Usable Assessments for Teaching Science Content and Inquiry Standards

Christopher J. Harris, Katherine L. McNeill, David J. Lizotte,

Ronald W. Marx, and Joseph Krajcik

School of Education

University of Michigan

Harris, C. J., McNeill, K. L., Lizotte, D. L., Marx, R. W. & Krajcik, J. (2003, in press). Usable assessments for teaching science content and inquiry standards. *Peers Matter*, 1 (1).

Abstract

This article reports on our approach to developing middle-school science assessments using an assessment-driven design model. We describe our design process for creating usable assessments that are aligned with curriculum and important science content and inquiry learning standards, then illustrate how one assessment tool, rubrics, can be used effectively by teachers and researchers during an instructional unit. Evidence from an enactment of a middle-school chemistry unit shows the initial success of our work as well as lessons learned from the real-world environment of an urban science classroom.

Usable Assessments for Teaching Science Content and Inquiry Standards

Assessment is a critical means for determining the extent to which students achieve learning goals. The closer student assessment is aligned with curriculum and classroom practice, the more likely assessment data will provide an accurate picture of learning (Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002). Large-scale, standardized assessments are often separated from real classroom learning (Shepard, 2000). Because these assessments are not aligned with classroom curriculum and instruction, they are less likely to be sensitive to changes in student learning. In contrast, assessments that closely align with an enacted curriculum and its learning goals may be more immediately usable by teachers and researchers for getting feedback about whether students are achieving goals and for adjusting curriculum and instruction accordingly.

In our work, we shift the focus of assessment toward the classroom, where the teaching and learning occurs. Assessments should be usable; they should be practical and informative for teachers and researchers. Yet, if not designed well, even classroom assessments may be inadequate. An assessment's effectiveness depends, in large part, on how well it aligns with curriculum and instruction to reinforce common learning goals (Pellegrino, Chudowsky, & Glaser, 2001). In this article, we describe our design process for creating usable assessments that are aligned with curriculum and important science content and inquiry learning standards. We then illustrate how one assessment tool— rubrics—can be used effectively by teachers and researchers during an instructional unit. The central goal of our research project was to narrow the gap between assessment, curriculum, and learning goals through the process of assessment-driven design.

Usable Assessment for Scientific Inquiry

Science education reform efforts call for students to develop scientific processes and skills through inquiry (American Association for the Advancement of Science [AAAS], 1993; National Research Council [NRC], 1996). The *National Science Education Standards* strongly emphasize developing students' inquiry abilities through which students should learn more than vocabulary definitions and content knowledge (NRC, 1996). Content knowledge is certainly valuable, but in reform-oriented, inquiry-based science classrooms, students are also expected to participate in scientific practices and apply scientific ideas through describing and explaining phenomena, carrying out experiments and investigations, and collecting and analyzing data (Krajcik, Czerniak, & Berger, 2002). In these science classrooms, knowing science does not simply mean that students remember facts, but includes many higher cognitive processes.

The kind of learning that arises from engaging in scientific inquiry is very different from the learning assessed with conventional testing methods. While many forms of instruction appear to have the same effect on student learning when the only measure is factual recall, instructional differences become visible when multiple measures including higher cognitive ones are used (Bransford, Brown, & Cocking, 2000). Using only multiple-choice assessment does not access the deep, rich understandings called for in reform. Assessment should measure what students learn and this learning should parallel the curriculum and the standards. Although alignment can be difficult to achieve, classroom assessments aligned with learning objectives and the curriculum hold promise for providing meaningful, informative, and useful feedback for teachers and researchers.

Design Process

Assessment-driven Design

Science teachers constantly design and redesign their lessons and assessments to match changing student populations. Science education researchers design instructional materials, such as curriculum and assessments, to meet the needs of changing school systems, administrators, teachers, and students. Both science teachers and researchers share the common goal of all designers to develop courses of action aimed at modifying existing situations into preferred ones (Simon, 1996); in this case, increased student learning. But what do we want students to learn and how do we assess this learning? The national standards (AAAS, 1993; NRC, 1996) provide a framework for what students should know and be able to do in science. They define what scientifically literate students should know at that end of grades 5, 8, and 12. While the national standards supply the desired learning outcomes, we still need to develop instructional materials to help students achieve these goals.

Our research group has been developing middle school science instructional materials based on national standards using an assessment-driven design model (Reiser, Krajcik, Moje, & Marx, 2003), similar to Wiggins and McTighe's (1998) backward design model. Often science teachers and researchers, with a general notion of what they want students to learn, will start with favorite activities or a textbook when they are designing curriculum. Instead, we began with our desired learning outcomes. Our commitment is that assessment-driven design can create better alignment between curriculum and assessment. We also think that this alignment can help us uncover more fine tuned changes in student learning.

Using Assessment-driven Design to Create an Instructional Unit

We used assessment-driven design to create a middle school inquiry-oriented chemistry unit. Our assessment-driven design process involved six steps. We: 1) identified and clarified national standards, 2) developed learning performances to meet standards, 3) created base and specific assessment rubrics linked to the learning performances, 4) identified learning tasks and assessment tasks, 5) produced a contextualized instructional sequence including both student and teacher materials, and 6) pilot tested materials and received feedback from external reviewers. Although these steps are listed linearly, in practice they were iterative. We found that the later steps of the design cycle, such as the assessments, informed previous steps, such as the learning performances. Figure 1 illustrates the iterative nature of the process.

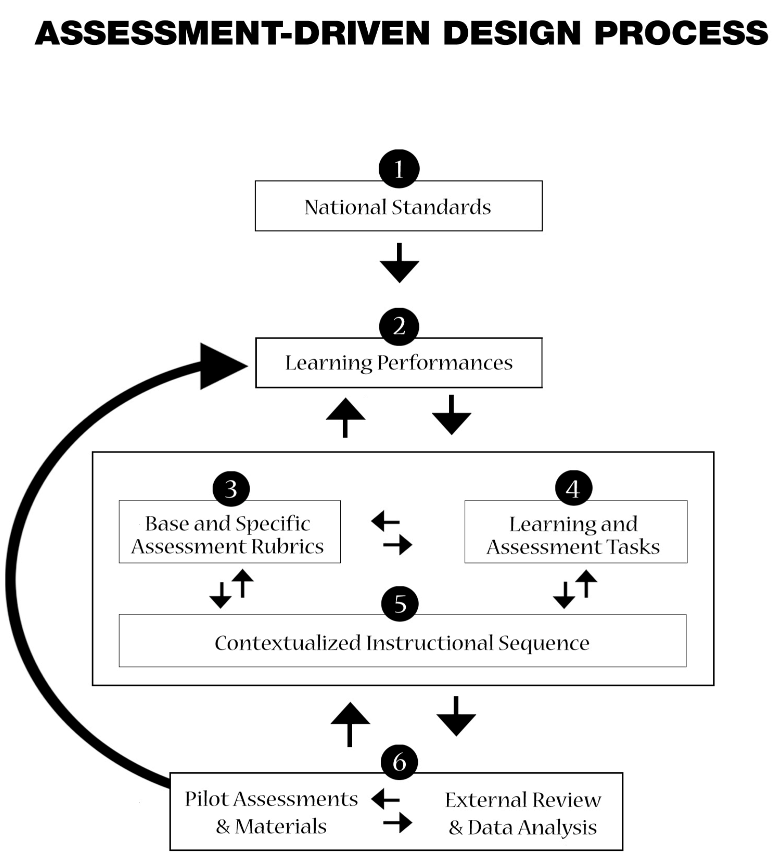

Figure 1: Assessment-Driven Design Process

As a first step in our design process, we began with national standards to identify key middle school chemistry ideas. Once we identified the relevant content standards, we unpacked these relatively succinct statements to clarify the science behind them. For the first four-week segment of the unit, we focused on the concepts of substance, property, and chemical reaction. These concepts are articulated in a national content standard about substances and properties (NRC, 1996, p. 154) and a national content standard about chemical reactions (AAAS, 1990, p.47) (see Table 1).

Table 1

From National Standard to Learning Performance

Standard	Clarifying the Standard	Learning Performance
A substance has characteristic properties, such as density, a boiling point, and solubility, all of which are independent of the amount of the sample (NRC, 1996, p. 154).	A substance is made of one material throughout. Substances have distinct properties that can be used to distinguish and separate one substance from another. Properties such as density, melting point, and solubility describe the unique characteristics of substances. The properties of a substance do not change regardless of the amount of the substance. Density is the ratio of mass per unit volume. Melting point is the temperature at which a solid changes to a liquid. Solubility is the capacity of a solid to dissolve in a liquid.	LP1 – Students identify substances and describe substances as being made of the same material throughout.
		LP2 – Students identify properties and describe that properties are unique characteristics that help identify a substance and distinguish one substance from another. These properties do not change regardless of the amount of the substance.
		LP3 – Students design an investigation to determine whether two objects are the same substance. They formulate questions or predictions, identify variables, control variables, and communicate scientific procedures.
		LP4 – Students conduct a scientific investigation to gather data about properties of substances, such as color, hardness, density, melting point, and solubility.
		LP5 – Students analyze and interpret data about properties to identify substances and distinguish one substance from another.
		LP 6 – Students construct scientific explanations stating a claim whether two items are the same substance or different substances, evidence in the form of properties, and reasoning that different substances have different properties.
When substances interact to form new substances, the elements composing them combine in new ways. In such recombinations, the properties of the new combinations may be very different from those of the old (AAAS, 1990, p. 47).	Substances have distinct properties and are made of one material throughout. A chemical reaction is a process where new substances are made from old substances. One type of chemical reaction is when two substances are mixed together and they interact to form new substance(s). The properties of the new substance(s) are different from the old substance(s). When scientists talk about "old" substances that interact in the chemical reaction, they call them reactants. When scientists talk about new substances that are produced by the chemical reaction, they call them products.	LP7 – Students identify processes (chemical reaction, phase change, mixing) and describe that a chemical reaction is a process in which old substances interact to form new substances with different properties than the old substances.
		LP8 – Students design an investigation to determine what combination of substances causes a chemical reaction. They formulate questions or predictions, identify variables, control variables, and communicate scientific procedures.
		LP9 – Students conduct a scientific investigation to gather data about properties before and after a process (chemical reaction, phase change, mixture).
		LP10 – Students analyze and interpret data for properties before and after a process to identify what type of process occurred (chemical reaction, phase change, mixing).
		LP11 – Students construct scientific explanations stating a claim for whether a chemical reaction occurred, evidence in the form of properties, and reasoning that a chemical reaction is a process in which old substances interact to form new substances with different properties than the old substances.

The focus on scientific inquiry in recent standards-based documents and science education research suggests that knowing science involves substantially more than recalling scientific facts. In order to articulate what we mean for a student to know the selected standards, we developed a range of *learning performances* that require different cognition from students (Table 1). This was the second step in our assessment-driven design process. Building on Perkins' notion of understanding performances (Perkins 1998), we moved from the standards, which are a description of the scientific ideas, to performances that represent the understanding and inquiry learning of those ideas. Our learning performances delineate multiple ways that students can demonstrate knowing the science content in a standard, derived from the recently revised Bloom's Taxonomy (Anderson & Krathwohl, 2001) as well as the scientific inquiry called for by reform initiatives (AAAS, 1993; NRC, 1996). Each learning performance addresses part of a single standard. A set of learning performances together addresses an entire standard. The learning performances focus on such ways of knowing as description, explanation, experimental design, and analysis of data.

The ways of knowing in the learning performances align with our assessment model via a set of *base* rubrics designed in step 3 of our process. For this step, we developed a set of base rubrics that correspond to the different cognitive processes articulated in our learning performances (description, explanation, experimental design, analysis of data). A base rubric articulates the different components of a particular way of knowing and the levels of those components. These base rubrics can be adapted to any science content and thus can be used across all science curricula. Table 2 presents our base rubric for explanation. The base explanation rubric can be used to evaluate students' explanations whether they are learning chemistry, biology, or physics. The base rubrics encourage greater alignment across units by emphasizing scientific practices that are consistently evaluated with the same criteria. Although we focus on our chemistry unit in this paper, we are also creating an ecology and evolution unit for middle school students with colleagues at Northwestern University. By focusing on the same "ways of knowing" and using the same base rubrics, we are encouraging greater alignment across the units.

Table 2

Base Explanation Rubric

Component	Level		
	1	2	3
Claim (An assertion or conclusion for a problem.)	Does not make a claim, or makes an inaccurate claim.	Makes an accurate but incomplete claim.	Makes an accurate and complete claim.
Evidence (Data that supports the claim.)	Does not provide evidence, or only provides evidence that does not support the claim.	Provides accurate but insufficient evidence to support the claim. May include some evidence that does not support the claim.	Provides accurate and sufficient evidence to support the claim.
Reasoning (An argument that links evidence to the claim.)	Does not provide reasoning, or only provides reasoning that does not link evidence to the claim.	Provides accurate but incomplete reasoning that links evidence to the claim. May include some reasoning that does not link evidence to the claim.	Provides accurate and complete reasoning that links evidence to the claim.

We used our base rubrics to develop *specific* rubrics for assessing students on each learning performance for our chemistry unit, also consistent with step 3 of our assessment-driven design process. A specific rubric has the same components and levels as a base rubric, but is tailored to a given learning performance. Because specific rubrics directly align with learning performances, they can only be used for a certain instructional unit and grade level. Table 3 presents a specific rubric for assessing students' explanations for learning performance 11 (see LP11 in Table 1). We then used the specific rubrics as a tool to structure assessment and learning tasks for the chemistry unit in step 4 of our design process. In this way, students could be evaluated consistently across the unit using the criteria of the specific rubrics, thus facilitating alignment with important science content and inquiry standards. The tasks provide multiple ways for students to engage in the science content and demonstrate knowing.

As a way of contextualizing the content in real world student experiences, we organized the learning tasks into a series of lessons that are linked together by a driving question. This was the fifth step in our design process. A driving question is a rich and open-ended question that uses everyday language to connect with students' authentic interests and curiosities about the world (Krajcik, Berger, & Czerniak, 2002). The driving question is carefully crafted as the central organizing feature that drives students' investigations. The driving question of the chemistry unit, "How can I make new stuff from old stuff?" addresses how new substances can be made from old substances. Specifically, students investigate how soap can be made from lard and sodium hydroxide. During the unit, students complete a number of investigations, each time cycling back to the driving question. The investigations allow them to experience scientific phenomena and processes by describing observations, designing and conducting experiments, gathering and analyzing data, and explaining scientific ideas that are instrumental to understanding important science content. Each cycle helps students delve deeper into the science content to initially understand substances, then properties, and finally substances interacting to form new substances (i.e. chemical reactions).

Explanation as a Way of Knowing

To illustrate how we employed our rubrics, we focus on one central way of knowing, *explanation*. Explanation is both a process in scientific inquiry and an important scientific practice, emphasized in the *National Science Education Standards* (NRC, 1996). A significant body of research treats explanation as a process of coordinating evidence and theory, and investigates how the use of explanation can provide students with opportunities to develop competency in this scientific practice (Driver, Newton, & Osborne, 2000; Kuhn, 1993; Sandoval, 2003; Wu & Krajcik, 2003). A number of science education researchers have examined students' explanations for the insight they provide into students' understanding of concepts, such as gears (Metz, 1991), natural selection (Sandoval, 2003), and light (Bell & Linn, 2000). These researchers share the view that explanation is more than a simple index of content knowledge. In accord with the research literature and the goals of the *National Science Education Standards*, we focus on explanation as both a process and a scientific practice.

The base rubric for explanation shown in Table 2 entails three components: a claim about a problem, evidence for the claim, and reasoning that links the evidence to the claim. The rubric defines a range of levels for completing each component of an explanation task and can be applied to learning performances regardless of content. Here we apply the base explanation rubric to learning performance 11 in Table 1. Table 3 presents the specific rubric for assessing students' explanations relevant to this learning performance. The three components of the explanation rubric allowed us to analyze separately a student's claim, evidence for the claim, and reasoning linking evidence and claim.

Using the Explanation Rubric in the Classroom

One of our goals was to create usable assessment tools for classrooms. Teachers can customize the base rubric for explanation to create a specific explanation rubric for any grade level or science content area. Teachers can chart the progress of their students' explanations by using the same rubric for assessment tasks throughout a unit or entire curriculum. They can adapt the base rubrics for the learning goals and student ability levels in their classrooms. In order to customize the rubrics, a teacher would need to go through a process of assessment-driven design similar to the one described previously in this paper. The teacher needs to determine the learning goals of the unit, then align the learning goals and the base rubrics to create specific rubrics.

Table 3

Specific Explanation Rubric for Learning Performance 11 (LP11)

Component	**Level**			
	1	2		3
Claim (An assertion or conclusion for a problem.)	No claim, or an inaccurate claim. Sample Response: "*Something happened after mixing.*"	[Does not apply.]		An accurate and complete claim. Sample Response: "*A new substance was formed after mixing substances together.*"
Evidence (Data that supports the claim.)	Does not provide evidence, or provides evidence that does not support the claim. Sample Response: "*When the substances were mixed together, I saw that they changed.*"	(a) Provides 1 piece of accurate evidence. May include some evidence that does not support the claim. Sample Response: "*When the substances were mixed together, I noticed that it made more stuff and changed color.*"	(b) Provides 2 pieces of accurate evidence. May include some evidence that does not support the claim. Sample Response: "*When the substances were mixed together, I saw that a gas formed and that it changed color.*"	Provides 3 of the following pieces of accurate evidence, if applicable: • changes color • changes hardness • changes odor • changes melting point • changes solubility properties • different density • different pH • produces a gas ("*bubbles*"; "*fizzes*") • produces a solid or precipitate • produces heat Sample Response: "*When the substances were mixed together, I noticed that the mixture bubbled, changed color, and became hot.*"
Reasoning (An argument that links evidence to the claim.)	Does not provide reasoning, or provides reasoning that does not link evidence to the claim. Sample Response: "*Because my evidence shows that a new substance was formed.*"	Reasoning includes: • Evidence shows that there are new properties after mixing. OR • Evidence shows that a new substance was formed after mixing. Sample Response: "*A new substance was formed because my evidence shows that the substances before mixing are different from the substances after mixing.*"		Reasoning includes: • Evidence shows that there are new properties after mixing. • These new properties show that a new substance was formed. Sample Response: "*A new substance was formed because my evidence shows that the properties after mixing the substances were different from before mixing them.*"

For example, our unit requires students to explain whether or not a chemical reaction occurred after different substances were mixed together (learning performance 11 in Table 1). The specific explanation rubric in Table 3 breaks down the explanation into the three components and the levels of response for each component. First students make a *claim*: an assertion about what they think is happening. After observing a chemical reaction, a student might claim, “A new substance was formed after mixing substances together.” This claim is accurate and would be rated a level 3 response (see Table 3). Next, students provide *evidence*: scientific data to support their claim. For example, a student could write, “This new substance has a different density, color, and melting point than the substances I started with.” Because this student provides three accurate pieces of evidence, the response would be rated a level 3 accordingly. Note that level 2 for the evidence component is divided into level 2a and 2b, allowing the teacher or researcher to differentiate between students’ responses that provide only one or two pieces of evidence. Splitting level 2 into 2a and 2b illustrates further how a specific explanation rubric may be adapted for a particular learning performance. Finally, students provide their *reasoning*: an argument articulating why the evidence supports their claim. For example, a student could write, “Because there are new properties (density, color, and melting point), I know there is a new substance. Different substances have different properties.” This reasoning would be rated a level 3. Separating a student’s explanation into these three components and considering the level of response for each component can provide greater insight into the evaluation of student understanding.

To illustrate how the specific explanation rubric can be used, we now describe how a middle school teacher used it and the issues that arose in the real-world environment of her urban science classroom.

The Teacher and Her Students

In spring 2002, Katheryne Frank enacted the chemistry unit with her seventh grade science class. Katheryne is an experienced science teacher at a public middle school in a large urban school district in the Midwest. Thirty-two students were enrolled in her class, although the attendance ranged from 19 to 27 students on any one day. The school, which is typical compared to other middle schools in the district, is comprised of approximately 470 students, primarily African-Americans from lower to lower-middle income families.

In order to provide a picture of how students' explanations changed, we focus on three students: Asha, Elena, and Bethany.[1] We discuss data from these students' explanations in order to demonstrate how the explanation rubric can be used for assessment. The three students reflected different levels of ability in writing explanations at the beginning of the unit and each showed unique growth. On occasion, we will also refer to examples from other students.

Students' Explanations on the Pretest

Using a rubric early in a unit can help teachers and researchers plan instruction to meet the particular needs of students. To illustrate this possibility, we applied the specific explanation rubric in Table 3 to students' explanations of a chemical reaction prior to starting the chemistry unit.

The students completed a pretest that included open-ended problems requiring explanation. The following problem dealt with a chemical reaction:

You have a clear liquid, a white powder, and a red powder.

1. Design an experiment to find out which two substances when mixed together will produce a chemical reaction.
2. Describe three pieces of evidence you would look for to determine if a chemical reaction occurred.
3. Why does this evidence support that a chemical reaction occurred?

---

[1] Names of students are pseudonyms.

We focus on students' responses to the latter two steps of the problem, because they correspond to the evidence and reasoning components of our explanation rubric. The claim was given: a chemical reaction occurred. The second step in responding to the problem required students to provide evidence for the claim and the third step required students to provide reasoning linking that evidence to the claim. Table 4 presents the evidence and reasoning of our 3 target students for this problem on the pretest, and for the same problem on the posttest.

Table 4

Pretest and Posttest Written Responses of Elena, Asha, and Bethany

Open-Ended Item: Design an experiment to find out which two substances when mixed together will produce a chemical reaction. Describe three pieces of evidence you would look for to determine if a chemical reaction occurred. Why does this evidence support that a chemical reaction occurred?				
Component	**Test**	**Elena**	**Asha**	**Bethany**
Evidence	Pretest	No response. (Level 1)	"The substances will have reacted, you will see a change, and you will just compare" (Level 1)	"A color change bubbling." (Level 2b)
	Posttest	"Change in color, or bubbles, powder dissolving." (Level 2b)	"The color, the hardness, and the density is evidence I would look for. Because you could look for the properties and see if they changed" (Level 3)	"bubbleing," "color change," "Density" (Level 2b)
Reasoning	Pretest	No response. (Level 1)	"This evidence supports that a chemical reaction occurred because you can follow the evidence and determine that it change." (Level 1)	"It will because it change. of formed" (Level 1)
	Posttest	"Because it is A it is A chemical change because new things happen." (Level 1)	"Because you could look for the properties and see if they changed" (Level 1)	"because when you mix two things they Just don't bubble it have to be something th- [illegible]" (Level 1)

Asha's explanation was representative of her classmates on the pretest. Her response did not include enough detail to count as evidence according to the specific explanation rubric. She wrote in general terms ("you will see a change") rather than citing particular pieces of evidence of a chemical reaction, such as a change in color or the production of a gas. Asha's evidence received a Level 1 rating.[2] Although Asha referred to both her evidence and the claim when asked to provide reasoning, she did not provide an argument articulating *why* her evidence supported the claim. Her attempt at reasoning was rated a Level 1.

Bethany provided a somewhat different explanation than did Asha by including two pieces of plausible evidence of a chemical reaction. We rated her evidence as Level 2b according to the rubric. Bethany was one of only a few students who provided appropriate evidence on the pretest. On the other hand, Bethany, like Asha, was unable to articulate a reason why her evidence supported the claim. Based on her response, it is not clear what she thought her evidence accomplished. In fact, reasoning was an impediment for all students on the pretest. Other examples of students' reasoning for this question included: "Cause that is what happen when a chemical reaction occurs." "It would no longer be a liquid." and "By just reacting the substances they had being mixed together." Each of these attempts received a Level 1 rating.

Elena did not answer the open-ended problem about a chemical reaction on her pretest, although she did complete a number of other questions. To characterize her explanations at the beginning of the unit, we examine her first opportunity to provide an explanation in the unit proper. In the first lesson, students recorded descriptions of two "unknown" materials. The students wrote a claim stating whether the two unknowns were the same or different and provided evidence for their claim. The task thus maps onto the claim and evidence components of the base explanation rubric. Elena claimed that the two unknown materials "are the same," a

---

[2] Ratings of levels in this section refer to those in Table 3.

perfectly reasonable claim, given that the materials were unknown. Yet providing appropriate evidence for a claim was a challenge; her evidence was that they "both leave grease spots," even though she recorded that only a *single* unknown was "greasy" in her descriptions. To support her claim that they "are the same," Elena needed to select appropriate evidence such as two similar characteristics of the unknown substances. She did not yet understand what counts as appropriate evidence.

In general, Ms. Frank's students successfully made claims at the beginning of the unit. However, the explanation rubric revealed that most students had difficulty providing evidence to support their claims, due to either inappropriateness, as exemplified by Elena, or a lack of detail, as exemplified by Asha. Most students' evidence was rated as Level 1, with students such as Bethany being the rare exception. The rubric also showed that all of Ms. Frank's students had difficulty providing reasoning linking their evidence and claims. Their early attempts at reasoning were generally poorly articulated and scored as Level 1.

Addressing Student Difficulties through Classroom Discussions about Evidence

On the fourth day of the chemistry unit, Ms. Frank used the term *evidence* for her bell work—what she called the "science thought" of the day. Ms. Frank selected this term from the unit materials and placed emphasis on it because she was concerned that her students would not understand how to use evidence in their explanations. The students had to create a diagram for the word *evidence* with four components: define what is meant by evidence, describe what evidence is used for, give examples of evidence and provide nonexamples of evidence (Ms. Frank's term for a counter example). Ms. Frank led a discussion about their diagrams of *evidence*.

Ms. Frank: Evidence. What does it mean?

Student: It shows what is true and what is not.

Ms. Frank: How have we used evidence this week?

Student: To observe things.

…..

Ms. Frank: You use evidence, ok. Not just in a courtroom, you use evidence when you want to say something… 'How do you know' is your evidence… What is a nonexample?

Student: An opinion.

Ms. Frank: That is a pretty good nonexample. Some people might think that just because they believe something that that's evidence. But the evidence is why you believe something. You've got to have reasons for believing something. That's your evidence.

On the next school day, Ms. Frank approached one of the researchers before class. Ms. Frank explained that while reviewing students' work from previous lessons, she noticed that her students seemed to be having a difficult time describing observations with appropriate evidence. For example, during a lesson in which students were given "unknown" materials and asked to carefully observe and write descriptions with appropriate evidence, they often overlooked the physical evidence in front of them. Ms frank commented that students were naming the materials (e.g. "looks like cotton candy") and giving their opinions (e.g. "it could be used for cooking chicken") instead of recording observable properties such as color and hardness.

In order to address the students' confusion, Ms. Frank gave her students the vocabulary word *observe* for their science thought that day. After she led a discussion of the students' diagrams for their science thought, she addressed a number of the students' difficulties. She told her students that she had read their descriptions of the unknown materials and that some of what they wrote were "not really observations."

Ms. Frank: Some of you wrote, 'It's soap'. Now if you wrote, 'It smells like soap,' that was ok because that's an observation based on prior knowledge. But if you told me,

> 'It was soap' that is not an observation. That is an opinion because I haven't told you what it was. You just looked at it.

By discussing *evidence* and *observe*, Ms. Frank helped her students understand what they needed to include in explanations. In order to help students create better explanations, they needed guidelines to help them in their thinking. One technique Ms. Frank used was to remind students to "CQAA" – Combine Question And Answer. This technique encouraged her students to think about what the question was asking and to include the question in their response. CQAA correlates with the concept of *claim* in the explanation rubric. Ms. Frank's use of CQAA is one example of how a teacher can customize the explanation rubric to match with one's own instructional strategies and the unique needs of students.

Because this technique was a part of the classroom culture prior to the start of the unit, it may be one reason why students consistently wrote appropriate claims even on the pretest. Ms. Frank had already set that as a precedent. On the other hand, the use of evidence and reasoning in the students' explanations was new. It is not surprising that her students had difficulty with these explanation components at the beginning of the unit. As the students become accustomed to being asked for evidence and after Ms. Frank's discussion of evidence, we begin to see improvement in the evidence component of the students' explanations.

Students' Explanations at the End of the Instructional Unit

A rubric can provide detail about the progress of individual students. Comparing student work at different times (i.e. beginning, middle, and end of a unit) can also help identify trends in student thinking and help uncover common student difficulties. To illustrate some of the changes in students' explanations by the end of the unit, we applied the specific explanation rubric to student explanations from the last lesson and the posttest.

The culminating lesson. In the final lesson of the unit, the students conducted experiments in which they made soap from lard and sodium hydroxide. Afterwards, students wrote a claim about whether they thought a new substance was formed as a result of their experiments, provided evidence to support their claim, and articulated their reasoning. Table 5 shows Asha, Bethany and Elena’s written responses for each explanation component. All three students made correct claims that a new substance was formed; their responses were scored a level 3 accordingly. Additionally, all were able to provide some appropriate evidence. Their explanations varied most on the reasoning component.

Table 5

Written Responses of Elena, Asha, and Bethany in the Culminating Lesson

Q: Do you think a new substance was formed after mixing the fat, the rubbing alcohol and sodium hydroxide? Provide 3 pieces of evidence to support your answer. Explain why the evidence supports your answer.			
Component	**Elena**	**Asha**	**Bethany**
Claim	“it is a new substance…because I don’t see any salt, or alcohol, or sodium hydroxide” (Level 3)	“a new substance was formed after mixing the fat, the rubbing alcohol and sodium hydroxide a chemical change” (Level 3)	“Yes because when they were not combined they was Just stuff some liquid some not but when they were mixed they made A hard or [‘soft’ or ‘solid’ – unclear] type of soap” (Level 3)
Evidence	“1. it is a new substance. 2. it has new propertys. 3. fat / soap hardness: soft squishy / semi-hard solubility: water - no / water - yes oil – yes / oil - no color: off white / milky white” (Level 3)	“3 pieces of evidence to support your answer is color (The color went from off white to milky white), smell (It went from stinky to a no smell), and the hardness” (Level 2b)	“The sodium hydroxide was [incomplete]… The salt was clear. The rubbing alcohol was clear. After you combine they the became A hard whit piece of soap. The Color went form off white to milkey white. The Hardness went form soft squishy to seim [semi] hard.” (Level 2b)
Reasoning	“Because it shows what each substance has in it.” (Level 1)	“the evidence support my answer because you could do the experiment and see that the same changes will occur and you will get a new substance from the other substances” (Level 2)	“It support my Answer because you see that After combining everything you get Different Color, Hardness, Denisty, Solubility, and ph. These Are changes and properties whitch mak a different substance.” (Level 3)

Asha's evidence included two properties that changed (color and odor). While she also stated "hardness", there was no mention of change, so it did not count as appropriate evidence. For this reason, the evidence component of her explanation rated a level 2b. Her reasoning statement received a level 2 because she mentioned that her evidence supported a change in substances, but she did not establish that this was due to a change in properties.

Bethany stated "The Color went form off white to milkey white. The Hardness went form soft squishy to seim [semi] hard." Her statement summarized two changes in properties and was rated level 2b. Bethany's reasoning was not characteristic of her classmates because she received a level 3. She was one of the only students to provide an argument that linked her evidence (changes in properties) to her claim (make a different substance).

Elena's response included three appropriate properties (hardness, solubility, and color) so she received a level 3 on the evidence part of the rubric. However, she also included "it is a new substance" as a piece of evidence suggesting that she was unclear about what counted as evidence. Her reasoning statement received a level 1 because she did not include an argument linking the properties of the substances to her claim that a new substance was formed.

Students' explanations on the posttest. The students took the same posttest as pretest so we could track their progress. Table 4 presents the evidence and reasoning of the three target students for an open-ended problem on the pretest and posttest.

Asha's evidence improved considerably. Whereas her pretest response did not include enough detail to count as evidence and received a level 1 rating, her posttest response (rated as level 3) included three pieces of accurate evidence. However, her reasoning component remained at level 1 on the posttest.

Bethany's explanation for the posttest problem was only slightly different from her pretest response. The evidence component of her explanation included 2 pieces of accurate evidence (bubbling and color change). While she also stated "density", there is no mention of change, so it does not count as accurate evidence according to the specific rubric. For this reason, her evidence still reflected a level 2b. Bethany's reasoning on the posttest also did not improve compared to the pretest. In both cases, her reasoning received a Level 1.

While Elena did not respond to the open ended questions on the pretest, her evidence on the posttest was considerably stronger than her evidence in Lesson 1. Elena's posttest included two pieces of accurate evidence (change in color, bubbles), so her response was classified as Level 2b. She also mentioned "dissolving," which is not evidence for a chemical reaction according to the specific rubric. By using the rubric to analyze student explanations, we observed that a number of students incorrectly stated that dissolving or mixing is a chemical reaction. Elena's reasoning received a level 1 because she did not provide an argument for why her evidence supported her claim.

Changes in Students' Explanations

By the end of the unit, the three target students had improved in constructing explanations. Elena improved in selecting evidence to support her claims, but her reasoning statements remained low in quality on both the culminating lesson and posttest. The development of Elena's explanations was characteristic of the majority of the students' in the class. Asha improved the quality of the evidence for her claims by providing more detail. She showed some improvement in reasoning on the culminating lesson, but did not sustain this growth on the posttest. Bethany was already adept at noting evidence at the beginning of the unit and was the only student to give a high level reasoning statement on the culminating lesson. On the other hand, her reasoning statement on the posttest did not reflect this improvement.

Conclusion

We expected students to articulate their reasoning in their explanations by the end of the unit and at first we were surprised that they did not. However, when we reviewed our unit materials and looked closely at how they were enacted in the classroom, we realized that they did not communicate to the students or teacher what counts as reasoning or explicitly state expectations that students include reasoning. It is not surprising that students' responses scored low for reasoning. While the first five steps of our assessment-driven design process resulted in greater alignment between the curriculum and the assessments, the sixth step in our design process—the pilot testing of the materials in the real-world environment of a science classroom— informed us of the gap between our expectations and the actual unit. Similarly, the materials did not communicate to the students or teacher what counts as evidence. Yet, Ms. Frank specifically discussed with her students what was meant by scientific evidence and what counted as scientific evidence. When we used the specific explanation rubric to evaluate students' responses from both the last lesson and from the posttest, we found that students scored high in their use of evidence. Using this and other data on student strengths and weaknesses from our pilot test, we have revised both the assessments and the activities in another cycle of assessment-driven design (McNeill, Lizotte, Harris, Scott, Krajcik, & Marx, 2003).

Assessments that tackle multiple ways of knowing with many components yield valuable information about student learning. A wide range of knowledge and understanding can arise from instruction, but student competencies and difficulties may not be revealed by assessments that treat knowledge as unitary. We broke down "knowing" with our rubrics. By parsing explanation as a way of knowing into three components, Ms. Frank found that her students became skilled at two of the components (making claims and providing evidence) but remained challenged by the third (reasoning). Assessments that incorporate a range of knowledge are likely to be more informative for teachers by further clarifying what it means for students to "know" science.

Our work suggests that this kind of assessment could be useful for teachers and researchers to see and follow the evolution of student learning across grade level or science content. The base rubric and its derivative specific rubrics were used to compare students' explanations at the beginning and end of the unit. The same base rubric could be applied to students' explanations in a subsequent science unit or even a different grade.

A strength of base rubrics, aligned with curriculum and important learning standards, is that a teacher or researcher can customize them for classroom use and obtain evidence of student learning immediately. Measures of student learning that are more distant from the standards and from the daily learning experiences of students cannot provide such data. Standardized tests have become the gold standard and have become a high priority for political and policy decisions, but they are not as immediately helpful for judging students' understanding. Indeed, results of large-scale standardized exams are rarely reported in a timely enough manner for practical and informative classroom use. In contrast, base rubrics and other forms of usable assessments enable teachers to take an active role in the analysis of student work. Such assessments can help reveal instructional gaps and suggest courses of action for teachers, allowing them to shift instruction in response to the information gained and support learning as students are progressing through a unit of study. This, we believe, holds tremendous promise for narrowing the divide between assessment and instruction, and ensuring that students meet the educational standards called for in science education reform.

Acknowledgements

We gratefully acknowledge the collaboration of Katheryne Frank and her students. The research reported here was supported in part by the National Science Foundation (REC 0101780 and 0830310A605). All opinions expressed in this work are the authors' and do not necessarily represent either the funding agencies or the University of Michigan.

Authors

Christopher Harris is a doctoral student in education and psychology. A former elementary school teacher and curriculum designer, his research interests are in student learning, assessment, and curriculum design. Katherine McNeill, a former middle school science teacher, is a doctoral student in science education. Her research interests include student learning and curriculum design specifically for scientific inquiry skills such as constructing explanations. David Lizotte is a doctoral student in psychology. Among his research interests is the cognition underlying students' scientific reasoning. Ronald Marx, Ph.D., is a Professor of Educational Psychology. His research and development activities focus on how to help teachers engage their students in thoughtful learning of science. Joseph Krajcik, Ph.D., a Professor of Science Education, focuses his research on designing science classrooms so that learners engage in finding solutions to meaningful, real world questions through collaboration and the use of learning technologies. The authors are at the University of Michigan's Center for Highly Interactive Computing in Education.

References

American Association for the Advancement of Science. (1990). *Science for all Americans: A project 2061 report on literacy goals in science, mathematics, and technology*. New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Longman.

Bell, P. & Linn, M. (2000). Scientific Arguments as Learning Artifacts: Designing for Learning from the Web with KIE. *International Journal of Science Education,* 22 (8), 797-817.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academy Press.

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

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77 (3),* 319-337.

Krajcik, J., Berger, C. F., & Czerniak, C. M. (2002). *Teaching science in elementary and middle school classrooms: A project-based approach* (2nd ed.). New York: McGraw Hill.

McNeill, K. L., Lizotte, D. J, Harris, C. J., Scott, L. A., Krajcik, J., & Marx, R. W. (2003, March). *Using backward design to create standards-based middle-school inquiry-oriented chemistry curriculum and assessment materials*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, PA.

Metz, K. (1991). Development of Explanation: Incremental and fundamental change in children's physics knowledge. *Journal of Research in Science Teaching*, 28 (9), 785-797.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Perkins, D. (1998). What is understanding? In M. S. Wiske (Ed.), *Teaching for understanding: Linking research with practice*. San Francisco, CA: Jossey-Bass Publishers.

Reiser, B. J., Krajcik, J., Moje, E., & Marx, R. (2003, March). *Design strategies for developing science instructional materials*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, PA.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, 39 (5), 369–393.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1), 5-51.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29 (7)*, 4-14.

Simon, H. A. (1996). *The Sciences of the Artificial* (3rd ed.). Cambridge, MA: MIT Press.

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

Wu, H.-K., & Krajcik, J. S. (2003, March). *Inscriptional practices in inquiry-based classrooms: How do seventh graders construct and interpret data tables and graphs?* Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Philadelphia, PA.

Resources

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousands Oaks, CA: Corwin Press. A practical guide to effective assessment for improved student learning.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academy Press. This book examines recent research findings about thinking and learning and considers implications for educational practice.

Brown, J. H., & Shavelson, R. J. (1996). *Assessing hands-on science: A teacher's guide to performance assessment*. Thousands Oaks, CA: Corwin Press. An assessment handbook for science teachers who use hands-on science curricula. Describes a variety of science performance assessments.

Enger, S. K., & Yager, R. E. (2001). *Assessing student understanding in science: A standards-based handbook*. Thousands Oaks, CA: Corwin Press. A comprehensive and practical handbook for assessing science learning linked to National Science Education Standards.

Investigating and Questioning our World through Science and Technology (IQWST) Project Web site: http://www-personal.umich.edu/~krajcik/IQWSTSite/Pages/Home.htm. IQWST is a collaborative effort between curriculum researchers at the University of Michigan and Northwestern University to develop middle school curriculum and assessment materials that support students in learning important science content based on national standards.

Krajcik, J., Berger, C. F., & Czerniak, C. M. (2002). *Teaching science in elementary and middle school classrooms: A project-based approach* (2nd ed.). New York: McGraw Hill. A thorough introduction to the theory and practice of project-based science. This book is keyed to National Science Education Standards and includes a chapter on assessing in project-based science classrooms.

Minstrell, J., & van Zee, E. H. (Eds.). (2000). *Inquiring into inquiry learning and teaching in science*. Washington, DC: American Association for the Advancement of Science. This book explores what is meant by inquiry learning and teaching, provides examples of inquiry practice in science classrooms, and addresses issues that arise in inquiry instruction.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press. An in-depth look at how new knowledge about student learning can inform the development of new kinds of assessments. This book presents a research-based approach to assessment of student learning, suggests principles for designing new kinds of assessments, and considers implications for education policy, practice, and research.

Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd edition). Columbus, OH: Merrill Prentice Hall. This comprehensive book for teachers illustrates how to create classroom assessments for documenting and improving student learning.

Wiggins, G., & McTighe, J. (1998). *Understanding by design.* Alexandria, VA: Association for Supervision and Curriculum Development. A book intended for teachers, curriculum developers, and administrators that offers a practical framework for designing curriculum, assessment and instruction.

Methods

Research Question

In what ways does our model of base and specific rubrics, aligned with curriculum and national standards, enable assessment of student learning throughout instruction?

Participants

The setting for our research was a public middle school located in a large urban school district in the Midwest. Our data collection focused on an experienced science teacher and her seventh grade science class. Thirty-two students were enrolled in the class, although attendance ranged from 19 to 27 students on any one day. The school, which is typical compared to other middle schools in the district, is comprised of approximately 470 students, primarily African-Americans from lower to lower-middle income families. Nearly all students qualify for free or reduced-price lunch.

Curriculum

We used the process of assessment-driven design to create an eight-week inquiry-oriented chemistry unit as a route to in-depth understanding of scientific concepts and processes specified in science education standards (AAAS, 1993; NRC, 1996). The chemistry unit focuses on increasing student understanding of substances, properties, and chemical reactions. Prior to the unit, the curriculum focused on scientific process skills, the earth's atmosphere, and the composition of matter including the states of matter and the particulate nature of matter.

Procedure

We piloted a 4-week segment of the unit in Spring 2002. An assessment was administered before and after the unit (pretest-posttest design). The pretest and posttest were the same and consisted of 24 multiple-choice and open-ended items. Open-ended items were assessed using specific rubrics. An example of an open-ended item is the following:

You have a clear liquid, a white powder, and a red powder.

1. *Design an experiment to find out which two substances when mixed together will produce a chemical reaction.*
2. *Describe three pieces of evidence you would look for to determine if a chemical reaction occurred.*
3. *Why does this evidence support that a chemical reaction occurred?*

In addition to pretest-posttest measures, students' written assignments during the unit were collected as artifacts and assessed using specific rubrics. Artifacts included work from students' science folders, such as written explanations based on observations of scientific phenomena and claims regarding experiments conducted by students.

Observers were present in the classroom daily during the unit enactment, taking field notes and videotaping class sessions.

Analysis Techniques

We developed specific rubrics for assessing students' written explanations on pre- and posttest open-ended items and student artifacts from the unit. A specific rubric defines a range of levels from 1 (low) to 3 (high) for completing each component of an explanation. Three raters independently assigned level ratings on pretest and posttest open-ended items for all students. The raters agreed on 91% of level ratings; disagreements were resolved through discussion. For the written explanations collected as artifacts during the unit and reported on in this paper, raters assigned level ratings through discussion and then agreement by consensus.

National Science Education Standards Link

Our paper reports our assessment-driven design process that takes national standards as the starting point for the alignment of assessment, curriculum materials, and instruction for middle school inquiry science.

The *National Science Education Standards* provide assessment standards that emphasize change in assessment practice (NRC, 1996, p. 100) consistent with the goals of our work. The following inquiry and physical science content standards were the focus of our alignment process:

Science as Inquiry

Content Standard A 5-8:

As a result of activities in grades 5-8, all students should develop

• Abilities necessary to do scientific inquiry

• Understandings about scientific inquiry

Excerpts from *Science as Inquiry*:

"Students in grades 5-8 can begin to recognize the relationship between explanation and evidence" (p. 143).

"The language and practices evident in the classroom are an important element of doing inquiries. Students need opportunities to present their abilities and understanding and to use the knowledge and language of science to communicate scientific explanations and ideas… These should be presented in a way that allows students to receive constructive feedback on the quality of thought and expression and the accuracy of scientific explanations" (p. 144).

Physical Science

Content Standard B 5-8:

As a result of activities in grades 5-8, all students should develop an understanding of

• Properties and changes of properties in matter

Content Standard B 5-8, 1A:

A substance has characteristic properties, such as density, a boiling point, and solubility, all of which are independent of the amount of the sample.

Excerpt from *Physical Science*:

"In grades 5-8, the focus on student understanding shifts from properties of objects and materials to the characteristic properties of the substances from which the materials are made. In the K-4 years, students learned that objects and materials can be sorted and ordered in terms of their properties. During that process, they learned that some properties such as size, weight, and shape, can be assigned only to the object while other properties, such as color, texture, and hardness, describe the materials from which objects are made. In grades 5-8, students observe and measure characteristic properties, such as boiling points, melting points, solubility, and simple chemical changes of pure substances and use those properties to distinguish and separate one substance from another" (p. 149).