

Informal Formative Assessment and Scientific Inquiry: Exploring Teachers' Practices and Student Learning

Maria Araceli Ruiz-Primo
School of Education
University of Colorado at Boulder

Erin Marie Furtak
School of Education
Stanford University

What does informal formative assessment look like in the context of scientific inquiry teaching? Is it possible to identify different levels of informal assessment practices? Can different levels of informal assessment practices be related to levels of student learning? This study addresses these issues by exploring how 4 middle school science teachers used questions as a method of informal formative assessment, and compares those practices to measures of student learning. The approach to exploring each teacher's questioning practices is based on viewing whole-class discussions as assessment conversations in which the teacher has the opportunity to draw out and act on students' evolving understanding. Assessment conversations are described as consisting of four-step cycles, where the teacher elicits a question, the student responds, the teacher recognizes the student's response, and then uses the information collected to student learning. Our results indicate that the teachers whose enactment of informal formative assessment was more consistent with this model had students with higher performance on embedded assessments. This trend was also reflected in the posttest scores. In addition, we found that teachers focused more on epistemic, rather than conceptual, features of scientific inquiry in their discussions. The study underlines the importance of informal formative assessment during scientific inquiry discussions for teacher training and professional development as a way to increase student learning.

Recent reforms in science education have emphasized the importance of teaching and learning through scientific inquiry. It has been argued that meaningful learning of science should involve not only performing investigations to collect data, but also the construction and evaluation of scientific claims (National Research Council [NRC], 1996; 2001). As Duschl (2000) wrote:

When students learn about what is known, without also learning how we have come to know it, and why this belief or conceptual scheme is better than another belief, it eliminates any chance of students' understanding the social, cognitive and epistemic dynamics that make science an objective way of knowing. (p. 187)

That is, if students are taught science in the context of inquiry, they will know *what* they know, *how* they know it, and *why* they believe it (Duschl, 2003). However, effective implementation of scientific inquiry poses significant instructional challenges to teachers (Tobin, Tippins, & Gallard, 1994; Welch, Klopfer, Aikenhead, & Robinson, 1981).

Among these challenges is being sensitive to students' present level of understanding so that instruction can be continuously modified while learning is still taking place. This means that continuous assessment of students' understanding to improve teaching and learning is required; that is, formative assessment, assessment for learning and not of learning, is needed (Bell, 2000; Bell & Cowie, 2001; Black & Harrison, 2001; Black & Wiliam, 1998).

Formative assessment involves gathering, interpreting, and acting on information about students' learning so that it may be improved (Bell & Cowie, 2001; see also Duschl, 2003). In a study of eight New Zealand teachers' classroom assessment practices, Bell and Cowie (2001) distinguished between two types of formative assessment: (a) formal or planned formative assessment, which focuses on obtaining information about student learning from a whole class; and (b) informal or interactive formative assessment, which focuses on obtaining information about student learning whenever possible, in any student-teacher interaction.

In this article, we focus on informal formative assessment for two reasons. First, the science education community is still beginning to understand the integral role of informal formative assessment in scientific inquiry teaching. Black and Wiliam's (1998) well-known meta-analysis of formative assessment focused primarily on formal methods of formative assessment. Second, defining more clearly what researchers mean by *informal assessment practices* can help them to understand how informal formative assessment may be involved with student learning. Thus, if we can better capture what effective informal formative assessment looks like in science classrooms, we can help other teachers improve their own practices through preservice training and professional development.

To that end, we focus on three questions:

1. What does informal formative assessment look like in the context of scientific inquiry teaching?
2. Is it possible to identify different levels of informal assessment practices?
3. Can different levels of informal assessment practices be linked to levels of student learning?

First, we present a framework for exploring teachers' informal assessment practices during discussions in scientific inquiry investigations. Then, we describe how the framework was applied to the videotaped discussions of four middle school science teachers implementing the same physical science units. Finally, we provide information about the learning of each teacher's students, not with the purpose to establish a causal relation, but with the intention of suggesting reasonable hypotheses for future research.

A FRAMEWORK FOR EXPLORING INFORMAL FORMATIVE ASSESSMENT IN SCIENTIFIC INQUIRY TEACHING

The framework we propose is based on the idea that informal formative assessment can take place at any level of student–teacher interaction in the course of daily classroom talk, whether whole class, small group, or one-on-one (Bell & Cowie, 2001; Duschl, 2003; Shavelson, Black, Wiliam, & Coffey, in press), and can help teachers continuously acquire information about their students' level of understanding.

We adopted the term *assessment conversation* (Duschl, 2003; Duschl & Gitomer, 1997) to refer to these daily instructional dialogues that embed assessment into an activity already occurring in the classroom. In contrast to initiation–response–evaluation (IRE) sequences that involve the teacher initiating a query, the student responding, and the teacher evaluating the student's contribution (Lemke, 1990), assessment conversations permit teachers to gather information about the status of students' conceptions, mental models, strategies, language use, or communication skills to guide instruction. We (Ruiz-Primo & Furtak, in press) have characterized these assessment conversations as consisting of four elements: The teacher asks a question to elicit student thinking, the student provides a response, the teacher recognizes the student's response, and then uses the information collected to support student learning. These components are aligned not only to the formative assessment components described (i.e., gathering, interpreting, and acting), but also to the moves (soliciting, responding, reacting) used to de-

scribe classroom discourse (Bellack, Kliebard, Hyman, & Smith, cited in Carlsen, 1991). Taken together, these four elements are called ESRU cycles.

Each step in the ESRU cycle serves a purpose toward collecting information about student learning, comparing it to the teacher's expectations, and taking action to move students toward learning goals. Eliciting requires the teacher to ask questions that allow students to share and exhibit their understanding as completely as possible (e.g., asking student to share their ideas). The question asked by the teacher gives the student an opportunity to make her thinking explicit. Then, the teacher recognizes the student's response by acknowledging her contribution in some way and comparing it, perhaps implicitly, to some learning goal (e.g., restating the student's comment to the class). Using information means that the teacher takes action on the basis of student responses to help students move toward learning goals (e.g., encouraging students to reach consensus on a particular explanation).

Differences between the ESRU and IRE patterns of teacher-student interaction are based on the nature of the statements made by the teacher in the ESRU cycle. First, the eliciting questions have the potential to provide the teacher with information on the evolving status of students' conceptions and understandings about scientific inquiry skills and habits of mind. Second, the teacher's recognizing statements do not evaluate what the student has said, but serve the purpose of the teacher validating each student's contribution as it is absorbed into the ongoing classroom narrative (Scott, 2004). Finally, the most important characteristic distinguishing ESRU patterns from IRE sequences is the element of using; that is, the ESRU pattern ends with the teacher taking action to increase student learning. For example, a teacher can provide students with specific information on actions they may take to reach learning goals or ask another question that challenges or redirects the students' thinking (Black & Wiliam, 1998; Duschl, 2000, 2003; Duschl & Gitomer, 1997; Minstrell & vanZee, 2003; NRC, 2001; Ramaprasad, 1983; Sadler, 1989, 1998).

In our framework, assessment conversations (i.e., ESRU cycles) are analyzed in the context of the three integrated aspects of science education originally proposed by Driver, Leach, Millar, and Scott (1996) and elaborated by Duschl (2000, 2003) as domains of assessment: *epistemic*, *conceptual*, and *social*. Epistemic structures are the knowledge frameworks that involve the rules and criteria used to develop or judge what counts as science (e.g., experiments, hypotheses, or explanations). Epistemic frameworks emphasize not only the abilities involved in the processes of science (e.g., observing; hypothesizing; experimenting; and using evidence, logic, and knowledge to construct explanations), but also the development of the criteria to make judgments about the products of scientific inquiry (e.g., explanations or any other scientific information). Conceptual structures involve deep understanding of concepts and principles as parts of larger scientific conceptual schemes. Scientific inquiry requires knowledge integration of those concepts and

principles, allowing students to use that knowledge in an effective manner in appropriate situations. Social processes refer to the frameworks involved in students' scientific communications while engaging in scientific inquiry, and can be oral, written, or visual. The social domain focuses on "the processes and form that shape how knowledge is communicated, represented, argued and debated" (Duschl, 2003, p. 42). Because any assessment conversation is, by nature, a social interaction between teacher and students, it is difficult to distinguish from the epistemic and conceptual domains. Therefore, ESRU cycles can only be classified as epistemic or conceptual because the social dimension is embedded in both.

In all but the most open, exploratory classroom settings, the teacher guides the course of discussions by eliciting information about student understanding. Thus, our framework for exploring assessment conversations in scientific inquiry primarily consists of the questions and statements made by the teacher. However, we acknowledge that the information obtained can also be interpreted and acted on by the students.

The questions and statements are organized as strategies according to the teacher elements of the ESRU cycle across epistemic frameworks and conceptual structures. These questions and statements were defined and selected from direct observation in multiple classrooms, as well as from literature on scientific inquiry (American Association for the Advancement of Science, 1993; Bransford, Brown, & Cocking, 2000; Ebenezer & Haggerty, 1999; NRC, 1996, 2001; Stipek, Salmon, & Givvin, 1998; White & Frederiksen, 1998). Many of the questions and statements included have been considered by others as strategies that promote conceptual change (e.g., Roth, 1996; Smith, Blakeslee, & Anderson, 1993), scientific thinking (e.g., Cocking, Mestre, & Brown, 2000), and indicators of good teaching practices (e.g., Carlsen, 1991; King, 1994; Roth, 1996). The questions and statements proposed do not form an exhaustive list of strategies that can be used, but an initial set that can be considered in the study of informal formative assessments. The ESRU question and statement strategies are presented in Table 1.

We applied our framework to the practices of four middle-school teachers during the implementation of four physical science investigations. We next describe the context in which this study is embedded and the methods used to collect the data across the four teachers and their students.

STUDY CONTEXT

During the 2003–2004 school year, the Stanford Education Assessment Laboratory (SEAL) and the Curriculum Research and Development Group (CRDG) at the University of Hawaii at Manoa collaborated in a study of the effects of formal formative assessment on students' learning (Shavelson & Young, 2000). The study embedded diverse formative assessments into the Foundational Approaches in

TABLE 1
Strategies by ESRU Cycle Phase

<i>Eliciting—Epistemic Only</i>	<i>Eliciting—Conceptual Only</i>
<p>Teacher asks students to:</p> <ul style="list-style-type: none"> Compare and contrast observations, data, or procedures Use and apply known procedures Provide observations Make predictions or provide hypotheses Formulate scientific explanations Provide evidence and example Interpret information, data, patterns Relate evidence and explanations Evaluate the quality of evidence Compare or contrast others' ideas, explanations Suggest hypothetical procedures or experimental plans <p>Teacher:</p> <ul style="list-style-type: none"> Promotes students' thinking Questions to check students' comprehension <p>Other epistemic strategies</p>	<p>Teacher asks students to:</p> <ul style="list-style-type: none"> Provide potential or actual definitions Apply, relate, compare, contrast concepts Compare or contrast others' definitions <p>Teacher:</p> <ul style="list-style-type: none"> Promotes students' thinking Questions to check students' comprehension <p>Other conceptual strategies</p>
<i>Recognizing—Both Dimensions</i>	<i>Using—Both Dimensions</i>
<p>Teacher:</p> <ul style="list-style-type: none"> Clarifies/elaborates on student response Takes votes to acknowledge different students' ideas Repeats/rephrases student's contribution Revoices student's words Captures/displays students' diverse explanations Provides evaluative responses Responds with a yes/no, fill-in-the-answer question Provides neutral response Responds with a minilecture not connected to the student's response Asks to repeat what other student said Answers her/his own question 	<p>Teacher:</p> <ul style="list-style-type: none"> Promotes student thinking with why/how questions Compares/contrasts students' alternative explanations Promotes consensus/come to an agreement Helps relate evidence to explanations Provides helpful feedback Models process skills Makes learning goals, expectations, standards explicit Connects to previous learning, investigations, conclusions Refers to nature of science Solicits and encourages diversity of opinions Promotes exploration of student's ideas Explores student's ideas—Meaning into matter Provides students with additional information (scaffolding) Acts on information accumulated in multiple ESRU cycles

Note. ESRU = teacher Elicits, Student responses, teacher Recognizes, teacher Uses.

Science Teaching (FAST) curriculum, a multidisciplinary middle-school science curriculum developed by the CRDG (Pottenger & Young, 1992) and aligned with the *National Science Education Standards* (NRC, 1996; Rogg & Kahle, 1997).

The SEAL/CRDG project focused on the first 12 investigations of the Physical Science (PS) strand of *FAST 1, The Local Environment*. In this strand, students investigate concepts such as mass, volume, and density, as well as the relationship between density and buoyancy. Students' learning was measured in a pre- and posttest design with multiple assessments of student learning (i.e., multiple choice, short answer, performance assessment, and predict–observe–explain).

Six matched pairs of FAST teachers were randomly assigned to experimental and control groups.¹ The control group teachers implemented the FAST investigation as usual, whereas experimental group teachers participated in a 5-day training program focusing mainly on how to implement formal embedded assessments and how to use information taken from these assessments to provide immediate feedback to students. The four teachers whose informal formative assessment practices are explored in this article were members of the experimental group in the SEAL/CRDG study.

METHOD

Participants

The four teachers studied in this article were selected for the following reasons: (a) Teachers represented a range of experiences and varied considerably in their teaching practices. (b) All teachers had been formally prepared by CRDG trainers in the FAST curriculum. (c) All teachers participated in the same training program for the experimental teachers in the larger project; therefore, if the program had any effect on their teaching and assessment practices, it should be expected to occur in the same direction. Finally, (d) focusing on the experimental teachers allowed us to have information on the students' learning at different points of instruction by using the information collected through the embedded assessments implemented in the larger project. All names have been changed to ensure the anonymity of the participants.

Data Collection

We collected information from teachers and students. The teachers (Adam, Carol, Diana, and Ray) were asked to videotape their classrooms in every science session

¹Ethnicity, free lunch, and student proficiency level were used to match the pairs as well as possible.

they taught during the 12 FAST 1 investigations. Each teacher was provided with a digital video camera, a microphone, and videotapes. All were trained on how to videotape their classes. We collected videotapes of 49 lessons across the four teachers over the four investigations.

Students were administered multiple types of assessments over the course of the implementation of the FAST investigations: a multiple choice test before and after the FAST investigations, and diverse embedded assessments after Investigation 4 (PS4), Investigation 7 (PS7), and Investigation 10 (PS10).

In this article, we focus on analyzing the informal assessment practices during the discussions of results that occurred at the end of four of the 12 FAST investigations: PS1, PS4, PS7, and PS10. The rationale behind this decision is based on three facts. First, the role of evidence in the development of scientific explanations becomes relevant in these discussions as the students and teacher actively construct meaning from data collected during the investigation (Cazden, 2001; Duschl, 2003; Edwards & Mercer, 1987; Lemke, 1990; Scott, 2004). Second, PS4, PS7, and PS10 were immediately followed by embedded assessments implemented as part of the SEAL/CRDG study. Third, in a previous study we found that the highest percentage (about 39%) of the assessment conversations occurred around the discussion of the investigation results (Ruiz-Primo & Furtak, 2004).

On the videotapes collected, no discussions of results were identified in 29 lessons, and one videotape did not have sound. Diana's tapes for PS7 did not have sound during the discussion portion of the investigation, and Carol did not submit her tapes for the discussion of PS10. This left 19 videotapes to be coded. The total time of discussions coded amounted to 388 min.

Instruments

ESRU coding system. Videotapes for every lesson taught at PS1, PS4, PS7, and PS10 were transcribed for each teacher. Transcriptions were segmented following the speaking turns between the teacher and the students (or student to student). All transcriptions were analyzed twice. The first analysis was used to determine which part of each transcript corresponded to discussion of results. To identify these discussions, we considered the following: The students had already conducted the investigation and constructed a graph of their results; the teacher and students took turns speaking; student responses were elicited by the teacher through questions; and the conversation took place in a whole-class setting (as opposed to a teacher working one-on-one with students or with small groups). The second analysis focused on coding the individual speaking turns. Transcripts were coded in tandem with watching the videotapes.

The framework for assessment conversations was then used to develop a coding system to capture the ESRU cycles in the transcripts. Each speaking turn was given a code that corresponded to an element of the cycle; complete cycles (i.e., elicit, student response, recognize, and use), incomplete cycles (e.g., elicit, student re-

sponse, and recognize; or only elicit and student response), and cycles initiated by students (e.g., student comment, recognize, use) were coded. Thus the coding system enabled us to (a) capture whether an eliciting strategy is followed by the other phases, (b) capture the specific strategies used in each phase of the cycle, (c) identify the strategies used more frequently, and (d) identify whether more than one cycle is being carried out with one student. The coding approach also allowed us to capture the sequence in which the strategies were used.²

We focused on determining whether two independent coders classified the same type of cycles (i.e., complete or incomplete) and number of cycles under the epistemic and conceptual domains. From the 388 min of videotape collected, 49 min were used to train the two coders in the coding system, and then 160 min (about 20% of the total number of minutes) were randomly selected and coded independently to assess the extent to which the two coders were able to consistently identify the ESRU cycles.

The averaged intercoder reliability coefficient across the three types of ESRUs (i.e., complete cycles—ESRU—and incomplete cycles—ESR and ES) was .89. The lowest reliability (.81) was found for the incomplete cycle ES, in which an elicitation (E) and student response (S) were identified without any follow-up in terms of recognizing or using.³ It is important to mention that in the sample selected for intercoder reliability purposes, none of the coders identified any cycles linked to the conceptual domain. Once consistency in coding was established, the 315 min of videotape were independently coded.

Student embedded assessments (EA). We used the information collected in the pretest, posttest, and the embedded assessments administered immediately following the investigations at PS4, PS7, and PS10 as measures of student learning. We have named the three prompts used in the embedded assessments EA4, EA7, and EA10 following the investigation in which each was administered. The EAs were implemented across multiple class sessions (usually around two to three, and up to five) and were administered to all students within a classroom at the same time.

Each embedded assessment involved four types of prompts: Graphing, Predict–Observe–Explain (POE), Open-Ended Response, and Predict–Observe (PO) Question. However, in this study we focus on only two of the assessments: Graphing and POE.⁴ The Graphing prompt focused on the skill of each student to summarize, represent, and interpret data. POE assessment focused on assessing students' understanding about density and relative density. In this assessment, stu-

²For further information about the coding system, please contact the authors.

³The intercoder reliability for ESRU = .91, for ERS = .96, and for ES = .81. The reliability for *E*, where only an elicitation was identified, was below the minimum accepted. Therefore we do not report information on this incomplete cycle.

⁴We found that the results from POE, PO, and the Open-Ended Response provided similar information. Therefore we focus on only the POE.

dents first predict (P) the outcome of some event related to sinking and floating and justify their prediction. Then students observe (O) the teacher carry out the activity and describe the event that they see. And finally, students reconcile and explain (E) any conflict between prediction and observation.

The scoring of the graph focused on issues related to the accuracy of the graph and its interpretation. The POE scoring focused also on two composites: the quality of the justifications for the predictions and the quality of the explanations provided after the observation. Predictions were not considered because almost all students answered them correctly, whereas the quality of the explanations better reflected the various levels of students' understanding. We calculated intercoder reliability for the EAs using total scores. The averaged intercoder reliability coefficient for the quality of explanations across EA4, EA7, and EA10 was .86.

RESULTS

In this section, we first describe informal formative assessment practices across all four teachers, and then we take a closer look at sample assessment conversations taken from each classroom. Finally, we provide information about student learning based on the information collected on the embedded assessments.

Description of Informal Assessment Practices Across Teachers

We begin describing the informal formative assessment practices of each teacher by providing information on two aspects of instruction: the characterization of assessment conversations based on the type of ESRU cycles observed, and the characteristics of the strategies used in the ESRU cycles. The purpose of this section is not to provide a detailed description of the teachers' informal assessment practices, but to note the general characteristics of informal formative assessment observed across the four teachers based on the proposed framework.

Characterization of assessment conversations. According to our framework, the type and number of complete and incomplete cycles observed during the discussions are indicative of teachers' informal formative assessment practices. We argue that the more cycles a teacher completes, the more likely it becomes that the information gathered from students was used in a helpful way for learning purposes. In addition, the quality of informal assessment practices can also be determined by diversity and relevance of the strategies (questions and type of statements) used by the teacher. For example, the use of only one type of question, such as asking students to provide their observations, is made more useful in combination with other types of questions; for example, asking students to evaluate the

quality of those observations. We focus first on the profile of cycles and then on the strategies. (Percentages of the strategies used by cycle phase and teacher are presented in Appendix A.)

The incomplete cycle ESR was the type of cycle observed most frequently (59%), followed by the ESRU cycle (26%), and the incomplete ES cycle (15%). This pattern was replicated by all the teachers but Ray, whose percentage of ESRU and ES cycles was the same (4%). The teacher with the highest percentage of complete cycles over the four investigations was Diana (32%), who was followed by Carol (28%) and Adam (24%).

In several cases, several iterations of incomplete cycles (ESR) were observed before a using phase, yet the using statement made by the teacher referred to information collected in the preceding iterations of incomplete cycles. An example of this type of practice is provided by Carol, whose total assessment conversations (340)—almost double the numbers for Diana (197) and Adam (178)—included 198 that were coded as ESRs. However, these incomplete cycles usually reflected scenarios in which Carol elicited information from multiple students or elicited multiple responses from the same student before acting on the accumulated information. In this way, Carol was able, in a sense, to take the pulse of several students before using the information she had collected, or to help to advance the learning of a particular student. This type of iteration was not observed in Diana or Adam in a consistent manner, and was never observed in Ray's assessment conversations.

The next way we characterized assessment conversations was to focus on the questions and statements teachers used in the ESRU cycles as being of an epistemic or conceptual nature. The overwhelming majority of cycles observed across all teachers, whether complete or incomplete, were of an epistemic nature (more than 95%). Overall, the eliciting questions most frequently observed were those related to the interpretation of data and identification of patterns (24%). Asking students for observations and for explanations were the questions with the second and third highest percentages (11% and 10%, respectively). Low percentages were found for eliciting questions asking students to relate evidence to explanations, evaluate quality of evidence, or compare and contrast others' ideas. Of the questions asked in the conceptual domain, the most common were those that asked students for definitions (5%). Overall, Diana and Carol were the teachers with more diversity in the type and relevance of eliciting strategies they used.

The most common recognizing strategy used by teachers was to revoice or rephrase the students' responses. Carol and Diana tended to use this strategy in connection with other strategies, such as elaborating on a student's response. Diana and Carol had a higher percentage of revoicing statements (40% and 39%, respectively) than Adam (32%).

The using strategy most frequently observed was asking why and how questions to challenge or redirect students' thinking. This type of question allowed teachers to get more information about the students' understanding easily. Helping

students to relate evidence to explanations or making criteria explicit about what counts in science were rarely observed.

Clearly, asking students “Why do you think so?” or “What does that mean?” are questions that can easily help teachers to explore in more detail their students’ level of understanding. But these questions are even more powerful when combined with other strategies, such as comparing and contrasting students’ ideas and explanations. Encouraging students to provide more information or to consider why they are making a statement can help students to recognize the elements of good scientific explanations so that they will have a better idea of what to include on the next occasion. Providing helpful feedback was not a strategy used frequently. Helpful feedback should assist students in developing the ability to monitor their learning progress, as well as to judge the quality of their own work (Sadler, 1989; 1998). Helpful feedback of this kind given to one student can also be a secondary source of feedback for the rest of the students.

A closer look at the assessment conversations. We have provided information about some characteristics of the teachers’ informal assessment practices and pointed out some differences among them. In this section, we present excerpts that are representative of each teacher’s practices. We hope these excerpts help to contrast the informal assessment practices across the four teachers and provide a rationale for the results observed in the students’ performance.

Ray. It is difficult to gather any definitive conclusions about Ray’s informal assessment practices in discussions, because he had so few during the investigations sampled; however, in those we did analyze, his questioning was characterized by asking questions (often implicit) and then repeating students’ words with minimal instructional intervention. Considering only PS1, Ray had a high percentage (60%) of asking for predictions. This type of question is aligned with the characteristics of this investigation in which students are asked to predict where the different vials will go (i.e., float, sink, subsurface float) in the cylinder with liquids. Ray frequently responded to students by repeating the students’ responses. On few of the cycles did he elaborate based on the students’ responses. In the excerpt provided in Table 2, Ray has brought out a graduated cylinder filled with liquids of different densities and is asking students to provide observations. Note that some of the eliciting questions are marked with parentheses, indicating that these questions were implicit; that is, the teacher called on a student to respond without repeating the original question that had been asked.

This excerpt illustrates a common pattern in Ray’s class, in which he asked an initial question, and then called on students to provide responses. Although students at times offered interpretations rather than observations (e.g., “It’s plastic”), Ray accepted all comments at face value and then called on the next student. In this

TABLE 2
Excerpt From Ray's Discussion During PS1(9/23/03)

<i>Speaker</i>	<i>Dialogue</i>	<i>Cycle</i>	<i>Code (Teacher ...)</i>
Ray	Okay, Jeff, what do you think, what do you observe? Oh, just a second. These are just basic observations, this is not right or wrong, and I want you to understand, I want you to feel comfortable in here whenever you come up with an idea, kind of like Claire said awhile ago when she said what she thought it was, that's her idea, it's not right or wrong, it's her idea and so we'll just go with that flow. But the thing is you want to be cautious not to intimidate other people by making remarks or anything like that about what their observations are. Okay? We want this to be a fair assessment of a science classroom so that anything you say pretty much goes. Jeff.	E S (E)	Asks students to provide data or observations (repaired question)
Jeff	There is a clear liquid.	S	
Ray	Okay, so there is a clear liquid. Okay. Mark, from the back row, give us an observation.	R E	Repeats student's words Asks students to provide data or observations
Mark	It was almost full of liquid.	S	
Ray	Almost full. Okay. All right, Ralph, what do you think from the front row, again?	R (E)	Repeats student's words
Ralph	The graduated cylinder is tall.	S	
Ray	Okay. Tall cylinder. Okay. Greg.	R (E)	Repeats student's words
Greg	It goes [inaudible] like milliliters or something.	S	
Ray	Okay. Almost to 1000 milliliters. Okay. That's ... Anna?	R (E)	Repeats student's words
Anna	It's plastic.	S	

(continued)

TABLE 2 (Continued)

<i>Speaker</i>	<i>Dialogue</i>	<i>Cycle</i>	<i>Code (Teacher ...)</i>
Ray	Okay. It's plastic. Any other observations?	R	Repeats student's words
	... back to Jeff over here.	(E)	
Jeff	[Inaudible.]		
Ray	It has what?		
Jeff	Lines on it.	S	
Ray	Okay. It's got lines on there.	R	Repeats student's words
Jeff	It has an opening at the top.	–S	
Ray	Okay, it's open. All right. Now, that pretty much takes care of the cylinder and it takes care of the material that's inside, and you did say that there's liquids inside; one of you said that it's possible two liquids inside, okay, so there is a possibility that there's more than one liquid.	R	Revoices students' words
			Talks/clarifies/elaborates <i>not</i> based on students' responses

Note. PS = Physical science; E = Teacher elicits; S = Student responses; (E) = Teacher elicits but the question is implicit; R = Teacher recognizes; –S = Student initiates the cycle without a teacher intervention.

way, Ray's teaching was almost completely neutral to student contributions, as he followed them with the comments they made without intervention.

Diana. Diana asked for many more scientific explanations (17%) and revoiced students' contributions more than any other teacher (40%). Diana was also the teacher with the highest percentage of evaluative responses (22%); however, this strategy was rarely used to end a cycle. She used it in conjunction with other recognizing or using strategies. Furthermore, as mentioned, Diana was one of two teachers who used a more diverse array of recognizing and using strategies, although those strategies were embedded in a pattern that was repeated several times. This pattern is illustrated in the discussion that is excerpted in Table 3. In this example, taken from PS10, students have made a graph that compares mass and total volume of sinking and floating objects. Diana is attempting to get students to notice whether there are different patterns in the sinking and floating objects.

The example in Table 3 illustrates a pattern common to many of Diana's conversations. She poses a question, a student responds, and then Diana repeats or revoices the comment, elaborates on the comment, and provides some kind of evaluative response. Repeating this sequence over and over during a discussion creates a conversation characterized by a good deal of guidance on the part of the teacher, as she asks questions that are easily answered by students, then she praises the students and adds more information. Occasionally, Diana will push a student to think more about her response, or to provide more information, but then will return to the pattern. In this way, Diana controls the flow of conversation and clearly lets students know if they have the answer for which she is looking.

Adam. Of the four teachers, Adam most frequently asked students questions about the interpretation of data (30%). He also used revoicing as a recognizing strategy, but not as much as Diana or Carol (only 32%). However, he was the second highest in using "why" and "how" questions (75%). Adam's discussions were characterized by asking multiple questions without rephrasing until one of a handful of regularly participating students responded, at which point Adam would ask several follow-up questions of that same student.

During his discussions in PS4 and PS7, Adam created instances in which students were asked to interpret graphs that they did not understand. During PS4, Adam began asking students to make interpretations of two graphs—one of number of ball bearings in a straw versus depth of sinking, and one of mass versus depth of sinking—before students pointed out to him that they had not made the graph of mass versus depth of sinking. Adam hesitated for a moment and then decided to push forward with the conversation. Alternating the two different graphs on the overhead projector, Adam continued to ask students about the two graphs (see Table 4).

The excerpt in Table 4 illustrates how Adam did not exhibit many characteristics of effective formative assessment when leading this discussion. To begin with,

TABLE 3
Excerpt of Diana's Discussion at the End of PS10 (11/12/03)

<i>Speaker</i>	<i>Dialogue</i>	<i>Cycle</i>	<i>Code (Teacher ...)</i>
Diana	... If you look at my graph, I want you to take note of something. Generally, where do my floaters appear? Take a look at where all my floaters are. Where do they generally appear? Ben?	E	Interpret a graph, data, patterns, best fit lines, or other information
Ben	On the right side.	S	
Diana	On the right side of the graph. Good. What about my sinkers? Where do they generally appear? [Carla]?	R	Revoices students' words
		E	Provides evaluative response
			Interpret a graph, data, patterns, best fit lines, or other information
Carla	On the left.	S	
Diana	On the left side. Good. Generally, you should have the same thing. All of your floaters should be towards what Ben defined as the right of the graph, and all of your sinkers should be to the left of the graph. Even if all of your information is grouped down here, all of your floaters should still be below your sinkers. Why? We're going to read it now. Look carefully. Read your axes, then read the data. Why are the floaters more towards the right? Why are the sinkers more towards the left? Kevin?	R	Revoices students' words
			Provides evaluative response
			Clarifies/elaborates/provides examples based on students' responses
		E	Formulate scientific explanations, provide reasons, or determine a causal relationship
Kevin	They mass more.	S	
Diana	What masses more?	(R) U	Promotes students' thinking
Kevin	The sinkers.	(E) S	

Diana	Okay, the sinkers mass more. So, Kevin's saying the sinkers mass more. So, why does it make sense then, Kevin, that the sinkers are to the left? He just said they mass more. So, Roxie, where are they appearing on the graph?	R	Revoices students' words
Roxie	The higher they are, the more [inaudible audio].	U	Promotes students' thinking
Diana	Excellent! You got it, Roxie. Roxie is talking in her little mouse voice. But, she said the higher the line, the more mass that it has. She knows that, she said, because the mass is on the vertical axis. I didn't even ask this, but you're right. The closer your line is to the vertical axis, the more what it has?	(E) S	Provides evaluative response Revoices students' words
Student	Mass.	E	Interpret a graph, data, patterns, best fit lines, or other information
Diana	Mass. Roxie hit it right on; she's right. So, let's flip-flop it. What about the floaters? What do we know, then, about the floaters? [Nate]?	S	
		R	Repeats students' words
		E	Provides evaluative response
			Formulate scientific explanations, provide reasons, or determine a causal relationship
Nate	They have more volume than mass.	S	
Diana	They have more volume than mass. Excellent! So, all of these lines that we drew all represent the density of that object. We're taking mass and volume, and you told me earlier that connection between mass and volume is density. Each line represents the density of the object. The floaters should be closer to the horizontal axis because they have more volume than mass. The sinkers should be closer to the vertical [inaudible audio].	R	Repeats students' words Provides evaluative response Clarifies/elaborates/provides examples based on students' responses

Note. PS = Physical science; E = Teacher elicits; S = Student responses; R = Teacher recognizes; (R) = Teacher recognizing the student response/contribution; U = Teacher using information collected from student; (E) = Teacher elicits but the question is implicit.

TABLE 4
Excerpt of Adam's Discussion at the End of PS4 (10/21/03)

<i>Speaker</i>	<i>Dialogue</i>	<i>Cycle</i>	<i>Code (Teacher ...)</i>
Adam	What did we measure in this graph?	E	Asks students to respond with fill-in-the-blank answers
Jessica	The depth of the straw.	S	
Adam	Right here.	(R)	
Jessica	I don't know.		
Adam	What did we measure on this one? First, what did we measure on this one, then Jessica? What did we measure in this one?	U (E)	Promotes students' thinking
Jessica	The ... I don't know, I can't see the top.	S	
Adam	Who told me earlier? Wait! Jessica's going to tell us.	E	Asks students to respond with fill-in-the-blank answers
Jessica	I am?	S	
Adam	What did we measure in this graph, guys? We already said it.	E	Asks students to respond with fill-in-the-blank answers
Student	The number of BBs.	S	
Adam	The number of BBs in the straw. Derek, you told us we measured in this graph. What did we measure in this graph?	R	Revoices students' words
		E	Asks students to respond with fill-in-the blank answers
Student	Mass.	S	
Adam	Mass of straw and BBs. So [Jessica]—	E	Asks students to respond with fill-in-the-blank answers
Jessica	I knew about that graph. I just didn't know about that graph.	S	
Adam	What did we not measure in this one?	E	Asks students to respond with fill-in-the-blank answers
Jessica	What did we not measure in that one? See I already forgot what that graph ...	S	

Note. PS = Physical science; E = Teacher elicits; S = Student responses; (R) = Teacher recognizing the student response/contribution; U = Teacher using information collected from student; (E) = Teacher elicits but the question is implicit; R= Teacher recognizes.

he was asking students to interpret a graph that they had not made and seemed not to understand. Then he continually asked the same question over and over again without rephrasing it, even though the student he was questioning before the whole class did not understand the question. In this case, Adam often did not seem to recognize what the student had said, but asked the same question again with little change in his question. It is also important to point out that Adam asked very general questions about what had been measured in the graph, with no questions about interpretations of the graph.

Carol. An analysis of Carol's teaching revealed the presence of informal assessment practices during all discussions. She was the teacher who used the most diverse strategies across all the ESRU phases; for example, she was the teacher who focused the most on cycles linked to the conceptual domain. She frequently used the strategy of responding to a student's contribution by asking another student to repeat it. This strategy helped to make a key idea more explicit in the discussion at hand. Carol was also the teacher with the most student-initiated cycles. Only Carol acted on information accumulated over multiple ESRU cycles, suggesting that acting on accumulated information about student understanding is possible while discussions are still in progress. Carol was also the only teacher who frequently provided students with additional information to help them move forward in their thinking.

She held discussions that were longer and more frequent than those of the other teachers. Carol frequently began a discussion by reviewing what had been addressed the previous day, and then posed open-ended questions to the class. She would frequently follow the course of students' comments rather than pursuing her own narrative for the class. In the excerpt in Table 5, Carol is leading a discussion during PS7. Students have been sharing their interpretations of data contained in a graph of displaced volume versus mass of floating and sinking objects, and one student raised the point that it was easier for things to float in salt water as compared to fresh water. Carol walked to the counter near the front of the classroom and retrieved two beakers, one filled with tap water and the other with saltwater, and two hardboiled eggs.

The transcript excerpt illustrates a number of features of interest in Carol's teaching. First, she is sufficiently familiar with the unit, and can therefore anticipate that students will raise particular issues (e.g., saltwater vs. freshwater, air in the egg). Second, Carol uses students' comments as opportunities to pull out these materials and make visible the students' own meanings; that is, to take action on their ideas by showing them immediately what they are thinking. This type of action shows the level of instructional responsiveness in Carol's teaching, where she leads a conversation by following students instead of controlling the course of the discussion. Third, Carol is constantly shaping the students' understandings, by revoicing their words and asking follow-up questions to promote their thinking

TABLE 5
Excerpt of Carol's Discussion at the End of PS7 (4/23/04)

Speaker	Dialogue	Cycle	Code (Teacher ...)
Carol	Um, do you mean like this? You said that there's ... it's easier for things to float in salt water?	R U	Provides neutral response Puts meaning into matter Acts on information accumulated in multiple cycles
	<i>(Teacher drops an egg into a beaker of saltwater and another egg into a beaker of fresh water. The egg in the salt water floats and the egg in the fresh water sinks)</i>		
Student	That's [amazing]		
Carol	Do you think the eggs are different kinds of eggs, and we should just use one?	E	Asks students to compare/contrast data, procedures, or observations
Student	One looks way bigger than the other.		
Student	Yeah.	S	
Student	It's 'cause of the [inaudible audio]	S	
Student	Maybe one's hardboiled.		
Carol	We just probably should use one egg to be sure that this is not a trick, right? Yeah, you wouldn't want me doing a trick on you. So I'll take this egg out of the fresh water, which is right here, and I'll put the egg in the salt water. Same egg. Remember, where did I take this egg from?	R U E	Clarifies/elaborates/provides examples based on students' responses Puts meaning into matter Asks students to answer fill-in-the-blank question
Students	Fresh water.	S	
Carol	Bottom of the fresh one. Same egg. Whoops! <i>(Teacher takes egg from fresh water beaker, drops into beaker filled with salt water. The egg floats)</i>	R	Revoices students' words
Students	No, uh-uh. [multiple voices, unintelligible]	-S	
Carol	Hey, the egg had air in it in here. <i>(Teacher points out at the egg)</i>	R U	Clarifies/elaborates/provides examples based on students' responses Promotes students' thinking
Student	Yeah, but ...		
Carol	But?		

Student	But, but ...		
Carol	But what? Yes?	(E)	
Student	My uncle went to Hawaii for his wedding and his honeymoon, and he said that when you lay down in the salt water the water is so salty there that you just float. Pretty much [inaudible audio].	S	
Carol	So the salt water's different. And was he saying that it's easier to float in?	R	Clarifies/elaborates/provides examples based on students' responses Responds with fill-in-the-blank question
Student	Yes.		
Carol	How come?	U	Promotes students' thinking
Student	It's way saltier than ...	(E) S	
Carol	What does salt have to do with it?	(R) U	Promotes students' thinking
Student	Salt makes it more dense.	(E) S	
Carol	What does that mean? Salt makes it more dense? What's that mean?	(R) U	Promotes students' thinking
Student	Thicker.	(E) S	
Carol	Thick is kind of, not a science word but, it's a description. Thicker. Let's go back to what Sandy was saying. She said something about, it has more what in it?	R	Clarifies/elaborates/provides examples based on students' responses Provides evaluative response
		U	Compares/contrasts students' responses
		E	Asks fill-in-the-blank question
Student	Matter.	S	
Carol	Matter of the universe in it. And I used the same amount?	R	Revoices students' words
	Maybe that has something to do with it.	U	Promotes students' thinking Models process skills

Note. PS = Physical science; R = Teacher recognizes; U = Teacher using information collected from student; E = Teacher elicits; S = Student responses; -S = Student initiates the cycle without a teacher intervention; (E) = Teacher elicits but the question is implicit; (R) = Teacher recognizing the student response/contribution.

(e.g., by pushing them to articulate what they mean by saltwater being thicker, and to articulate what they mean in scientific terms).

The ESRU framework for assessment conversations identified differences between teachers' informal formative assessment practices. It suggests that Carol and Diana conducted more assessment conversations than Adam or Ray. It highlights the requests for explanations and evaluative responses that Diana made more often than the other teachers. Furthermore, it illustrates how Carol, unlike the other teachers, emphasized concepts and took action on information collected across multiple cycles. It is important to note that these differences in practices do not map cleanly onto common explanations of differences in teachers' practices. For example, the two teachers who had undergraduate degrees in science (Adam and Ray) showed fewer ESRU cycles than the two teachers who did not (Diana and Carol); Ray and Carol were more experienced teachers than Adam and Diana, yet Diana's and Carol's informal formative assessment practices were more similar than Adam's and Ray's.

Student Performance

In this section, we link the teachers' informal assessment practices with student performance. We asked three questions to approach the linking: (a) Does students' average performance differ by teacher on the pretest? (b) Do students' average gain scores from pretest to posttest differ by teacher? and (c) Do students' average gain scores on the posttest controlling for initial status (pretest) differ by teacher? Table 6 provides the information by group in the pre- and posttest. The 38-item multiple-choice test (average $\alpha = .86$; Yin, 2005) was used to assess whether students across the four groups were similar in their knowledge about relative density before Investigation 1 (PS1) and after the 12 investigations were taught.

We acknowledge the nested nature of the data collected for this study. Therefore, to answer the first question, we tested differences between the four groups with a one-way ANOVA and calculated an intraclass correlation coefficient to ex-

TABLE 6
Pre- and Posttest Multiple-Choice Test Means
and Standard Deviations by Teacher

<i>Teacher</i>	<i>Pretest (Max = 38)</i>			<i>Posttest (Max = 38)</i>		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Ray	26	16.00	4.58	22	18.27	6.72
Diana	25	15.52	3.61	23	22.83	4.06
Adam	19	14.55	5.08	28	19.89	6.33
Carol	29	13.16	4.49	17	25.64	5.44

plore independence of the observations (i.e., the homogeneity of the observations within teachers, relative to observations between teachers). The one-way ANOVA indicated no significant differences between the four groups, $F_{(3, 95)} = 1.68$, $p = .176$, and the magnitude of the intraclass correlation coefficient ($\rho = .027$) indicated that the variability in student performance accounted for by teacher was very low in the pretest. We concluded that, on average, students within each of the groups participating in the study were similar in their understanding level of relative density before the first FAST investigation was implemented, and that students' performance at the beginning was not associated with teacher.

To respond to the other two questions, we used a general linear model approach to analyze gain scores. We estimated two successive models; one with teacher fixed effects only and another with teacher fixed effects and pretest scores (see Table 7). The first model showed significant differences on the average gain scores between teachers, $F_{(3, 86)} = 17.91$; $p < .001$; $R^2 = .385$. The second model, which included students' initial status (pretest), suggested that gain scores were not related to the pretest scores, $F_{(1, 85)} = 1.62$; $p = .205$, but were still dependent on teachers, $F_{(3, 85)} = 16.23$; $p < .001$; $R^2 = .396$. Therefore, the teacher appears to be a better predictor of the gain scores than initial status of the student. It is important to point out that the gain scores are directly related to the regressions coefficients. The highest gain scores were observed for Carol (M gain score = 12.70), followed by Diana (M gain score = 8.93), Adam (M gain score = 5.79), and Ray (M gain score = 2.51). Figure 1 provides information on both pre- and posttest (Panel A) and gain scores (Panel B) across teachers.

TABLE 7
Regression Results by Model

	<i>Student Performance Model</i>			
	<i>Model 1: Teacher</i>		<i>Model 2: Teacher and Pretest</i>	
	<i>Coefficient</i>	<i>SE</i>	<i>Coefficient</i>	<i>SE</i>
Intercept	12.706*	1.073	14.504*	1.769
Ray	-10.251*	1.429	-9.852*	1.458
Diana	-5.488*	1.415	-5.118*	1.440
Adam	-7.170*	1.361	-6.973*	1.364
Carol	0.0 ^a	—	0.0 ^a	—
Pretest			.109	
R^2	.385	.396		
R^2 adjusted	.363	.368		
F -change	17.91*	1.629		

^aParameter set to zero because it is redundant.

* $p < .005$.

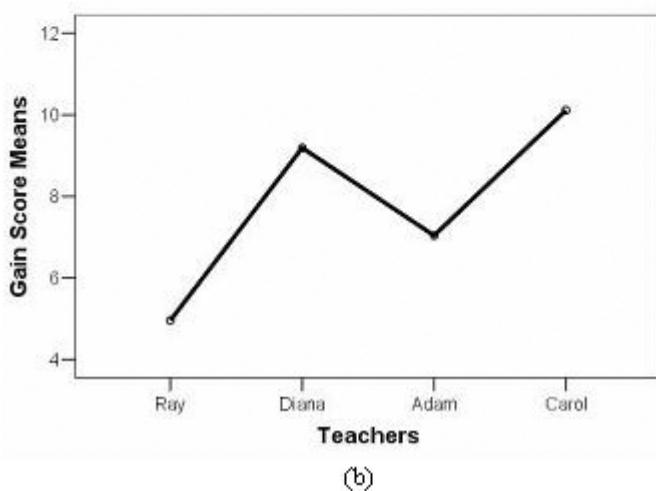
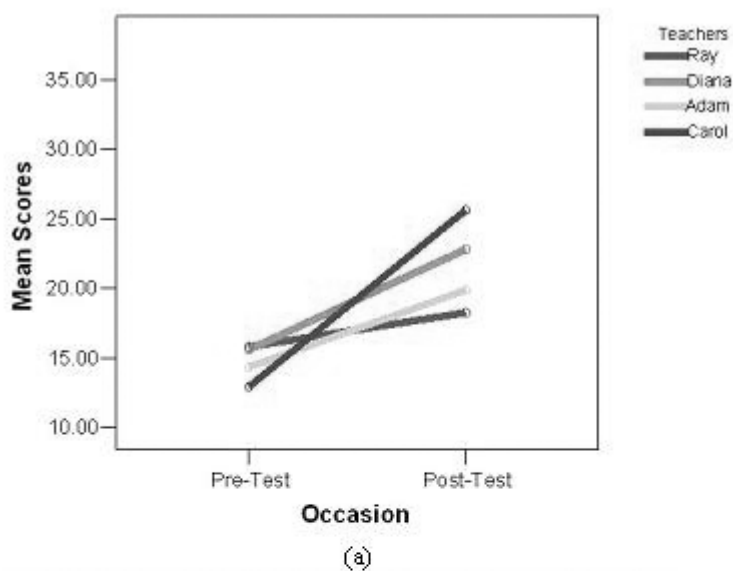


FIGURE 1 Students' performance in the pretest and posttest (a) and gain scores (b) by teacher.

This pattern was consistent with students' performance on the two embedded assessments across the three occasions in which the embedded assessments were implemented (PS4, PS7, and PS10). Mean scores across type of assessments and occasions are provided in Appendix B. Figure 2 shows the mean score in the graphing prompt. In this embedded assessment, we found that the proportion of variance accounted for by teacher increased at each occasion: $R^2_{(EA4)} = .188$; $R^2_{(EA7)} = .257$; and $R^2_{(EA10)} = .341$.

Figure 3 shows the mean score in the POE embedded assessment. In this prompt, a similar but not exact pattern was observed; the proportion of variance accounted for by teacher tended to increase from the first embedded assessment at PS4 to the last one at PS10: $R^2_{(EA4)} = .205$; $R^2_{(EA7)} = .194$; and $R^2_{(EA10)} = .244$. Across both embedded assessments, Carol's and Diana's students consistently showed a higher performance than Adam's and Ray's. Although there was no information from Ray's students for the POE after Investigation 4 (EA4), his students consistently showed the lowest mean performance in all the assessments.

Without venturing to make a causal relation, we believe that the evidence provided suggests that the students' mean performance is consistent with the levels of the quality of the informal formative assessment practices observed across the four teachers analyzed. Although we acknowledge that other factors could influence the students' observed performance, we believe that one of those possible factors is the type and the quality of assessment conversations in which these teachers were engaged with their students. However, we also acknowledge that opportunities for

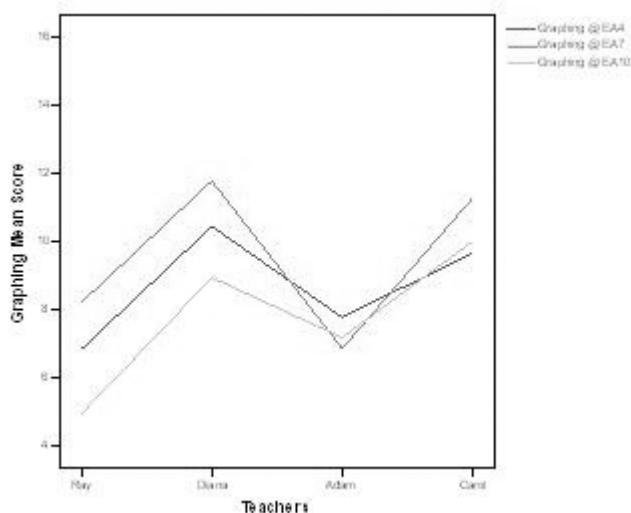


FIGURE 2 Students' performance on the graphing embedded assessment by teacher and investigation.

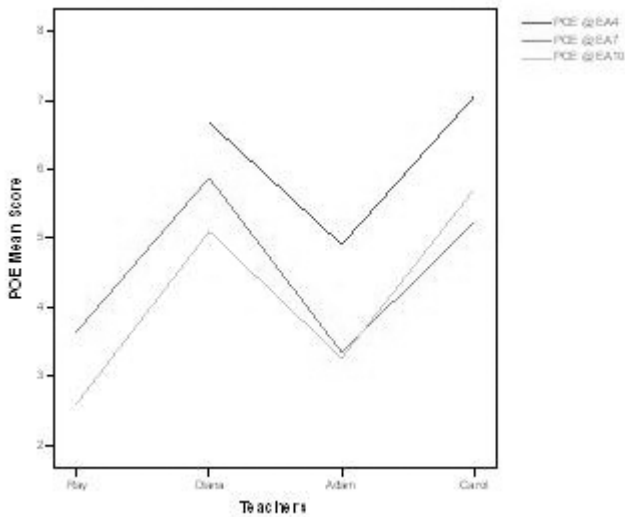


FIGURE 3 Students' performance on the POE embedded assessment by teacher and investigation.

learning were lost in each classroom. In fact, it is important to notice that students' performance across the assessments was far from reaching the maximum possible scores, indicating that students still had a lot to learn. For example, a closer look at the embedded assessments indicated that students were better at plotting the data points in the graphs than interpreting them. In addition, although students were able to provide accurate predictions, they could not accurately justify them or provide an accurate explanation of what they observed. In fact, the mean score decreased at each embedded assessment, perhaps due to the fact that an appropriate explanation at PS4 becomes inappropriate at PS10.

CONCLUSIONS

In this article, we propose a framework to study informal formative assessment practices based on the formative assessment components (i.e., gathering, interpreting, and acting) and science inquiry domains (i.e., epistemic, conceptual, and social). The approach was developed to capture, during everyday teacher–student interactions, the different ways that teachers collect information about students' understanding on a continuing and informal basis. Using this framework, we studied the informal assessment practices of four teachers during the implementation of four science investigations and tracked their students' learning during these investigations with diverse embedded assessments.

The results indicate that the informal formative assessment practices of the four teachers were considerably different from each other. Although student performance on the pretest was not significantly different between the classrooms, as the unit progressed differences began to emerge. These differences followed a pattern similar to what was observed in the teachers' informal formative assessment practices: Those teachers whose assessment conversations were more consistent with the ESRU cycle had students with higher performance. This difference was also reflected on the posttest. The teacher whose students had the highest performance on our tests was the teacher who held the most discussions, asked the most concept-eliciting questions, and employed the greatest diversity of strategies that used information she had gained about student understanding.

These findings help provide evidence about the distinction between our ESRU cycles and the IRE/IRF models of discussion by showing the relevance of the final step in the cycle, using. We have differentiated the IRE/IRF (Cazden, 2001) and the ESRU model based on three aspects: (a) the characteristics of the eliciting strategies that initiate the cycle (e.g., questions that can help to make explicit the student's level of understanding); (b) the role of recognizing the student's response (e.g., to acknowledge or empower the student's contribution in the construction of knowledge); and (c) the nature of using, which is more specific than the meaning assigned to feedback in the IRF sequence.

We acknowledge that the number of teachers is small, but lessons were sampled from each teacher over time, a necessary tradeoff when coding transcripts in great detail. We also acknowledge that our coding system does not completely capture the diverse quality of the eliciting questions. For example, we combined under the same category—interpretation of graph, data, and patterns—any eliciting question related to the interpretation of graphs and tables. This methodological decision did not allow us to provide information on the exact types of questions teachers asked when guiding students in the interpretation of data; neither did it yield clues about how teachers help students view data as evidence. Whereas some teachers focused on asking students for one data point, others focused on asking students to identify patterns. We think that these distinctions are important; however, in the end, they were not distinguished in our analysis. Some of the categories under the epistemological domain should be considered under a different domain more related to the procedural or methodological aspects of conducting the investigation. Future studies will revise the scientific inquiry domains.

The evidence we have provided about the teachers' informal assessment practices should be considered in designing assessment courses for student teachers and for professional development. It is important to provide teachers with simple tools that can help them gather information about students' learning in ways that allow them to respond in an immediate and effective way. The data we have presented provide evidence that it is possible for teachers to collect information about students' understanding during their everyday interactions, and we have proven

that high-quality informal assessment practices can be linked to increases in students' performance.

Future research should build on the methodology and evidence presented in this study by using the ESRU coding technique to explore discussions that took place in the context of the EAs. Further interpretation of complete and incomplete ESRU cycles, in association with information about the quality of student responses, could supply valuable information about the characteristics of scientific inquiry taking place in these classrooms. Furthermore, exploring the different conceptual levels addressed by teachers and students during discussions throughout the investigations, as well as in the embedded assessments, will further illuminate the distribution of student conceptions on the diverse assessments involved in this study.

ACKNOWLEDGMENTS

Funding for this research reported in this article was provided by the National Center for Research on Evaluation, Standards, and Student Testing/Award 0070 G CC908-A-10 and National Science Foundation/Award ESI-0095520.

REFERENCES

- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- Bell, B. (2000). Formative assessment and science education: A model and theorizing. In R. Millar, J. Leach, & J. Osborne (Eds.), *Improving science education. The contribution of research* (pp. 48–61). Buckingham, UK: Open University Press.
- Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht, The Netherlands: Kluwer.
- Black, P., & Harrison, C. (2001, April). *The science teacher's role in formative assessment*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7–74.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Carlsen, W. S. (1991). Questioning in classrooms: A sociolinguistic perspective. *Review of Educational Research*, 61, 157–178.
- Cazden, C. B. (2001). *Classroom discourse. The language of teaching and learning* (2nd ed.). Portsmouth, NH: Heinemann.
- Cocking, R. R., Mestre, J., & Brown, A. L. (2000). New developments in the science of learning: Using research to help students learn science and mathematics. *Journal of Applied Developmental Psychology*, 21, 1–11.
- Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham, UK: Open University Press.
- Duschl, R. (2000). Making the nature of science explicit. In R. Millar, J. Leach, & J. Osborne (Eds.), *Improving science education. The contribution of research* (pp. 185–206). Buckingham, UK: Open University Press.

- Duschl, R. A. (2003). Assessment of inquiry. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41–59). Arlington, VA: National Science Teachers Association Press.
- Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, 4, 37–73.
- Ebenezer, J. V., & Haggerty, S. M. (1999). *Becoming a secondary science teacher*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London: Routledge.
- King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31, 338–368.
- Lemke, J. L. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex.
- Minstrell, J., & vanZee, E. (2003). Using questioning to assess and foster student thinking. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 61–73). Arlington, VA: National Science Teachers Association Press.
- National Research Council. (1996). *The national science education standards*. Washington, DC: National Academy Press.
- National Research Council. (2001). *Inquiry and the national science education standards*. Washington, DC: National Academy Press.
- Pottenger, F., & Young, D. (1992). *The local environment: FAST 1 foundational approaches in science teaching*. Honolulu: University of Hawaii Manoa, Curriculum Research and Development Group.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4–13.
- Rogg, S., & Kahle, J. B. (1997). *Middle level standards-based inventory*. Oxford, OH: Miami University.
- Roth, W.-M. (1996). Teacher questioning in an open-inquiry learning environment: Interactions of context, content, and student response. *Journal of Research in Science Teaching*, 33, 709–736.
- Ruiz-Primo, M. A., & Furtak, E. M. (2004, April). *Informal assessment of students' understanding of scientific inquiry*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Ruiz-Primo, M. A., & Furtak, E. M. (in press). Exploring teachers' informal formative assessment and students' understanding in the context of scientific inquiry. *Journal of Research in Science*.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education*, 5, 77–84.
- Scott, P. (2004). Teacher talk and meaning making in science classrooms: A Vygotskian analysis and review. In J. Gilbert (Ed.), *The RoutledgeFalmer reader in science education* (pp. 74–96). London: RoutledgeFalmer.
- Shavelson, R. J., Black, P., Wilam, D., & Coffey, J. (2003). On aligning summative and formative functions in the design of large-scale assessment systems. *Educational Evaluation and Policy Analyses*.
- Shavelson, R. J., & Young, D. (2000). *Embedding assessments in the FAST curriculum: On the beginning the romance among curriculum, teaching and assessment*. Unpublished manuscript, Stanford University, Stanford, CA.
- Smith, E. L., Blakeslee, T. D., & Anderson, C. W. (1993). Teaching strategies associated with conceptual change learning in science. *Journal of Research in Science Teaching*, 30, 111–126.
- Stipek, D. J., Salmon, J. M., & Givvin, K. B. (1998). The value (and convergence) of practices suggested by motivation research and promoted by mathematics education reformers. *Journal for Research in Mathematics Education*, 29, 465–488.
- Tobin, K. G., Tippins, D. J., & Gallard, A. J. (1994). Research on instructional strategies for teaching science. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 45–93). New York: Macmillan.

- Welch, W. W., Klopfer, L. E., Aikenhead, G. S., & Robinson, J. T. (1981). The role of inquiry in science education: Analysis and recommendations. *Science Education*, 65, 33–50.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16, 3–118.
- Yin, Y. (2005). *The influence of formative assessments on student motivation, achievement, and conceptual change*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

APPENDIX A
Percentage of Used Strategies by Cycle Phase

<i>Strategies by Phase</i>	<i>Ray</i>	<i>Diana</i>	<i>Adam</i>	<i>Carol</i>	<i>Total</i>
Eliciting—Epistemic					
Compare and contrast observations, data, or procedures	0	3	2	2	2
Use and apply known procedures	0	0	12	1	4
Provide observations	16	21	3	23	11
Make predictions or provide hypotheses	60	4	4	0	6
Formulate scientific explanations	0	17	9	7	10
Provide evidence and example	4	3	0	7	3
Interpret information, data, patterns	0	20	30	17	24
Relate evidence and explanations	0	0	0	3	1
Evaluate the quality of evidence	0	4	0	3	2
Compare or contrast others' ideas, explanations	0	1	0	0	0
Suggest hypothetical procedures or experimental plans	0	0	0	1	0
Promote students' thinking	0	1	1	2	1
Questions to check students' comprehension	4	4	9	1	6
Other epistemic	8	7	20	19	15
Eliciting—Conceptual					
Provide potential or actual definitions	0	11	6	3	5
Apply, relate, compare, contrast concepts	0	1	3	3	4
Compare or contrast others' definitions	0	4	0	1	3
Promote students' thinking	0	0	0	1	0
Questions to check students' comprehension	8	0	0	0	1
Other conceptual	0	0	0	2	1
Recognizing					
Clarifies/elaborates on student response	8	17	11	19	17
Takes votes to acknowledge different students' ideas	21	1	1	3	3
Repeats/rephrases student's contribution	30	15	16	12	15
Revoices student's words	20	40	32	39	35
Captures/displays students' diverse explanations	0	1	0	1	1
Provides evaluative responses	3	22	9	7	12
Responds with a yes/no, fill-in-the-answer question	0	1	1	7	3

(continued)

APPENDIX A (Continued)

<i>Strategies by Phase</i>	<i>Ray</i>	<i>Diana</i>	<i>Adam</i>	<i>Carol</i>	<i>Total</i>
Neutral response	9	3	9	7	6
Responds with a minilecture not connected to the student's response	9	0	12	3	5
Asks to repeat what other student said	0	0	1	1	1
Teacher answer her/his own question	0	0	7	1	2
Using					
Promotes student thinking with why/how question	100	65	75	55	62
Compares/contrasts students' alternative explanations	0	11	2	5	4
Promotes consensus/come to an agreement	0	0	0	1	0
Helps relate evidence to explanations	0	1	0	0	2
Provides helpful feedback	0	8	0	2	4
Models process skills	0	0	0	2	0
Makes learning goals, expectations, standards explicit	0	3	4	2	6
Connects to previous learning, investigations, conclusions	0	0	4	2	1
Refers to nature of science	0	3	0	4	1
Solicits and encourages diversity of opinions	0	5	2	6	4
Promotes exploration of student's ideas	0	3	2	3	4
Explores student's ideas—meaning into matter	0	0	2	8	2
Provides students with additional information (scaffolding)	0	3	8	11	9
Acts on information accumulated in multiple ESRU cycles	0	0	0	1	0

APPENDIX B
Mean and Standard Deviation for Each Embedded
Assessment by Teacher

	<i>Max.</i>	<i>Ray</i>			<i>Diana</i>			<i>Adam</i>			<i>Carol</i>		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
PS4													
Graphing	16	23	6.82	3.25	23	10.34	3.03	26	7.34	3.65	15	9.66	2.41
POE	8	—	—	—	25	6.68	2.09	22	4.91	1.71	18	7.06	1.69
PS7													
Graphing	16	21	8.33	4.94	24	11.83	3.06	25	6.88	3.49	18	11.56	3.27
POE	8	17	3.65	2.20	24	5.82	2.32	23	3.35	2.01	18	5.22	2.57
PS10													
Graphing	13	24	4.96	2.01	21	9.19	2.11	24	7.04	3.35	18	10.11	3.57
POE	8	20	2.60	2.08	20	5.10	2.26	23	3.26	2.22	17	5.71	2.46

Note. PS = FAST Physical Science Investigation. Graphing = graphing assessment. POE = Predict–Observe–Explain assessment. — indicates that no data are available.