Design principles for scaffolding inquiry-based science assessment

Gabrielle Matese, Miriam Gamoran Sherin, and Daniel C. Edelson
Department of Learning Sciences, Northwestern University, 2115 North Campus Drive, Evanston, IL 60208-2610
Tel: 847-467-3808, Fax: 847-491-8999
Email: matese@northwestern.edu, msherin@northwestern.edu, d-edelson@northwestern.edu

Abstract: In this paper we examine the need for and design of professional development that supports teachers' efforts to use reform-minded assessment practices in inquiry-based science classrooms. We begin by describing our motivation for this work and what these reform-minded assessment practices might look like. We then detail some of the challenges teachers face in implementing these practices, based on both the research literature and our empirical work. To conclude, we outline two design principles for professional development that we believe will help teachers to productively address these challenges.

The need for reform-minded assessment

National efforts to reform science education call for the development of science inquiry skills (National Research Council [NRC], 1996). Toward this end, much effort by teachers and researchers has focused on what students should learn and how instruction can support that learning. Less emphasis, however, has been placed on how teachers can then assess students' learning. In fact, some research has found that though teachers use inquiry-based instruction, their assessment practices are not aligned. This is problematic because inquiry skills, such as *designing a scientific investigation* or *organizing, analyzing, and interpreting data*, are not readily captured by traditional classroom assessments. Thus new science inquiry standards require not only new instructional practices but also new forms of assessment. This is particularly important given that new forms of assessment have the potential to lead to improved learning outcomes for students (Black & Wiliam, 1998a; Wilson & Sloane, 2000; White & Frederiksen, 1998).

What are these "new forms of assessment?" In contrast to traditional forms of classroom assessment, in which a quiz or test is given at the end of a unit or a lab report is graded and returned, assessment reformers promote classroom assessments that incorporate challenging tasks to elicit higher order thinking, and that look not just at outcomes but at developing learning processes (Shepard, 2000). Ideally, teachers make classroom expectations visible to students, and students take an active role in their own assessment. The assessment that reformers describe is ongoing and formative, providing feedback both to students about their learning and to teachers about their instruction. It takes the form of "embedded assessment," in which "opportunities to assess student progress and performance are integrated into the instructional materials and are virtually indistinguishable from the day-to-day classroom activities" (Wilson & Sloane, 2000, p. 182). In short, new forms of assessment call for teachers to pay close attention to the development of student ideas, and to tightly integrate assessment into instruction (Shepard, 2000; Wilson & Sloane, 2000). In this view of assessment, teachers must learn new strategies for collecting and interpreting evidence about what their students know, what their students are able to do, and what their students are learning.

New assessment requires professional development

As professional development designers, we recognize that the innovative forms of assessment described above make new demands of teachers' knowledge and skills. They call for teachers to develop strategies for collecting new types of information, to collect this assessment information on an ongoing basis, to interpret assessment information about students' cognitive processes, and to make use of that information to revise instructional goals and practices and provide meaningful feedback. However, teachers attempting changes in classroom assessment often bring with them incompatible beliefs, knowledge, and practices (Bliem & Davinroy, 1997; Borko, Mayfield, Marion, Flexer, & Cumbo, 1997). Stiggins (1991) cautions that many teachers lack the "assessment literacy" required for implementing authentic assessments. Teachers may fail to collect information on student thinking, to analyze or interpret that information, or to use the information to provide feedback or improve instruction (Black & Wiliam, 1998b; Duschl & Gitomer, 1997; Shepard, 2000).

Research Questions and Design

Our research agenda focuses on the question: How do we design professional development to support teachers' use of innovative assessment practices, specifically in an inquiry-based science context? This paper addresses that question in two parts. We first look at factors influencing teachers' assessment practices. In the Findings section, we ask the questions: *What are teachers' beliefs about assessment in inquiry-based science? How do those beliefs support or conflict with the assessment practices promoted by assessment reform?* We then discuss the implications of these findings for the design of professional development in the next section, Design Principles.

We begin with a description of the context of the research: the assessment approach promoted, and the professional development and classroom enactments in which teachers were engaged. We follow with a brief description of how we collected and analyzed our data.

The context

The Assessment Approach

Inquiry-based learning places new demands on teachers for assessing student progress. In response to this need, the Center for Learning Technologies in Urban Schools (LeTUS) created a professional development class to support teachers in the development of ways to assess students' inquiry learning. The class was designed to help teachers implement a specific approach to embedded performance assessment (Wilson & Sloane, 2000). The professional development promoted a view of assessment in which students' conceptual understanding and fluency with inquiry skills are tracked over time as they develop. In this approach, "progress variables" for the year (for example, "Designing and Conducting Investigations") are identified. Next, class activities are designed that provide assessment information about students' progress in this area. In addition, scoring guides are developed that reflect the expected learning trajectory of students toward those goals. Student performance in these areas is assessed and documented using the scoring guides in order to track, understand, and further support student development.

The professional development

The instructor for the course was a local teacher, who was assisted by two researchers from Northwestern University. Participants in the professional development class were recruited by staff of the Center for LeTUS, and received incentives for taking part, including graduate credit and funds to use towards the purchase of technology for their classrooms. Participation was limited to teachers who, prior to the assessment seminar, had taught a LeTUS curriculum and taken a related professional development class. They were teaching one of three LeTUS curricula concurrently with taking the professional development course. The specific curriculum they were teaching served as the basis for the formation of discussion groups within the professional development class. Teachers and researchers came together over ten sessions to discuss, design, and revise performance assessments for their particular inquiry-based science curriculum. Teachers taught the unit and implemented their assessment designs during the same time frame. Additionally, participants engaged in "moderation sessions" in which they analyzed and discussed student work from their own and others' classrooms (Wilson & Sloane, 2000). In weekly reflections and in class discussions, the teachers made connections between their practice and larger assessment ideas. As a culminating project teachers created a portfolio containing reflections from across the course, and specifically, their analyses of student work.

The classroom enactment

Data for this paper were drawn from one of the three groups of teachers participating in the professional development. These teachers were teaching a new 7th grade inquiry-based science curriculum designed by the LeTUS Center called Planetary Forecaster, in which students examined factors affecting global temperature variation. The unit is divided into four phases: a) eliciting students' prior understandings and generating initial predictions about the factors that affect global temperature using computer-generated temperature data; b) generating specific hypotheses about factors to investigate; c) investigating and explaining factors that may or may not affect global temperature differences based on data from computer-based and hands-on investigations; and d) applying the principles of temperature variation developed in phase C investigations to the temperature prediction for a hypothetical world. As part of the curriculum, content and process learning goals were identified and embedded assessments were designed into the activities. Teachers also used a software tool called the Progress Portfolio, designed to support students' inquiry practice (Loh et al., 2001), as part of the curriculum. The Planetary Forecaster group consisted of four teachers. Two of the teachers each taught the curriculum to one of their seventh grade science classes. The other two teachers were team-teaching the curriculum to a fifth grade class.

Data collection and methods

As described above, the data for this paper were collected from the group of Planetary Forecaster teachers participating in the professional development. As part of this study we interviewed these teachers to inquire about their classroom assessment practices and their beliefs about classroom assessment in the context of inquiry-based science. The semi-structured interviews were conducted during the summer after the professional development using the same set of open-ended questions, and were audio-taped and transcribed. The interviews ranged from 45 minutes to two hours, depending on the teacher. The interview data is supplemented by observations of classroom practice and artifacts from classrooms (including teacher-designed materials and student work). For the purposes of this paper, we focus primarily on the interview data.

From that group of teachers, we selected two to describe in these cases. We chose to describe the assessment stories of one fifth-grade and one seventh-grade teacher because they help to document diverse variations and to identify important common patterns (Kuzel 1992; Patton 1990). These teachers were selected as examples of the challenges facing assessment reformers. We selected Marian[i] because of the stark contrast between her assessment beliefs and the assumptions of the assessment approach. Marian was skeptical of the value of the assessment techniques promoted in the professional development class, and was clear in her reasons for rejecting them. In contrast, Beth provided a case of a teacher who, although many of her beliefs were aligned with the assumptions of the assessment approach, still faced challenges in her classroom assessment.

Coding themes were suggested from the design principles of the professional development and from the research literature. The professional development had a particular vision of assessment as its goal (Wilson & Sloane, 2000; Shepard, 2000), and that implied certain beliefs about assessment. We identified these as the following:

1. The teacher considers the development of student thinking, looking at prior and current conceptions as well as the potential learning trajectory.
2. Student development of content understanding, inquiry skills, and metacognitive skills can and should be monitored and documented. This documentation provides information for use in feedback to both students and teacher, for use in furthering that development.
3. Assessment should be formative in addition to summative. Students should receive feedback about where they are now, where they need to get, and how they can get there. Assessment should also be used to inform instruction.

While not a comprehensive list of beliefs, we felt that these represented some of the core ideas of the assessment approach promoted in the professional development. These beliefs we categorized as related to the *goals and nature of assessment*.

Next, we examined the transcripts of the final interviews and made note of patterns, especially those that might suggest new themes. As a result, two additional themes emerged from the data: *beliefs about assessment constraints* and *assessment self-confidence*. We then returned to the transcripts and coded for teacher beliefs related to all three themes. In this paper we present findings related only to the first of these themes, *beliefs about the goals and nature of assessment*. In particular, we focus on four specific sub-themes in this area: (a) teachers' *assessment values*—what is important to them to assess, (b) teachers' beliefs about the *purpose for assessment*, (c) teachers' beliefs about the *audience for assessment*, and (d) teachers' *view of student development*. We culled the interviews for examples that reflected these themes. From those examples we created our cases.

Findings

Through this work, we found that the underlying causes of challenges to teachers' reform-based assessment practices were deep-rooted and difficult to address. In case studies of two teachers involved in our professional development, we found difficulties in two key areas (see Matese, Griesdorn & Edelson, 2002 for a more detailed description of the cases summarized here). First, we saw *conflict between teacher beliefs and the principles of the assessment system*. Second, even when teacher beliefs were aligned with the assessment principles, we found that *teachers face challenges operationalizing those beliefs.*

Conflict between teacher beliefs and the principles of the assessment system

In the case of the former we found that one of our case study teachers had underlying beliefs about the nature of science, pedagogy, and assessment that were fundamentally different than the principles of the assessment system

we were trying to support. Table 1 below summarizes our findings on these teachers' assessment beliefs (Matese, Griesdorn & Edelson, 2002).

Table 1: Case study teachers' beliefs about inquiry-based science assessment

	Beth	Marian
What assessments are valued?	Values both content and inquiry skills. Engages in both informal and formal assessments. Practices formative assessment.	Inquiry and metacognitive skills not as valued as content knowledge. Formal summative assessments of content, such as lab reports or quizzes, are most important.
What is the purpose of assessing students?	Guide instruction, provide feedback to students. Sees feedback to students and instruction as an important goal of assessment.	Ensure students receive correct content; produce grades.
Who is the audience?	Students	Parents, students, administrators
What is the view of student development?	Goal is to track student development over time (such as their development of hypotheses), but finds the actual practice challenging.	Student development is understood and tracked in terms of grades.

A quick glance at the table reveals stark differences in these teachers' beliefs about assessment in inquiry-based science. While Beth's are in line with the principles of the assessment system, Marian's conflict in fundamental ways. These differences can be summed up in their attitude toward a key aspect of the assessment system: the use of scoring guides to understand, document, and communicate to students their progress toward key learning goals. For Marian, the costs in time and energy of using scoring guides outweigh the expected benefits. With an eye to her assessment audience of parents and administrators, Marian felt discomfort with assessment by scoring guides, describing it as too subjective and difficult to defend: "Even in the way these [scoring guides] are now, there's just such a degree of subjectivity to it. If you were going to use it to support why you were giving a kid a certain grade, because that's really the only reason why you'd have to use them, to support a grade. But then you'd have the parent coming and saying, what makes you think that wasn't a "serious effort"? There's just no, no, I wouldn't use it. I don't know how I would." For Beth on the other hand, scoring guides provide clarity in her assessments, direct her attention to the important features of student performance, and are "hands down" the ideal way to assess students.

Challenges operationalizing beliefs

Though Beth's beliefs fit well with the principles of the assessment system, we see evidence of yet another challenge in our analysis of her case. We sometimes saw inconsistencies between Beth's espoused beliefs and her practice. For example, in contrast to Marian, Beth clearly articulated process learning goals in addition to content. In describing her assessment of a particular activity, she describes the range of her learning goals for the whole curriculum:

> [T]o me, the four causes for changes in temperature were not the only things I wanted to focus on as we went through this. There were many other science skills that I wanted to get rolling for them, including data gathering and filling out these charts and making comparisons and drawing conclusions based on comparisons, so there was a lot happening for me.

When asked about the value of the curriculum's scoring guides, Beth again emphasizes the value of the skills they capture, and articulates why she believes this: "Because analysis of data is one of those higher-order thinking skills. And I think that's really important." In her description of what she feels is important to assess, Beth's beliefs are clearly aligned with the assumptions of the assessment system. But when Beth actually describes her assessment, it seems these other skills, and the use of the scoring guides to assess them, fall through the cracks:

> And when I graded, sometimes I graded letter grades and sometimes I used my check plus, check, check minus, check minus minus based on the amount of understanding I deemed---and it was kind of my own, I have a mental rubric, you know. I can see that this person understands the two or the three concepts. They've explained it pretty clearly---that was a check plus. I can see that this person understands some of the concepts and---you know, maybe they're having a little bit of

difficulty explaining it, but I can see there's threads in there that are coming together. That's a check. This person is like, missing chunks. That's a check minus.

Her first thoughts in describing her assessments are to refer to students' content understanding. Later, when pressed by the interviewer, Beth recalls that she looked at inquiry skills in assessing their labs, but this is an afterthought in her description of her practice. There remains a gap between what Beth would like to do and what she is able to do. Throughout the interview she reflects on things she would like to do differently in her assessment. For her, beliefs were not the challenge but operationalizing them in her particular context was.

Design principles for professional development to support inquiry-based science assessment

Because authentic assessments are critical to inquiry-based science curriculum and instruction we pose these challenges as design problems to be addressed by professional development designers. In this section we outline our ideas concerning design principles for professional development that scaffolds reform-minded assessment practices. We discuss the principles and how they relate to our assumptions about assessment and about teacher learning, along with illustrations from our pilot test experience. Specifically, we propose the following principles:

1. Teachers need opportunities to problematize their current inquiry-based science assessment practice.
2. Teachers need opportunities to operationalize their understanding of assessment tools.

The intention is not that professional development include either one or the other, but that it provide teachers with opportunities for both kinds of experiences.

Certain recurring themes in the design of professional development (Wilson & Berne, 1999) run throughout these principles and are not isolated for discussion here. First, we assume that this professional development is situated in the context of teachers' classroom practice. Second, we assume that the professional development affords opportunities for collaboration, discussion, and exchange with peers. Finally, we assume that the professional development is ongoing over an extended period of time (for example, over the course of a school year or longer).

1. Teachers need opportunities to problematize their current inquiry-based science assessment practice.

Teachers need opportunities to reflect on their assessment practices in light of the goals of inquiry. How do their assessments fit with their learning goals and with the instructional strategies they are using? We refer to this as opportunities to problematize current assessment practices.

For example, as part of the professional development, the teachers experimented with an inquiry-based science curriculum in which assessment, instruction, and technology supports were tightly integrated. In particular, tasks incorporated specially designed technology scaffolds that fit with the assessment and that served as a direct reminder to the teacher of what was being assessed. As suggested, Marian used the technology to provide a different kind of feedback to her students. In her interview she cites this as the communication with students that "worked best." Instead of her typical feedback in which she looked at a snapshot of their content understanding and filled in what is missing ("you should have seen this or you should have made a connection"), she provided feedback in the Progress Portfolio software with the expectation that students would respond to her comments and would continue to develop their ideas. Furthermore she looked not just at whether the student was right or wrong in the moment, as she did when grading labs, but saw an opportunity to point to ways in which the work might be improved. She recognized that for her this was an anomaly, but she liked the kind of feedback she was able to provide in the software. Though she changed her practice in this one instance, once past the experimentation period she returned to her previous assessment practices.

Marian's experience represents both a success and a failure. Despite the opportunity to try new practices, her current assessment practices were not problematic enough to lead to sustained change. She tried new kinds of assessment, but her beliefs about what was important to assess and the methods for doing so remained unchanged. In the context of inquiry-based science, Marian's learning goals for her students were quite limited. To problematize her assessment practice, her beliefs needed to be elicited as objects for reflection (Bliem & Davinroy, 1997). The contrast between her current practices and the new practices, and the potential implications of these differences, should have been explicitly discussed. Perhaps concrete models and examples of instruction and assessment from expert teachers in other inquiry-based science classrooms could have prompted Marian to consider her beliefs and

their implications for establishing learning goals and designing inquiry-based instruction and assessment. Whatever the particular strategy, problematizing beliefs underlying assessment practices is a formidable task and one that cannot be achieved without ongoing professional development over an extended period of time.

With Marian, core beliefs underlying her learning goals and the assessment of them needed to be problematized and reconsidered. For others in the professional development, the need was in considering how well their assessments aligned with their existing learning goals and instruction. Teachers' learning goals for students imply particular instructional and assessment decisions. Some participants in the professional development had ambitious learning goals for their students, and were attempting new forms of inquiry-based instruction in line with those goals. Though for the most part the teachers were implementing inquiry-based science activities, they often neglected the assessment aspects of them. In these cases, teachers' attention needs to be drawn to the lack of alignment in their practice. Reflection on this alignment (or lack thereof) may problematize assessment practices for these teachers.

Problematizing assessment practices is difficult, to say the least. Opportunities for teachers to problematize their assessment practice are unlikely to occur in professional development that is disconnected from their classroom context, such as a summer workshop. Nor are opportunities likely to occur frequently enough, or be powerful enough, to effect any real change in a short period of time. Instead, teachers need multiple opportunities over an extended period of time to try inquiry-based science assessments in the context of their classrooms and to reflect on their practice.

2. Teachers need opportunities to operationalize their understanding of assessment tools.

As we saw earlier, Beth faced a challenge in operationalizing her beliefs about assessment. Beth's ability to discuss an assessment design with her peers, test it out in her classroom, and then talk to other teachers about the resulting student work offered her an opportunity to reflect on and change her practice. As a first step in operationalizing their assessment tools, teachers need to construct an understanding of what learning goals are appropriate to pursue in the context of a particular curriculum. In Beth's final interview she expressed her struggle to do that in this new curriculum:

> [T]he next time I do [the curriculum] I'm going to keep more careful track of their revising, particularly in the revision of hypotheses. I wasn't able to focus on that as much as I would have liked to. There's just so many aspects to this that when I do it again, I'm going to make a list of all these aspects. And you know, really, now that I understand how it fits together . . . I'm going to need to make some kind of a checklist for myself of the various things I want to touch in it and. . . where to address them, how to keep track of them. So that I do wind up hitting all of them.

A tool similar to what Beth describes was provided in the curriculum, but she needed support to construct her own understanding of this tool in the context of her classroom.

Beth cites a second challenge in operationalizing assessment tools, once an understanding of the scope of the learning goals is reached: "Really deciding if they were understanding the concepts. . . . So designing questions that don't just ask 'what is tilt?'; 'How do you know tilt happens?' You know, like really, straightforward just content fill-in-the-blanks or you know, one-sentence answers. But designing something that allows them to demonstrate that they understand what the concept is instead of just telling you what the concept is. That was the challenge." Once the teacher has a sense of *what* they would like to assess, they then need to figure out *how* to assess that. They must come up with assessment designs that provide pertinent information about student learning—not always an easy task.

For Beth, learning often happened in the course of her interaction and collaboration with others in the professional development: "It gives you another perspective on things that I think is a positive thing." Beth received support in these efforts through opportunities to share and discuss her assessment designs. In the interview Beth tells an anecdote about a conversation with one of the professional development facilitators before class one evening:

> I was in the process of developing an assessment. And [a facilitator] walked in and said, "first of all, what are you looking for and second of all, are you actually getting it based on what you're doing here?" [This] rings a bell to me. Ok, then I decided that I am, in fact—because originally I was gonna pitch [the assessment design] —no, I am, in fact, looking for this. Not what I originally thought I was looking for. And I do want to know it. So I was able to come to some kind of decisions. Some kind of more enlightened decisions about exactly what I was looking for, why I

was looking for it and did I think it was important or not, or is it just extraneous, you know?

The conversation led her to reconsider the design at hand and reevaluate what she was actually assessing with it. While she was not assessing what she originally thought she was, she decided that the information elicited from the design was just as important to her. The conversation provided her with a model for understanding and evaluating her assessment designs.

Collaboration was helpful not just in designing assessment tools to collect information about student thinking, but also in developing the key skill of interpreting that information (National Research Council, 2001). For example, in an activity called "moderation" (Wilson & Sloane, 2000), teachers brought selections of student work for the group to assess according to shared scoring guides. Scores were compared and discussed. This experience allowed teachers to further their understanding of what is meant by particular learning goals, and to build their skills in interpreting the assessment information they gather in student artifacts. Beth notes that conversations with the other teachers in the group about her students' work have prompted her to think more carefully about her interpretation of that work: "When we did the moderation I realized that I was making inferences that were unfounded. That I shouldn't be assuming this based on some sentence that they made. So after that I was much more careful."

Conclusion and future work

In this paper we have offered two design principles for professional development for inquiry-based science assessment. We want to be clear, however, that no one instructional design is perfect for all audiences. Descriptions of reform-based teaching emphasize that teachers must modify their instructional plan in response to the needs of their students, and the same is true of professional development. However, we suggest that the principles we propose can serve as important guides for the design of professional development to support teachers' learning about assessment of inquiry-based science. For there to be change in teacher practice, professional development in support of inquiry-based science assessment will need to provide teachers with opportunities for both problematizing their current assessment practices and operationalizing assessment tools in the context of their own classrooms.

A goal of future work is to examine how teachers *develop* assessment practices in the context of inquiry-based science. Thus far we have described ways of supporting teachers' efforts to assess inquiry-based science. We want also to describe the paths along which these changes take place. Like classroom teachers, we expect professional development facilitators to draw on many kinds of knowledge to interpret and respond to the needs of participants. Just as we provide teachers with potential learning trajectories for students in different content areas, we want to be able to provide professional development facilitators with potential learning trajectories for teachers in assessing inquiry-based science. We currently do not have many examples of models or developmental trajectories depicting the process of teacher learning (notable exceptions include research by Simon (1995) and Franke, Fennema, & Carpenter, (1997)). Thus, a goal of this research is to understand what the trajectory might be for teachers' learning of reform-minded assessment in inquiry-based science. The design principles presented here are the first steps in that process.

References

Black, P. J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principle, Policy, and Practice, 5*(1), 7-74.

Black, P. J., & Wiliam, D. (1998b). Inside the black box: Raising standards through formative assessment. *Phi Delta Kappan, 80*(2), 139-148.

Bliem, C. L., & Davinroy, K. H. (1997). *Teachers' beliefs about assessment and instruction in literacy*. (Center for the Study of Evaluation Report 421). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing

Borko, H., Mayfield, V., Marion, S. F., Flexer, R. J., & Cumbo, K. (1997). *Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development*. (Center for the Study of Evaluation Report 523). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment 4*(1), 37-73.

Franke, M. L., Fennema, E., & Carpenter, T. (1997). Teachers creating change: Examining evolving beliefs and classroom practices. In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 255-282). Mahwah, NJ: Erlbaum.

Kuzel, A.J. (1992). Sampling in Qualitative Inquiry. In B.F. Crabtree & W.L. Miller (Eds.), *Doing qualitative research* (pp.31-44) (Research Methods for Primary Care Series, Vol.3). Newbury Park, CA: Sage.

Loh, B., Reiser, B.J., Radinsky, J., Edelson, D.C., Gomez, L.G., & Marshall, S. (2001). Developing Reflective Inquiry Practices: A Case Study of Software, the Teacher, and Students. In K. Crowley, C. Schunn, & T. Okada, (Eds.), *Designing for Science: Implications from Everyday, Classroom, and Professional Settings*. Mahwah, NJ: Erlbaum.

Matese, G., Griesdorn, J., & Edelson, D. C. (2002, April). *Professional development in support of assessing inquiry-based science: A design research approach.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

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

National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Patton, M.Q. (1980). *Qualitative evaluation and research methods* (2nd ed.) Newbury Park, CA: Sage.

Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning* (Center for the Study of Evaluation Report 517). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114-145.

Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappa, 72*, 534-539.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition & Instruction, 16*(1), 3-118.

Wilson, S. M. & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education, Vol. 24* (pp. 173-209). Washington, DC: American Educational Research Association.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*(2): 181-208.

Acknowledgments

We are grateful to our collaborators in this research: Edys Quellmalz and Chris Hoadley from SRI; Mark Wilson and Deborah Peres from the Berkeley Evaluation and Assessment Research Center at the University of California, Berkeley; Laura Baumgartner and Herb Thier from Lawrence Hall of Science; and Jacqueline Griesdorn, Louis Gomez and Brian Reiser from Northwestern University. We would also like to thank the teachers who participated in this research. This material is based upon work supported by the National Science Foundation under Grants No. REC-9730377 and REC-0087751. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[i] Teacher names are pseudonyms.