teachers recognized a task’s potential to assess student understanding. We feel that the existence of this reasoning will present challenges for efforts targeted on improving classroom assessment practices. Second, the type of change in assessment practices called for by educational researchers noted at the beginning of this article will be challenged by the variety of teachers’ interpretations of factors related to characteristics of the task. For example, prospective and experienced teachers must be provided opportunities to consider their beliefs about and interpretations of evidence of student understanding. Third, we feel that the findings from this exploratory study suggest a productive area for future research on teachers’ reasoning about assessment. For example, is the variation in factor interpretation related to teachers’ limited beliefs about the purposes of assessment (Radnor, 1994; Shepard, 2000)? Or is the variability associated with teachers’ poor understanding of the scientific concept being assessed? In other words, if a teacher has only a formula-based understanding of the concept of density, is that teacher likely to discount a predict-and-justify task as a simple math problem? Future research on teachers’ formative assessment reasoning is needed for the improvement of assessment practices in our schools (Shepard, 2000).

This work was supported in part by grants from the National Science Foundation (DUE-0088046 and PHY-0108787) and the Howard Hughes Medical Institute (71195-521304). The authors would like to thank these sponsors for their support. We would also like to thank the prospective teachers in the College of Science Teacher Preparation Program and our experienced teacher colleagues who allowed us to examine their thinking about assessment.

References


Journal of Research in Science Teaching


