Case Studies in Teacher Content Learning in a Problem-Based Learning Professional Development Setting

Tom J. McConnell
Ball State University

Joyce M. Parker
Michigan State University

Jan Eberhardt
Michigan State University

A paper presented at the
National Association for Research in Science Teaching Annual Conference,
Indianapolis, IN, March 27, 2012.

This material is based upon work supported in part by the National Science Foundation, under special project number ESI – 0353406 as part of the Teacher Professional Continuum program. Any opinion, finding, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of any of the supporting institutions.

Assessment of teachers' content knowledge provides an important type of information needed by teacher educators to inform design of professional development (PD) curricula for teachers as well as for accountability (Howley & Howley, 2005; Mundry, 2005). However, assessment of teacher content knowledge poses several challenges. Educators and researchers need a strategy for assessing teacher content knowledge that is efficient but authentic, offers some level of standardization, can be applied to multiple content areas, and provides insight into teachers' deep understandings of science concepts, even when applied to small sample groups. This article describes a model for assessment of content knowledge used by researchers in the Problem-Based Learning (PBL) Project for Teachers of Science, an NSF-funded research project that provided professional development for 206 science teachers for grades K-12.

Having deep and coherent knowledge of science content is critical for science teachers (Jeanpierre, Oberhauser, & Freeman, 2005; Traianou, 2006; Darling-Hammond & Richardson, 2009; Hill, Rowan, and Ball, 2005; Mundry, 2005). It is the foundation for giving clear explanations and for identifying relevant and accurate examples of concepts. Without understanding of the science concepts, teachers are not able to organize and implement meaningful curriculum that includes multiple representations and models of the concepts in ways that strip away details from the essential concepts. A deep understanding of science concepts also allows teachers to develop application assessment tasks that reveal students' understanding and identify misconceptions (Ball, 1997; Ma, 1999; Thames & Ball, 2010; Traianou, 2006).

Content Knowledge Assessment Strategies

Several strategies have been used to assess changes in teachers' science content knowledge. These include concept inventories (Hauslein, Good, & Cummins, 1992; Savinainen & Scott, 2002; Tretter, Brown, Bush, Saderholm, & Moore, 2007), concept maps and drawings (Da-Silva, Mellado, Ruiz, & Porlan, 2007; Trundle, Atwood, & Christopher, 2006; Weizman *et al.*, 2008), and interviews and observations (Traianou, 2006). **Concept inventories** are the easiest to administer and analyze. They allow researchers to see changes in teachers' ideas and compare data across groups of teachers. The concept inventories assess a person's ability to recognize accurate descriptions or explanations, but do not reveal how teachers explain concepts in their own words, a critical part of their role (Nehm & Schonfeld, 2008). Also concept inventories vary in their scope and detail making comparisons across content areas difficult. An additional problem with inventories is that they may only include a few questions that are aligned with the goals of a specific professional development activity.

Concept maps and drawings are open-ended and can be used to examine the subjects' ability to connect concepts and to connect ideas to real-world contexts. However, this type of assessment usually requires additional questioning to elicit detailed explanations of the diagrams before interpretation can take place. Researchers may have to make assumptions and inferences based on incomplete descriptions of the learners' ideas. As with concept inventories, concept maps do not mirror the type of explanation a teacher would have to give or analyze in the classroom.

Other researchers have employed different approaches to gain access to teachers' understanding of science concepts. Traianou (2006) used a variety of qualitative approaches to assess the development of teachers' content knowledge, including **interviews, correspondence, teachers' writing, and classroom observation**. Her approach is effective in revealing the depth of teachers' science understanding, but this strategy is time consuming and difficult to use when trying to assess science knowledge of more than a few teachers or to compare knowledge of teachers in different content areas.

PBL Project for Teachers of Science. In this paper, we discuss an approach that combines different types of extended-response questions to gauge the science content knowledge of teachers participating in the Problem-Based Learning (PBL) Project for Teachers of Science (Authors, 2008). This was an NSF-funded project (#ESI-0353406) that offered professional development to over 200 K-12 science teachers over a span of four years. Planners and facilitators for the PBL Project developed and tested an assessment strategy that included two types of open-response questions. The first type of question asked participants a general knowledge question in which they wrote what they knew about a topic. One or more application questions then followed each presenting a specific scenario that related to the topic in the general knowledge question. Participants were asked to explain why or how this phenomenon works. These two types of questions assessed both the breadth of teachers' understanding of a topic and the depth of teachers' ability to use the concepts in specific, authentic contexts. In this paper, the authors describe the types of information that can be revealed by this assessment strategy used in the 4th cohort of teachers with the PBL Project for Teachers of Science.

Research Questions

The goals of the PBL Project included the evaluation of the assessment instrument described in this paper to determine how well it met the needs of teacher educators as a tool for measuring content knowledge and teacher learning. The research questions that guided this research included:

1. What information about teachers' content knowledge can this method yield?
2. What are the strengths and limitations of this assessment approach?

Methodology

Participants

The participants in the 4th cohort of the PBL Project for Teachers of Science included 78 individuals who taught science in schools across central Michigan. The cohort included 32 teachers from elementary grades, 28 from middle schools, and 18 from high school classrooms. The ages of the participants ranged from 26 to 69, and averaged 40 years. The teachers came to the project with years of experience teaching ranging from 1 to 43, with an average of 11 years. Only 9 of the 78 teachers were male. All teachers who applied to the PD program were accepted. They participated in a content strand of their choosing.

Professional Development Design

The PBL Project had three main components: the Professional Working Conference, the summer Focus on Practice, and the academic year Focus on Practice.

The Professional Working Conference (PWC). The PWC spanned seven days during the summer. The first three days included problem-based learning activities designed to help teachers learn science content. Project planners developed content "strands" that addressed the teachers' selected topics. In the summer of 2008, the strands included Geosphere for elementary teachers, Geosphere for secondary teachers, Genetics, Cell Biology, Ecology, Astronomy, and Weather. Both PBL activities and assessments were organized around "Big Ideas" (Brooks & Brooks, 1993) or key concepts that are central to an accurate and coherent understanding of the science subject.

The major learning activities were "dilemmas" or problems used in a problem-based learning format. A PBL dilemma is an authentic, ill-structured problem that requires learners to develop and apply specific concepts as they construct possible solutions to the problem (Mikeska, Koehler, Weizman, & Lundeberg, 2007). Participants worked in groups exploring the dilemmas using a PBL analysis strategy (Torp & Sage, 2002). This framework encouraged an evaluation of known information, generation of learning issues and hypotheses, research using literature and/or experimentation with materials to gain a deeper understanding of the concepts related to the problem, and discussion of application of these concepts to the problem. The content learning was followed by four days of curriculum development where teachers used their new content knowledge to plan an instructional unit to be used in the following school year. This work was followed by three days in the summer workshop and monthly meetings during the academic year where groups of teachers worked to study and improve their teaching as they implemented their new units.

Data Collection

The focus of this paper is on the three days of content learning, and the pre- and post-assessment of teachers' content knowledge within each content strand. The pre-assessments were administered during an orientation meeting in April before the beginning of the summer PWC. The same assessments were administered at the end of the PWC following three days of content learning and four days of unit plan construction for use in their classrooms. In each case, teachers wrote for about half an hour. Responses ranged from 3 lines to 3 pages.

Design of the Assessments – General and Application Qs

In order to reflect content knowledge for teaching, two types of open-response questions were designed for each content area by the content experts who designed the corresponding learning activities for that content strand.

General essay questions. One or more questions were included in each assessment that elicited teachers' general knowledge of a concept. These questions were intended to be broad enough that teachers would be able to display whatever knowledge they had. However, at least some part of the question or set of questions explicitly asked for explanations. The intent for these questions was to assess the breadth of teachers' knowledge. Because omissions in responses are difficult to interpret (did the teacher not know about or simply not write about a particular concept?), many of the general questions included specific instructions about the desired scope of the response. The following is an example of a general essay question from the Ecology strand: "What factors might influence the future biodiversity in Michigan ecosystems? Provide examples and explanations. How might the loss of biodiversity affect human welfare?" The complete set of questions used is shown in the supplemental materials.

Application questions. Assessments in the first cohorts revealed that, when asked a general essay question, teachers often gave a brief list of examples relating to a concept without significant explanations. This trend made it difficult to discern if teachers did not know the explanations or were just being overly brief in their answers. In the 2008 cohort, we addressed this by including one or more questions that assessed the teacher's ability to explain why specific events happen in a given context. These application questions included a scenario that described a specific situation and included reference to a least one important science concept, then asked the writer to explain the phenomenon described in the scenario. Earlier research had suggested that teachers were more likely to reveal a more complete picture of their understanding in their written responses to this type of question. The intent of the application questions was to assess the teachers' ability to accurately and completely use the concepts in a specific context. An example of an application question from the Weather Strand was: "Using the diagram, explain why Holland (MI) experiences greater lake-effect snows than Lansing. In your explanation discuss phase changes in water, air pressure and density, land water and air temperatures, and movement of air."

Scoring and analysis

Pre- and post-assessments blind scored. The assessments and content dilemmas were planned so as to focus on a limited set of "Big Ideas" that are essential for understanding the concepts addressed in the PD. The Big Ideas were aligned with the state and national standards for each topic (MDOE, 2008; NRC, 1996; AAAS, 1993). Project facilitators who were experts in each specific content area wrote ideal responses for each assessment item. The ideal response served as the benchmark to which teachers' responses were compared.

Qualitative coding scheme. Teachers' responses to the pre- and post-tests were analyzed using a qualitative coding scheme (Sanders, 1993). The coding scheme described here was designed to work across content strands and types of questions and was suited to characterization of teachers' incoming content knowledge and change in content knowledge associated with the PD. The coding also allowed identification of Big Ideas that were more challenging for the participants. Through three iterations of PD, we developed the coding scheme shown in Table 1 to score for the presence and quality of each Big Idea in a teachers' response. For further explanations of codes for sample teacher responses, see the supplemental materials.

Table 1
Coding scheme for analyzing open-response items

Code	Definition	Description
NP	Not Present	Concept is not present in the response.
I	Inaccurate	Concept is inaccurate; exhibits misconception.
C	Confused	Answer is confused, vague, or offers too little information to understand what respondent knows about the concept.
AI	Accurate, but incomplete	Response is accurate, but lacks important information; answer is incomplete.
AC	Accurate and complete	Response is accurate, and completely addresses the Big Idea.

Coding. The analysis of responses began by transcribing, de-identifying, and scrambling all responses (pre and post PD) to allow blind coding. Teams of three researchers then coded each response. The teams included at least two experts in the content covered by the strand. The experts were faculty members from a local university and community college, and included the facilitators who led the summer workshops and administered the assessments. For some groups, the third researcher was an expert from a related field of science who was not one of the strand facilitators. Through trial and error, we found that a team of three coders worked most effectively, and it was important to include one team member who was from outside the content strand but from within the project. Discussions prompted by questions from this individual usually led to explicitly stated rationales for assigning a code.

The data included responses to pre- and post-tests, for seven different content strands. Coding of responses began with individual team members assigning a code for each Big Idea associated with a response. The team members then discussed each response and reached a consensus on codes for each response. We chose to ask teams to come to a consensus score for each response rather than coding separately and determining inter-rater reliability primarily for practical considerations. First, there were fewer than a dozen teachers in each strand with an average of nine teachers per strand. This meant that there were too few responses to do two rounds of inter-rater calibration and still have a significant number of responses to score and test inter-rater reliability. In addition, after coding responses from three groups early in the process, it became clear that the nature of the assessments and the content strands would require an inordinate amount of time to train coders (or "recalibrate") about the interpretations we used to code responses. The experts performing the coding quickly developed an efficient process for discussing answers and reaching a consensus on the codes to assign. The consensus scores and the team's explanatory comments were recorded for further comparison.

Compiling data

Compilation by teacher. After coding teachers' responses, the items were re-identified and entered into a spreadsheet to display all the codes assigned to a single teacher's responses to the pre- and post-PD questions. See for example, Table 3. These spreadsheets allowed us to easily view patterns in a teachers' incoming knowledge as well as changes in the teacher's content knowledge over time.

Compilation by content strand. The spreadsheets for individual teachers were grouped by content strand, and counts of responses coded in each category and changes in scores from pre- to post-test were compiled by Big Idea. These data allowed researchers to identify patterns within the strands. For instance, if a specific Big Idea was scored "NP" for all assessments in a strand, questions were raised about the learning activities and the assessment questions to see if they addressed the Big Idea adequately. Researchers were also able to identify Big Ideas that were consistently difficult for teachers to explain, revealing areas of need for further learning.

Results

Four Case Studies

In the following section, we share examples of teachers' coded responses to assessment questions and the compiled data to illustrate the kinds of inferences afforded by the open-ended assessments and the coding scheme described above. The research questions addressed are: 1) What information about teachers' content knowledge can this method yield? 2) What are the strengths and limitations of this assessment approach? We present four cases selected to illustrate different patterns of teacher content knowledge and learning that are evident from the data analysis. In addition, we note conclusions about the teachers' content knowledge that can be drawn only from the raw data. The four cases describe teachers, using pseudonyms, from three different content strands and from elementary and middle schools.

Mrs. Brinkman's case. At the time of this study, Mrs. Brinkman was a sixth grade teacher in an urban middle school. She was 29 years old and had six years of experience. She was a biology major with additional certification in general science who taught earth science because of changes in the state's curriculum requirements. She elected to participate in the weather strand. Table 2 includes her responses to a general question about how thermal energy and radiation create weather patterns along with the Big Ideas expected in the "ideal" response written by the activity designers. Codes for each Big Idea assigned to the responses are summarized in the right hand column. The portions of the responses most relevant to each code are indicated with brackets.

Table 2

A general essay question, the Big Ideas expected in ideal response, Mrs. Brinkman's responses & assigned codes

Q. Using the words thermal energy and radiation explain why we have weather? (Think about seasons, the water cycle, air masses and predicting the weather.)

Big Ideas expected in an ideal response

1. The sun's radiation is absorbed unevenly by the atmosphere, land & water.
2. The tilt of Earth's axis causes different latitudes to receive different light intensities (due to differences in angle) and duration.
3. Water changes phase due to absorption of the sun's radiation or transfer of thermal energy.
4. Thermal energy is redistributed when water changes phase.
5. Transfer of thermal energy between land, air & water affects air pressure and density.
6. Air moves from high to low pressure.
7. When air masses with different characteristics meet, weather changes.
8. Air masses take characteristics of area over which they develop.
9. Severe weather occurs when 2 colliding air masses have greater differences in pressure, temperature & humidity.
10. In the US, weather moves west to east.
11. Air pressure, density, & temperature are related. Cooler air is denser & has higher pressure than warmer air.

Mrs. Brinkman's pre-PD response	**Scores**
[The atmosphere gets warmed by the sun's radiation and conduction(?)**]** ***(BI#1)*** **[**Warm air is less dense and rises. Cool air is more dense and sinks.**]*****(BI#11)*** In high pressure areas – cool air sinks and is warmed. The air can hold more moisture. In low pressure areas – warm air rises and cools and can't hold as much moisture. I know an air mass is a body of air with a certain temperature and fronts are what separate air masses.	AC/AI (BI#1,11) NP (BI#2-10)
Mrs. Brinkman's post-PD response	**Scores**
[The Earth is heated unevenly. Land, water and air absorb the sun's radiation differently and convert it to thermal energy.**]*****(BI#1)*** **[**The heating of these influences air temperatures above them.**]*****(BI#5)*** **[**As air is heated and it cools, warm air rises and cool air sinks. Cool air is denser and has a higher pressure than warm air.**]*****(BI#11)*** **[**Air moves from an area of higher to lower pressure and this movement globally affects weather (jet streams, westerlies, easterlies).]***(BI#6)*** **[**Air masses meet along fronts and this meeting causes weather to occur.**]*****(BI#7)*** **[**The bigger the differences in air mass temperature, humidity and air pressures, the greater the chance of severe weather.**]*****(BI#9)*** [The Earth is tilted as it revolves around the sun. This causes the seasons to occur because the Earth is receiving different amounts of direct and indirect sunlight, depending on the latitude of a location.**]*****(BI#2)***	AC/AI (BI#1,2, 5-7, 9,11) NP (BI#3,4,8, 10)

Table 3 shows codes assigned to Mrs. Brinkman's answers across both general and application assessment questions. This table is essential in creating a summary of each teacher's understanding of the science concepts. In this table, scores from both types of questions were compiled into a cumulative score that allows researchers to quickly identify patterns. In her pre-PD responses, Mrs. Brinkman had very little incoming content knowledge. She demonstrated "accurate but incomplete" knowledge (coded "AI") for Big Ideas 1 & 11. She was able to connect the sun's energy and factors such as temperature and moisture to weather patterns. She also gave an unclear explanation (coded "C" for Big Idea 3) of what makes water evaporate. Other Big Ideas that were relevant to the question were coded as "NP," indicating that they were not included in her pre-PD response.

Table 3

The scores by Big Idea (BI) assigned to Mrs. Brinkman's responses to general (Gen) and application (App) questions pre- and post-PD

	Pre			Post		
BI\Q Types	Gen	App	**CUM**	Gen	App	**CUM**
1	AI	NP	**AI**	AC	NP	**AC**
2	NP		**NP**	AI		**AI**
3	NP	C	**C**	NP	AC	**AC**
4	NP		**NP**	NP		**NP**
5	NP	NP	**NP**	AC	NP	**AC**
6	NP	NP	**NP**	AC	AC	**AC**
7	NP		**NP**	AC		**AC**
8	NP		**NP**	NP		**NP**
9	NP		**NP**	NP		**AC**
10	NP	NP	**NP**	NP	C	**C**
11	AI	NP	**AI**	AC	AI	**AI**
12	NP		**NP**	NP		**NP**
13	NP		**NP**	NP		**NP**
14		NP	**NP**		NP	**NP**

Note. Scores: AC = accurate & complete; AI = accurate but incomplete; C = confused; I = incomplete; NP = not present. Blank cells indicate that the Big Idea was not expected in the response to that particular question. CUM = scores tabulated across a test

In contrast, after the professional development, Mrs. Brinkman wrote more clearly about one Big Idea and accurately used five additional ideas. Two of these ideas included explanations of the energy sources that drive weather events. It was unclear whether she understood the role of the prevailing winds in local weather. She never addressed two ideas: that thermal energy is redistributed when water changes phase and how air masses take on the characteristics of the place in which they form. With very few exceptions, Mrs. Brinkman expressed her ideas clearly, both before and after the PD. A careful reading of her responses reveals that she structured her responses as a series of cause and effect statements.

Mr. Cooper's case. Mr. Copper was 32 with 8 years of teaching experience. He taught fifth grade in a middle school in a rural district. He had elementary certification with a science specialization. He elected to participate in the Astronomy strand. In this strand, planners included one general and five application questions in the content assessment. Mr. Cooper's responses to these assessment questions are summarized in Table 4. Questions addressed the causes and patterns of the changing phases of the moon, solar and lunar eclipses, and planetary and stellar motion.

In the pre-test, Mr. Cooper was able to use 12 Big Ideas. However his responses were short and his explanations of about half of the ideas were unclear (Coded as "C") or inaccurate (I). After participating in the professional development, Mr. Copper used 14 Big Ideas accurately and usually completely. This suggests that Mr. Cooper's learning did not take the form of adding many new ideas, but he was able to clarify and correct several concepts he was already familiar with. This pattern differed from that illustrated in Mrs. Brinkman's case, but both are important examples of teacher learning.

To understand how Mr. Cooper's ideas changed, it is helpful to examine specific questions in more detail. Table 5 gives a detailed description of the coding for his responses to the general essay question. This question addressed four of the fifteen Big Ideas. In this question, teachers were asked to include a diagram, and Mr. Cooper's drawings are included with the responses.

Table 4
The scores by Big Idea (BI) assigned to Mr. Cooper's responses to general (Gen) and application (App) questions pre- and post-PD

BI\			**Pre**								**Post**				
Qtype	Gen	App	App	App	App	App	CUM	Gen	App	App	App	App	App	CUM
1	C						C	AC						AC
2	I						I	AC						AC
3			C				C			AC				AC
4	I						I	AC						AC
5			AI				AI			AC				AC
6			C				C			AC				AC
7	NP		AI				AI	AI		AC				AI
8				NP			NP				AC			AC
9				AI			AI				AC			AC
10				I			I				AC			AC
11			NP				NP			NP				NP
12						AI	AI						C	C
13		AC			NP		AC		AC			NP		AC
14						AI	AI						C	C
15					NP		NP					AC		AC

Note. Scores: AC = accurate & complete; AI = accurate but incomplete; C = confused; I = incomplete; NP = not present. Blank cells = Big Idea was not expected in the response to that particular question. CUM = scores tabulated across a test.

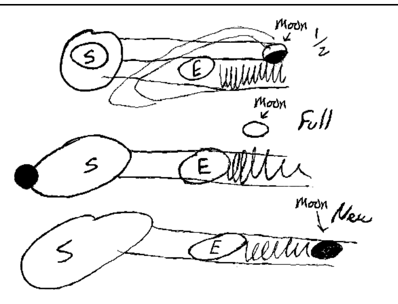
Table 5
A general essay question, the Big Ideas expected in an ideal response, Mr. Cooper's responses and assigned codes. In the responses, the particular portions that were used to determine the codes for each Big Idea are indicated with square brackets.

Q. Why do we see the moon go through phases, and why do they change? Explain clearly in words and support your explanation with a well-labeled diagram.

1. The Moon shines by reflected sunlight
2. Observers on the Earth only see the part of the Moon that is illuminated by the Sun *and* that is facing the Earth.
4. The phases of the Moon change because the Moon is revolving around the Earth.
7. The phases of the Moon (its appearance, position in the sky, and rise and set times) change in a regular and predictable pattern.

Mr. Cooper's Pre-PD Response — **Scores**

[Sun/Earth/Moon alignment creates phases in moon.]**(*BI#4*)** [Earth's shadow is cast at different angles onto moon's surface.] **(*BI#2*)**



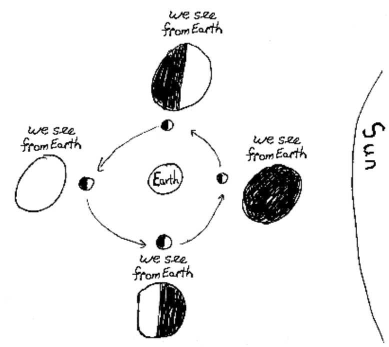
C (BI#1)

I (BI#2, 4)

NP (BI#7)

Mr. Cooper's Post-PD Response — **Scores**

Phases are caused because 1) [the Moon isn't always in the same location with respect to the Earth — AC/AI

& Sun,]*(**BI#4**)* 2) **[**the moon reflects light from the sun, not produce its own light.**]*(BI#1)*** **[**Phases are caused because the lighted part of the moon faces the sun & we see different amounts of it due to position.**]*(BI#2)***		(BI#1,2,4) NP (BI#7)

Mr. Cooper's pre-PD writing and diagrams revealed some misconceptions about the phases of the moon, including the idea that the phases of the moon are caused by the earth's shadow falling on the moon. His written response was brief, but clearly stated that the earth's shadow cast on the moon results in different phases. In his diagram, you can see the light from the sun meeting the earth, with the moon being partially blocked. The angle of the light side of the moon does not relate to the angle at which the moon is viewed from the earth.

Post-PD, his responses were not longer, but they were clearer and more accurate. His statement that phases "are caused because the lighted part of the moon faces the sun and we see different amounts of it due to position" suggests an important shift in his conceptual understanding compared to his pre-test. This was true of both his descriptions of what happens and his explanations of why they happen.

Ms. Schuler's case. Ms. Schuler was 30 and had 5 years of teaching experience. She earned a degree in elementary education, and was certified to teach grades K – 8. She taught seventh grade in a rural/suburban charter school. She had participated in the professional development program the two previous years in different content strands. In her third year, she chose to enroll in the Astronomy strand, in part because of a change in the state science standards that would require her to teach more geology and astronomy than in previous years. Her responses reveal a third pattern seen in the PD program. Table 6 shows a summary of codes assigned to her responses across all the assessment questions.

Table 6

The scores by Big Idea (BI) assigned to Ms. Schuler's responses to both the general (Gen) and application (App) questions pre- and post-PD

Q type /BI	**Pre**							**Post**						
	Gen	**App**	**App**	**App**	**App**	**App**	**CUM**	**Gen**	**App**	**App**	**App**	**App**	**App**	**CUM**
1	NP						NP	AC						AC
2	NP						NP	AI						AI
3			NP				NP			I				I
4	C						C	AC						AC
5			C				C			NP				NP
6			I				I			I				I
7	NP		C				C	AC		AC				AC
8				NP			NP				NP			NP
9				AI			AI				I			I
10				NP			NP				C			C

11		NP				NP		NP				NP
12					NP	NP					I	I
13	C			NP		C	I			NP		I
14					NP	NP					I	I
15				I		I				C		C

Note. Scores: AC = accurate & complete; AI = accurate but incomplete; C = confused; I = incomplete; NP = not present. Blank cells = Big Idea was not expected in the response to that particular question. CUM = scores tabulated across a test.

On the pre-PD assessment, she used six Big Ideas, but only one of these was clear and accurate. This was a descriptive idea that people in different parts of the US are able to see the same lunar eclipse. The other five Big Ideas she discussed reflected inaccurate or confused ideas. Her post-PD responses showed evidence of learning, including the addition of some new ideas and clarification of concepts that had been inaccurate or confused. However, some of her new ideas, such as Big Ideas 3 and 10, were confused or inaccurate. This pattern suggests that Ms. Schuler was still working to assimilate the new science concepts, and that the ideas were still tentative or incomplete.

A more detailed examination of Ms. Schuler's responses shows the development of her ideas. Table 7 quotes the question and Ms. Schuler's responses to the general question from the Astronomy strand assessment instrument.

Table 7

A general essay question, the Big Ideas expected in an ideal response, Ms. Schuler's responses and assigned codes

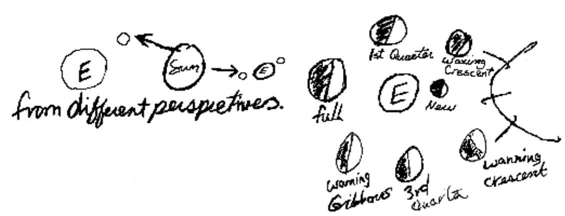
Q. Why do we see the moon go through phases, and why do they change? Explain clearly in words and support your explanation with a well-labeled diagram.

Big Ideas expected in ideal response

1. The Moon shines by reflected sunlight.
2. Observers on the Earth only see the part of the Moon that is illuminated by the Sun *and* that is facing the Earth.
4. The phases of the Moon change because the Moon is revolving around the Earth.
7. The phases of the Moon (its appearance, position in the sky, and rise and set times) change in a regular and predictable pattern.

Ms. Schuler's Pre-PD response	**Scores**
We see the moon from different **[**positions as we rotate. The moon rotates around us.**]** (***BI#4***) I have no idea. Sorry.	NP (BI#1,2,7) C (BI#4)

Ms. Schuler's Post.-PD response	**Scores**
The moon goes through phases because here on earth **[**we see the side of moon that is lit by the sun**]** (***BI#1***) so when the three are aligned we see no moon which is the new moon phase. **[**As we rotate, the moon revolves around us**]** (***BI#4***) and **[**we see the moon lit from different perspectives. **]** (***BI#2***)	AC/AI (BI#1,2,4,7) Score for BI#7 based on diagram



Ms. Schuler's pre-test responses suggested that she knew very little about why we see moon phases or how to explain the motion of other heavenly objects. Table 7 shows that she mistakenly associated moon phases with the fact that the earth rotates. Her statement that "The moon revolves around us" implies knowledge of some part of Big Idea #4, but does not clearly associate the movement of the moon with the changing phases. This response was coded as "confused or ambiguous." She also clearly stated that she did not know much about the topic, and could not draw a diagram to illustrate her ideas.

Her post-test response to the same question includes more accurate information, suggesting that she corrected some of the inaccurate ideas seen in her earlier response. She also wrote about concepts not addressed in the pre-test. In her response to this question, she used the new ideas accurately. She drew an accurate diagram explaining the phases of the moon. However, in response to a different question, she was not able to apply her new knowledge of the moon phases to explain which phases people in different places and times would see. Her understanding of planetary and constellation motion remained inaccurate. Her case illustrates science concepts that are beginning to develop, but are fragile.

Mrs. Turkus'case. Mrs. Turkus was 56 with 14 years of teaching experience. She had elementary certification and taught sixth grade in a rural school. During the professional development, she participated in the geology strand for secondary teachers. In this strand, the assessment included one general question and two application questions. The questions were connected to ideas about the morphology of rocks and minerals as a way to understand plate tectonics and the formation of rocks. Table 8 is a summary of her coded responses on the three assessment questions.

Table 8
The scores by Big Idea (BI) assigned to Mrs. Turkus' responses to both the general (Gen) and application (App) questions pre- and post-PD

	Pre-test				**Post-test**			
BIs\ Q type	**Gen**	**App**	**App**	**CUM**	**Gen**	**App**	**App**	**CUM**
1	NP			NP	NP			NP
2	C			C	I			I
3	NP			NP	I			I
4	NP			NP	C			C
5	NP			NP	C			C
6		I		I		AI		AI
7		NP		NP		AI		AI
8		C		C		NP		NP
9		NP		NP		NP		NP
10		NP	NP	NP		NP	NP	NP
11			NP	NP			AI	AI
12			I	I			C	C
13			NP	NP			NP	NP

Note. Scores: AC = accurate & complete; AI = accurate but incomplete; C = confused; I = incomplete; NP = not present. Blank cells = Big Idea was not expected in the response to that particular question. CUM = scores tabulated across a test.

Mrs. Turkus' pre-test responses suggest that she had very little prior knowledge about the concepts addressed in the strand. She used only 5 of the 13 Big Ideas and these were confused or inaccurate. After participating in the professional development, she was able to use 3 ideas accurately. One of these ideas (Big Idea #1) was a clarification of an idea she had used before. The other two (Big Ideas #7 and 11) were new concepts that she had not used previously.

A more detailed look at Mrs. Turkus' responses led to some insights into her understanding. Table 9 shows the question, Big Ideas and responses to a general question about the historical information that can be inferred from physical features of a rock.

Table 9

A general essay question, the Big Ideas expected in an ideal response, Mrs. Turkus' responses and assigned codes

Q. It is said that the history of the Earth is written in the rocks. How do rocks tell the geologic history of an area? Include the rock type, age, texture of the rock (at the hand specimen level as well as regional scale), the general composition, relationship to other rocks in the environment, and the rock forming processes involved.	
Big Ideas expected in ideal response 2. Igneous, metamorphic and sedimentary rocks are indicators of geologic and environmental conditions and processes (cooling and crystallization, weathering and erosion, sedimentation and lithification, metamorphism). 3. Rocks tell us about their origin and history through texture. For example, Grain size indicates cooling rates, distance from source. The texture of grain indicates degree of metamorphism. 4. Composition of rocks provides clues to the tectonic setting. 5. Different rock assemblages exist in different tectonic settings.	
Mrs. Turkus' response pre-PD Information is trapped in many ways within the earth. There are different types of rock based on its origin and the process it went through. These processes report happening by trapping the different types in different layers at different times during different conditions. **[**Metamorphic, Sedimentary, Igneous - All vary because of makeup, pressure, texture, etc.**]*(BI#2)*** These groups can be subdivided into additional groups based on characteristics of each type. Layering occurs due to pressure and exposure to the environment and the materials surrounding the rock.	**Scores** C (BI#2) NP(BI#3-5)
Mrs. Turkus' response post-PD Rocks tell the age of geologic history because they show how a particular rock formed. You can identify the type of rock and how old it is because of the sequence it went through in the rock cycle.**[** If it was cooled quickly or slowly it will indicate how the particles in the rock are "cemented" together. If it was "hot hot" very few particles.**](*BI#3*)[** If sedimentary, many particles gathered over time.**](*BI#2)*** **[**You can look at a rock in a given area and identify if it was pushed and shoved into that particular place or did it arrive via hot molten lava to spread out over that surface.**](*BI#4*)** **[**You may also find combinations of these when a hot spot erupts through the crust possibly due to plates moving and subducting, converging or diverging that material to new heights**](*BI#5*)** (ex. Oceanic-continental plates create subducting and rift building). Over billions of years earth continues to change. As the earth continues to cycle core, mantle, crust, nothing replaced just changed and moved to a new job or a new spot.	**Scores** I (BI#2,3) C (BI#4,5)

In her pre-PD responses, Mrs. Turkus wrote a number of vague references to concepts, but this seemed to be a recapitulation of ideas presented in the stem of the question. The only idea she addressed was about the three types of rocks, but she did not provide an explanation of how the environment in which the rock was formed can be inferred from the rock type. This Big Idea was coded "C" because of the ambiguous answer.

In the post-PD responses, Mrs. Turkus wrote more directly about answers to the question. More specifically, she was clearer on what was occurring at convergent and divergent plate boundaries. She used new ideas dealing with volcanic activity at both types of plate boundaries. She used an additional three new ideas inaccurately or in a confused way. When explaining these ideas, she made erroneous connections between phenomena and explanations or used vague language. For example, she associated rapid cooling with cementation of rock particles rather than crystal size. She said that you could "look at" rocks and tell something about their tectonic history without specifying what specific rock features indicates. These problems may indicate learning in progress or learning aimed at a narrative rather than a mechanistic explanation.

There were also clues that Mrs. Turkus struggled with vocabulary. On her pre-test, she appeared to confuse convergent and divergent boundaries. Even after the PD activities, she wrote on her post-test, "I just can't keep metamorphic and igneous straight." The data in the cumulative coding table (Table 8) cannot illustrate what she found difficult, but they do point to the responses that may expose the nature of the difficulty.

Summary of Cases

The four cases described above illustrate the type of inferences afforded by the content analysis methodology explained in this paper. The compilation of teacher scores facilitates assessment of the strength of teachers' incoming knowledge and changes in their knowledge both in terms of number of Big Ideas and the clarity, accuracy, and completeness of that use. For example, Mrs. Brinkman came with very little knowledge of weather. She used only three of fourteen Big Ideas initially. This is in contrast to Mr. Cooper's incoming knowledge of astronomy. He used twelve of fifteen ideas, but his explanations for half of them were not clear.

The four cases were representative of common patterns of change in teachers' content knowledge. Mrs. Brinkman's learning took the form of addition of new ideas. She went from using three ideas to accurately using eight ideas. Mr. Cooper's content knowledge improved mainly through clarification of ideas that were already in his repertoire. He used six ideas clearly and accurately that he had previously described in a confused or inaccurate way. He also used two new ideas that were not in his earlier writing, but his description of two ideas that he used accurately on his pre-test were confused on the post-test. Ms. Schuler and Mrs. Turkus are examples of teachers whose new knowledge was confused. Ms. Schuler used four ideas accurately (primarily having to do with phases of the moon) that she had not used or had used in a confused way on her pre-test. However, she used four new ideas inaccurately or in an unclear way. Mrs. Turkus made smaller gains in her content knowledge. She used two new ideas accurately, clarified one idea, and added three confused or inaccurate ideas.

When scores are compiled by Big Idea, they can alert teacher educators to Big Ideas that are difficult for teachers and Big Ideas not adequately addressed by the PD or the assessment. For example, like Ms. Schuler, six of the nine teachers in the astronomy strand did not accurately or clearly use the idea that because of the earth's rotation, the moon appears to move from east to west across the sky daily. This idea was clearly called for in one of the application questions and participants had discussed the idea during their problem-based learning sessions. This suggests that they needed more opportunities to work with the idea. The situation was somewhat different in the weather strand. Only one teacher explained how changes in temperature affect air density at the molecular level. While this might mean that participants needed more experiences with this idea, it is likely that the problem lies with the wording of the question (see Table 2) which directed teachers to include seasons, the water cycle, air masses and predicting the weather in their responses, but did not ask for a molecular explanation.

Each of the four cases also illustrates the type of inferences that do not arise directly from this analysis strategy. These are more subtle characterizations of the ways in which teachers are thinking about content. For example, Mrs. Brinkman formulated all of her responses in the form of cause and effect statements. Mr. Cooper clarified his responses, not by writing more, but by using more precise language and in one case, clearing up a misconception. In her response to the general question on the post-test, Ms. Schuler seemed to have a strong grasp of why we see changing phases of the moon, since she could draw the correct configuration of the sun, moon, and earth from two different perspectives. However, she was unable to transfer this knowledge to an application question about lunar eclipses. Mrs. Turkus appeared to struggle with vocabulary.

Discussion

The cases described above are presented to illustrate the usefulness of the assessment strategy implemented in the PBL Project for Teachers of Science. For professional development planners and science teacher educators, the task of revealing teachers' science content and the concepts learned during PD activities is important. We posit that the assessment strategy discussed here offers a method for studying the depth of teachers' content knowledge that can be applied to a range of science topics and a range of levels of understanding (for example, the understanding of elementary teachers as well as of secondary teachers.) While these cases are from three content strands, the data collected in the PBL Project includes similar cases from a total of seven strands.

In this section, we discuss inferences and assertions about the utility of the assessment strategy, as well as some of the limitations and questions raised from our analysis that present opportunities for future analysis and research. The discussion address the two research questions described earlier in this article.

What can this assessment tell us about teacher content knowledge?

The assessment and analysis method used in the Professional Working Conference of the PBL Project revealed several types of information about teacher content knowledge that are useful in planning and implementing teacher professional development. The data are helpful in characterizing the science knowledge of teachers before the professional development activities began. In addition, comparison of pre- and post-PD responses using the coding scheme allows researchers and educators to assess the new science ideas learned or clarified. When patterns across a content strand are compared, the data also highlight Big Ideas that are consistently difficulty for teachers to learn during the PD activities.

Beginning knowledge of teachers entering a PD program. In any educational setting, it is essential to access learner's prior understandings (Chiapetta & Koballa, 2010). This type of "assessment FOR learning" (Stiggins, Arter, Chappuis & Chappuis, 2006) allows the teacher to plan instruction to address the needs of the learners. Just as this principle is addressed in educational methods courses, professional development planners also need to base their instruction on the needs of learners. In the PBL Project, the assessment questions were given prior to the planning of the summer learning activities, and an informal review of responses revealed the nature of teacher content knowledge for each specific strand. This information was used to tailor the instructional problems so that participating teachers would have enough initial knowledge to get started on the problems, but the problems would also deepen their understanding of the content.

The two types of open-ended assessment questions were effective in revealing the depth of understanding of teachers prior to the PBL learning activities. Unlike a concept inventory, the written responses gave insights into the breadth of concepts upon which teachers could draw when answering the general questions. The application questions then asked for a more detailed explanation of specific concepts that incorporated some of the Big Ideas for the strand. Teacher responses to these questions elicited enough detail to expose misconceptions, confused ideas, and gaps in teacher understanding that multiple-choice questions might miss. For instance, in Mr. Cooper's case, the general question asked why the moon changes phase. In a concept inventory, an incorrect answer might indicate that the he did not have an accurate understanding of the phases of the moon, but because the answer might be a guess, it fails to give a clear picture of the ideas he actually held. In his pre-test response, Mr. Cooper wrote that the "Earth's shadow is cast at different angles on the moon's surface." This makes it very clear that he attributes the phases to the earth's shadow, not the relative position of the sun, moon and earth. Once the teacher educators know this information, they can plan activities that directly address this concept.

Changes in teacher content knowledge after intervention. One of the most important ways to assess the efficacy of a professional development program is to demonstrate that learners have learned. The assessment strategy used by the PBL Project accomplished this by administering the assessment questions several months prior to the summer workshop and after seven days of professional development. Between the pre- and post-tests, participants engaged in a series of problem-based learning activities over three days. By comparing the codes for pre- and post-PD responses to the same questions, we were able to see how teachers ideas changed, new ideas were introduced, or in some cases, ideas remained stable.

An example of the pre-post comparison provided by the assessment instruments can be seen in the example of Mrs. Brinkman. In her case from the Weather strand, her general question response on the pre-test showed that she only included the sun's energy and the density of air as it changed temperatures as factor that cause weather. In the post-test, she accurately included five other Big Ideas and applied four of these to a specific situation.

Ms. Schuler's case illustrates that not all of her learning was as robust as Mrs. Brinkman's. In her Astronomy pre-assessment, she wrote that phases of the moon change as the earth rotates. In the post-test, she was able to draw an accurate diagram to illustrate the phases as a result of the position of the earth in relation to the moon and sun. While this clearly revealed that she had learned what causes the phases, her written response to an application questions suggested that she could not extend her understanding to predictions about the visibility of the moon's phases from different places. Not only did we see that she had learned some important concepts, but the scoring helped point out areas for continued learning. Both of these are important outcomes for teacher educators to identify.

Which ideas are difficult for teachers? The assessments also revealed the "Big Ideas" that were most difficult for teachers to learn. This was evident by examining responses for individual teachers and identifying patterns in responses for the entire group of participants in each content strand. When most or all of the teachers were unable to improve their responses to include certain ideas, the researchers examined the assessment questions and the learning activities to explain the apparent lack of learning. When the assessment addressed an idea that had been included in learning activities, but still did not result in changes in the teachers' understanding, the PD activity was reviewed to find possible explanations.

Many of these challenging Big Ideas are concepts that depend on prerequisite knowledge that the teachers were developing during the learning activities. For instance, teachers in the Astronomy strand found it difficult to predict the phases of the moon over time or in different places. The ability to predict the appearance of the moon depends on first understanding what causes the phases, the movement of the moon around the earth, and the rotation of the earth on its axis. There is a clear learning progression (Plummer & Slagle, 2009), and the teachers are still assimilating new ideas about the cause of the moon's phases. They may not be ready to learn related, but more advanced, ideas about the movement of the solar system. Knowing which ideas pose the biggest challenge for teachers and professional development leaders is important because it reveals areas of need. These needs can then be addressed by further professional development efforts and supplementary curricular materials.

What are the strengths and limitations of this assessment approach?

Because of the variety of content areas represented in the different strands, one of the challenges facing PD planners is the development of assessments that are similar enough to allow a comparison of results across groups, yet flexible enough to meet the varied assessments needs for a range of science topics. The design of the assessments described here has been effective in finding a balance between flexibility and consistency.

By structuring the pre- and post-tests to include two types of questions (broad open-response questions, application questions), we were able to examine patterns across groups in ways that let us draw inferences about our professional development model. This layered approach to assessment offers a useful framework for researchers who need to evaluate the effectiveness of a specific teaching strategy, such as problem-based learning.

Validity. Early versions of the project's assessment instruments included multiple-choice items from concept inventories and standardized tests. The results showed that it was extremely difficult to find questions that targeted the Big Ideas associated with the learning goals. Standardized selected-response questions are better suited for assessing recognition of a wide range of goals, and not as valid for revealing a teacher's ability to explain and apply a small set of Big Ideas.

One of the strengths of the PBL Project's assessment strategy is the emphasis on making the assessments valid. The content experts who designed the learning activities first identified the learning goals for the activity, then developed assessment items that would target these ideas. The same planners then wrote ideal responses that revealed the Big Ideas that reflect deep understanding of the content.

Because of the strong validity of this approach, we posit that this assessment strategy is a promising stepping stone that can be used to build assessments that examine the ideas emphasized in the new Core Standards, especially scientific reasoning and the models that teachers use to explain concepts. The written responses teachers generated in the project resemble the types of information teachers need to be able to generate in a classroom situation. The ability to explain a concept and apply it to a specific context is central to designing and delivering instruction, and in evaluating students' work for content accuracy.

Reliability. Another important consideration is reliability of the coding process. We found that because of small sample sizes and the small number of test items, inter-rater reliability was difficult to assess. By the time a group of coders reached a common view for interpreting answers, all the responses in a group had been coded. As a result, researchers elected to have the group of three coders reach a consensus on codes assigned to responses. When other coders familiar with the content area revisited the codes later only a very small number of discrepancies were found. This suggests that the consensus coding is a reliable method for analyzing responses.

Limitations of the Assessment Strategy

While the assessment strategy offers several advantages over selected-response tests, there are limitations to the method. Among these is the time needed to code responses. Open-ended assessments are, by their nature, more time-consuming to score than selected-response tests. Analysis requires more attention to individual responses. This strategy would not be appropriate for comparing content knowledge across a very large sample. Coding requires a thorough reading of responses and discussion among coders. However, the strategy chosen for the PBL project offers insight into aspects of teacher thinking that selected response questions cannot provide. PD providers sometimes must choose between efficiency and depth of information provided by various data sources.

There may also be limits to the degree to which the assessments allow diagnose of the source of misconceptions and gaps in understanding. This is illustrated in Mrs. Turkus' case. Comparing her coded responses from pre- and post-tests showed that she struggled to accurately use the terms "divergent" and "convergent" in explanations about plate tectonics. The codes do not tell if the difficulty is a misconception or a lack of familiarity with vocabulary. Still, the coding tables do provide cues to the researcher or teacher educator that can lead to a more thorough examination of written responses.

Even with multiple forms of assessments given at intervals, we were unable to get a perfect picture of what teachers know. Many of the participants in the PBL Project responded to questions based on what they think we "want" to see in their answers, just as many of their students have learned to do. Others were not motivated to write more than a minimal answer. The short responses failed to completely answer the assessment questions, leading to many scores of "Not Present," even though discussions during learning activities suggested the teachers knew much more about a concept than was revealed in their writing. Despite instructions to write complete sentences, some teachers only provided bulleted lists of words or phrases. Nehm and Schonfeld (2008) also wrote about that participants' "aversion to writing" resulting in "limited responses or errors of omission, and poor writing skills may hamper clear communication, preventing the instructor from recognizing the extent of the student's knowledge..." (p. 1175). Alternatively, after the demands of the PD which included a lot of writing, teachers may have experienced writers' fatigue and shortened their post-PD responses. While these teachers may have a deep and accurate understanding of a concept, their writing sometimes does not reflect their understanding.

The limitations are important in making decisions about the use of the assessment strategy we describe. As in any learning situation, the purpose of the assessment and the setting in which it will be used should guide choices made by the educator. Our findings support the use of the strategy described here as an effective method for assessing science content knowledge.

Further Questions Arising from the Cases

While the cases presented here illustrate some of the information that can be collected from the assessment strategy described above, there are also questions raised by the cases. In the following section, we discuss some of the questions still to be explored using the data collected in the PBL Project for Teachers of Science. Some of the questions foreshadow future publications, and others point to some limitations of the assessment strategy. Still other questions suggest a need for future research in contexts not addressed by the PBL Project.

What types of content understandings do teachers possess? In the analysis of teachers' responses, we noted that the ideas we were assessing represented different types or levels of scientific knowledge. As described by Windschitl (2009), some of the Big Ideas we expect teachers to know are descriptive in their nature, while others are explanations for the patterns and events observed in nature. Both of these types of concepts are represented among the Big Ideas addressed in the assessment instruments across all the content strands in the PBL Project for Teachers of Science. A question still to be explored is whether teachers have stronger prior content knowledge of one type of idea or the other, and which type of idea teachers are more successful in learning. For example, some teachers may be able to list the types of rocks (igneous, metamorphic and sedimentary), and even classify rocks into those categories. Questions that fit this category can be considered descriptive. Others may understand not only the list, but be able to explain how each type is formed. The same teachers may not be able to explain how the rocks were formed, or how the environment in which they were formed influences characteristics like crystal size. These ideas are explanatory in nature. Teachers should have a strong understanding of both types of information in order to design instruction that address all these Big Ideas. Understanding the patterns in the types of ideas teachers know can shape the design of PD activities. This will entail coding each Big Idea as either a descriptive or explanatory concept and comparing the coded responses to reveal any patterns. The assessment strategy and coding process described here facilitates this analysis because the information is already compiled. However, researchers need only add these descriptors of the Big Ideas to examine this question.

What are the patterns in learning across a larger sample of teachers? The cases presented above focus on individual teachers. The data reveal what each participant knew or learned. Other questions still to be explored will require analysis of patterns in teacher learning across a larger sample of teachers. The questions may focus on variables that might influence learning, such as teaching experience, educational background, incoming content knowledge and differences between each content strand.

The cohort from which the cases described above were drawn included 78 teachers ranging in experience from 1 to 43 years. Some participants like Mrs. Brinkman, held degrees in science. Others, like Mrs. Shuler and Mrs. Turkus, were certified in elementary education, with little or no science coursework, but were assigned to teach middle school science. With such variation among teachers, it may be useful to examine the correlation between these characteristics and the learning outcomes of the teachers. Do teachers with science backgrounds learn more effectively in a problem-based learning activity than non-science majors? Do teachers with less experience learn more quickly? Answers to these questions may help PD planners develop programs and strategies that have the greatest impact or that target teachers who need more support.

The variables above may also be related to differences in the incoming knowledge of participants. One question yet to be addressed is whether teachers with very little prior knowledge struggle to assimilate new concepts. It is also possible that participants who entered with a high level of knowledge may demonstrate a ceiling effect and not learn new concepts. Use of the methodology described here to answer these questions may yield insights that guide designers of PD programs to effectively meet the needs of particular groups of teachers.

Analysis of a larger sample of teachers may also reveal patterns that differ by content strand. The types of Big Ideas, the difficulty of concepts, and the number of teachers in each strand varied. There may also be patterns attributable to the types of learning activities or the strategies employed by different instructors. Examination of these variables may inform both the design and delivery of instruction aimed at changing teachers' understanding of science concepts.

Implications for Science Teacher Educators

Content knowledge is an important characteristic of effective science teachers (Ball, 1997; Ma, 1999; Thames & Ball, 2010; Traianou, 2006). Teacher educators who plan and implement professional development designed to enhance science content knowledge must be attentive to both incoming knowledge and changes in teachers' ideas. In order to understand what teachers' knowledge of science concepts, teacher educators need a tool that can reveal the depth of understanding that teachers possess. Concept inventories and other selected-response assessments may not reveal the degree to which teachers' knowledge is fragmented or interconnected to practical applications.

The strategy described here serves as a sample methodology for assessing content knowledge, especially when used with limited numbers of teachers. The strategy we have described here is flexible enough to be applied to many different science topics, and reveals multiple dimensions of content knowledge that are important to effective science teaching. It includes a systematic scheme for coding responses and compiling data in ways that allow analysis of prior knowledge, changes in knowledge, and trends within and across groups of teachers.

Another application of this assessment strategy is to study content learning among teachers. Professional development providers need to collect evidence of changes in teachers' ideas as part of the program evaluation process. While concept inventories can provide concise information, they lack the ability to reveal the details about teachers' understanding. The two types of assessment questions provide information about both the breadth and depth of teachers' ideas related to specific concepts.

An important purpose for using any kind of assessment of learning is to inform the design of instruction and documentation of learning. In the PBL Project, this goal has included a flexible design that uses pre-assessment data to customize instruction to the needs of the participating teachers. We cannot assume that teachers come to professional developments programs with homogeneous needs and prior content knowledge. Because of their varying needs, teacher educators need to be responsive to the audience they serve. The strategy we have presented offers a practical tool for this purpose.

References

Akerson, V. L. (2005). How do elementary teachers compensate for incomplete science content knowledge? Research in Science Education, 35, 245–268.

AAAS. (1993). Benchmarks for Science Literacy. New York: Oxford University Press.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluations of the conceptual inventory of natural selection. Journal of Research in Science Teaching, 39(10), 952–978.

Authors. (2008) Information removed for blind review.

Ball, D. (1997). What do students know? Facing challenges of distance, context, and desire in trying to hear children. In B. J. Biddle, et al. (Eds.), International Handbook of Teachers and Teaching, vol. II, (pp. 679–718). Dordrecht, Netherlands: Kluwer Press.

Brooks, J. G., & Brooks, M. G. (1993). In search of understanding: The case for constructivist classrooms. Alexandria, VA: Association for Supervision and Curriculum Development.

Chiapetta, E. L., & Koballa, T. R, Jr. (2010). Science instruction in the middle and secondary schools: Developing fundamental knowledge and skills for teaching. (7th Ed). Boston, MA: Allyn & Bacon.

Czerniak, C., & Chiarelott, L. (1990). Teacher education for effective science instruction - A social cognitive perspective. Journal of Teacher Education, 41(1), 49–58.

Da-Silva, C., Mellado, V., Ruiz, C., & Porlan, F. (2007). Evolution of the conceptions of a secondary education biology teacher: Longitudinal analysis using cognitive maps. Science Education, 91, 461–491.

Darling-Hammond, L., & Richardson, N. (2009). Teacher learning: What matters? Educational Leadership, 66(5), 46–53.

Goldhaber, D., & Brewer, D. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. Journal of Human Resources, 32(3), 505–523.

Harlen, W. (1997). The teaching of science in primary schools. London, David Fulton.

Hauslein, P. L., Good, R. G., & Cummins, C. L. (1992). Biology content cognitive structure: From science student to science teacher. Journal of Research in Science Teaching, 29, 939–964.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. American Educational Research Journal, 42(2), 371–406.

Howley, A., & Howley, C. B. (2005). High-quality teaching: Providing for rural teachers' professional development. The Rural Educator, 26(2), 1–5.

Jeanpierre, B., Oberhauser, K., & Freeman, C. (2005). Characteristics of professional development that effect change in secondary science teachers' classroom practices. Journal of Research in Science Teaching, 42(6), 668–690.

Ma, L. (1999). Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States. Mahwah, NJ: Lawrence Erlbaum Associates.

Michigan Department of Education. (2008). Michigan merit curriculum. Retrieved from http://www.michigan.gov/mde/0,1607,7-140-38924---,00.html

Mikeska, J. N., Koehler, M. J., Weizman, A., & Lundeberg, M. A. (2007, April). Designing teaching dilemmas for problem-based learning professional development. Roundtable discussion at the 2007 Annual Meeting of the American Educational Research Association, Chicago, IL.

Mundry, S. (2005). Changing perspectives in professional development. Science Educator, 14(1) 1–15.

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

Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. Journal of Research in Science Teaching, 45(10), 1131–1160.

Plummer, J. D., & Slagle, C. (2009, June). A learning progression approach to teacher professional development in astronomy. Paper presented at the Learning Progressions in Science Conference. Iowa City, IA.

Sanders, M. (1993). Erroneous ideas about respiration: The teacher factor. Journal of Research in Science Teaching, 30(8), 919–934.

Savinainen, A., & Scott, P. (2002). The force concept inventory: A tool for monitoring student learning. Physics Education, 37(1), 45–52.

Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2006). Classroom assessment for student learning: Doing it right – using it well. Portland, OR: Educational Testing Service.

Thames, M. H., & Ball, D. L. (2010). What mathematical knowledge does teaching require? Knowing mathematics in and for teaching. Teaching Children Mathematics, 17(4), 220–225.

Torp, L., & Sage, S. (2002). Problems as possibilities: Problem-based learning for K-16 education. (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Traianou, A. (2006). Teachers' adequacy of subject knowledge in primary science: Assessing constructivist approaches from a sociocultural perspective. International Journal of Science Education, 28(8), 827–842.

Tretter, T. R., Brown, S. L., Bush, W., Saderholm, J., & Moore, B. (2007, April). Valid and reliable physical, life, and earth science content assessments for middle school teachers. Poster presented at the National Association for Research in Science Teaching Annual Conference, New Orleans, LA.

Trundle, K. C., Atwood, R. A., & Christopher, J. E. (2006). Preservice elementary teachers' knowledge of observable moon phases and pattern of change in phases. Journal of Science Teacher Education, 17, 87-101.

Weizman, A., Covitt, B. A., Koehler, M. J., Lundeberg, M. A., Oslund, J. A., Low, M. A, Eberhardt, J., Urban-Lurain, M. (2008). Measuring teachers' learning from a problem-based learning approach to professional development in science education. Interdisciplinary Journal of Problem-based Learning, 2(2), 29–60.

Windschitl, M. (2009, February). Cultivating 21st century skills in science learners: How systems of teacher preparation and professional development will have to evolve. Paper presented at the National Academies of Science Workshop on 21st Century Skills, University of Washington.