Informal Formative Assessment of Students' Understanding of Scientific Inquiry

CSE Report 639

Maria Araceli Ruiz-Primo & Erin Marie Furtak
School of Education, Stanford University/CRESST

August 2004

Center for the Study of Evaluation (CSE)
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 2.2: Classroom and Teachers' Assessment. Strand 2: Curriculum Embedded Assessments.
Richard Shavelson and Maria Ruiz-Primo, Project Directors, Stanford/CRESST

The work reported herein was supported, in part, by the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

INFORMAL FORMATIVE ASSESSMENT OF STUDENTS' UNDERSTANDING OF SCIENTIFIC INQUIRY

Maria Araceli Ruiz-Primo & Erin Marie Furtak

Stanford University/CRESST

Abstract

This paper provides information on an exploratory study about informal formative assessment practices in three science classrooms implementing a physical science curriculum focusing in buoyancy. We provide a framework for examining these practices based on three components of formative assessment (eliciting, recognizing and using information) and three domains linked to science inquiry (epistemic frameworks, conceptual structures, and social processes). We developed a coding system to track strategies teachers used across the three informal formative assessments components. The coding system could capture differences in assessment practices across the three teachers. Furthermore, based on three questions used for assessing students' performance we linked students' level to the quality of the teachers' assessment practices. We concluded that the strategies provided important information about the teacher informal assessment practices.

The success of science-education reform relies on the alignment of the three central elements to any educational enterprise: curriculum, instruction, and assessment. These three elements should function toward the same ends and reinforce each other rather than work towards different purposes. Ideally, an assessment should measure what students are actually being taught, and what is actually being taught should parallel the curriculum one wants students to learn (Pellegrino, Chudowsky, & Glaser, 2001).

In this paper we focus on the third element, assessment. More specifically, we focus on the *informal formative assessment practices* that teachers implement in their classroom to gather information about students' understanding. The rationale is that formative assessment is essential to good teaching and learning (Black, 1993). If teachers do not monitor students' understanding, their efforts to help students to improve their learning are limited.

In this paper we provide information on the informal assessment practices of three teachers implementing the four first lessons of an inquiry-based physical science curriculum on buoyancy. We link their informal assessment practices with student performance on three assessment tasks implemented as formal embedded assessments. In what follows, we first present a framework for analyzing informal assessment practices in the context of science inquiry. Then we describe the means by which we collected the information, and finally, we present our findings.

On Formative Assessment

Formative assessment involves gathering, interpreting, and acting on information about students' learning so that it may be improved (Bell & Cowie, 2001). That is, formative assessment involves gathering and interpreting information to be used as feedback to modify teaching and learning activities in order to reduce the gap between desired student performance and observed student performance (Bell & Cowie, 2001; Black & Wiliam, 1998; Shavelson, Black, Wiliam, & Coffey, 2003). Notice that this definition considers formative assessment as encompassing all those activities undertaken by both teachers and students (Black & Wiliam, 1998). Teachers can use the information to monitor their teaching and students to monitor their own learning (NRC, 1999). The process of formative assessment also includes students since they need to recognize, evaluate, and react to their own learning and/or others' assessments of their learning (Bell & Cowie, 2001; Zellenmayer, 1989).[1]

Formative assessment can be *formal*—a planned act designed to provide evidence about students' learning, or *informal*—where evidence of learning is generated in the course of a teacher's day-to-day activities (Bell & Cowie, 2001; Duschl, 2003; Shavelson et al., 2003). Each type can be characterized in a different manner. Formal formative assessment focuses on obtaining information from the whole class. It usually starts with students doing/carrying out an activity designed or selected by the teacher so that information may be more precisely collected (gathering). The activity enables teachers to step back at key points during instruction, check student understanding (interpreting), and reflect on the next steps

[1] The focus of this paper is confined to the teachers' actions during the process of formative assessment. The reason behind this decision is more practical than conceptual. Since information was collected on videotape, requiring the teachers to use a microphone, students' comments and questions were not always captured.

they must take to move forward their students' learning (acting). Commonly, formal formative assessment practices come in the form of curriculum embedded assessments that focus on some specific aspect of learning (e.g., students' knowledge about why objects sink or float), but it can also take the form of direct questioning, quizzes, brainstorming, generation of questions, and the like (Bell & Cowie, 2001). Teachers plan in advance the implementation of this type of assessment and it can take place at the beginning, during, or end of a unit.

Informal formative assessment can take place in any student-teacher interaction. It has the potential to occur at any time and can involve whole class, small group or one-on-one interactions. It can arise out of any instructional/learning activity at hand (gathering), and it is "embedded and strongly linked to learning and teaching activities" (Bell & Cowie, 2001, p. 86). Although said not to be planned because it can happen at any time and there is no specific activity designed for students, it is still possible for teachers to prepare in advance for this type of formative assessment. Certainly, teachers cannot predict exactly when they will be able to gather evidence about students' understanding in the course of non-planned activities, but they can make available diverse opportunities for doing so (e.g., by creating more interactions in class, group discussions, or informal observations). The information gathered during informal formative assessment is transient (Bell & Cowie, 2001; e.g., students' comments, responses, questions) and many times goes unrecorded. It can also be verbal (students' questions) or non-verbal (based on teacher's observations). The time frame for interpreting and acting is more immediate when compared with formal formative assessments. A student's incorrect response or unexpected question can trigger an assessment event by making a teacher aware of a student's misunderstanding. Acting in response to the evidence found is usually quick, spontaneous, and flexible since it can take different forms (e.g., responding with a question, eliciting other points of view from other students, conducting a demonstration when appropriate, repeating an activity).

Both formal and informal formative assessments involve gathering, interpreting, and acting on information. They differ in terms of how much planning is done and the type of planning required. Teachers' content knowledge and pedagogical content knowledge are important factors that will determine the interpretation and the acting aspects of the process. Figure 1 provides a schematic

representation of the two types of formative assessment and their relationship. We distinguish between the processes involved in formal and informal formative assessment by using different words to characterize them: *gathering*, *interpreting*, and *acting* for formal formative assessments; and *eliciting*, *recognizing*, and *using* for informal formative assessments. Other authors have used different names to describe the same processes (Bell & Cowie, 2001; Duschl, 2003). The black boxes between units in Figure 1 represent specific points in the curriculum in which the formal formative assessments are implemented. Both formal and informal formative assessments are connected through the general purpose for formative assessment. The continuous line between the units and informal formative assessment is intended to indicate the continuous nature of this type of assessment.

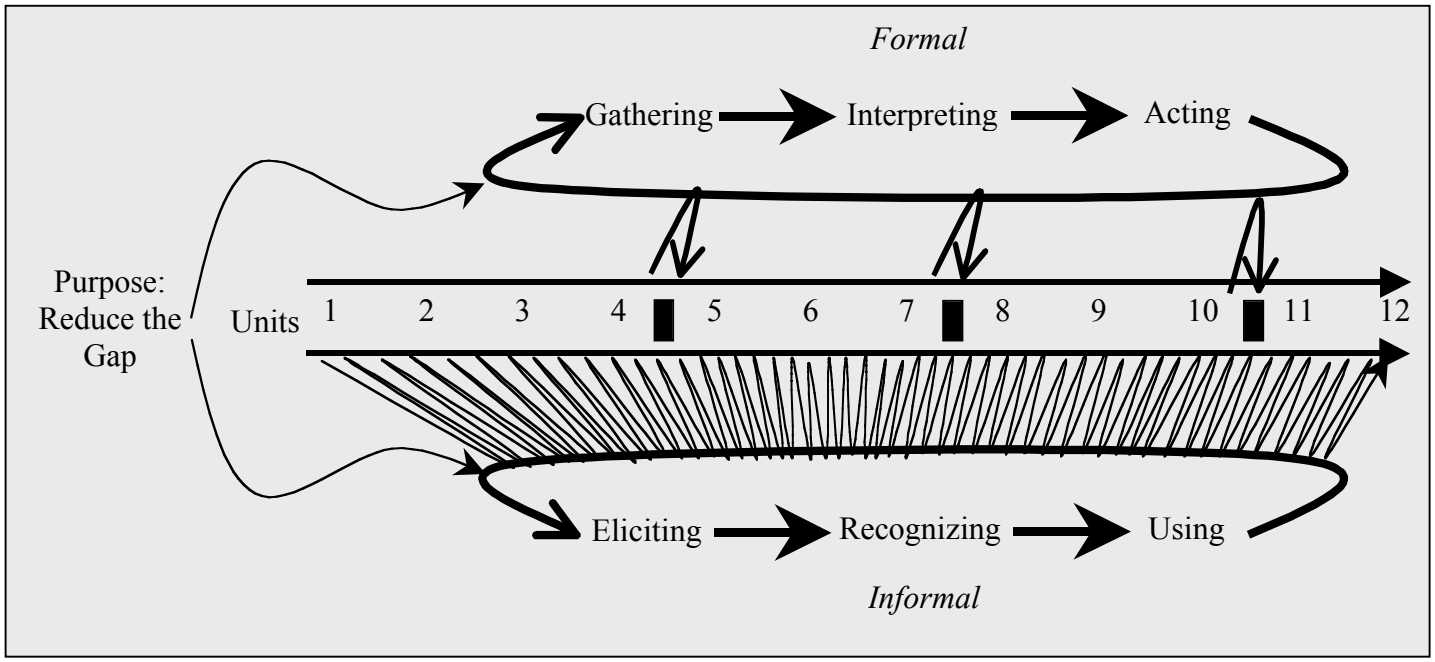

Figure 1. Graphical representation of formal and informal formative assessment.

Informal Formative Assessment in the Context of Science Inquiry.

A goal of the present science education reform is the development of thinking, reasoning, and problem solving skills to prepare students in the development and evaluation of scientific knowledge claims, explanations, models, scientific questions, and experimental designs (Duschl & Gitomer, 1997). The National Research Council

(2001) describes these habits of mind as fundamental abilities for students as they engage in the learning of science through the process of scientific inquiry.

Frequent and ongoing assessment activities in the classroom can help to achieve these habits of mind by providing information on the students' progress towards the goal and helping teachers to incorporate this information to guide their instruction (Black & Wiliam, 1998; Duschl & Gitomer, 1997; Duschl, 2003). Ongoing formative assessment is more likely to occur in a classroom learning environment that helps teachers acquire information on a continuing and informal basis.

Effective questioning and posing of challenges by teachers have been considered as a natural strategy to assess students through informal conversations. This type of classroom conversation has been termed *assessment conversations* (Duschl and Gitomer, 1997; Duschl, 2003), or an instructional dialogue that embeds assessment into an activity already occurring in the classroom. Assessment conversations permit teachers to recognize students' conceptions, mental models, strategies, language use, or communication skills and allow them to use this information to guide instructional activity. In classroom learning environments in which assessment conversations take place, the boundaries of curriculum, instruction, and assessment should blur (Duschl & Gitomer, 1997). For example, an instructional activity suggested by a curriculum, such as a discussion of the results of an investigation, can be used as an assessment conversation to find out about how students evaluate the quality of evidence and how they use evidence in explanations.

Assessment conversations have the three characteristics of informal assessment previously described: *eliciting*, *recognizing*, and *using* information. Eliciting information requires the use of strategies that allow students to share and make visible or explicit their understanding as completely as possible (e.g., sharing their thinking to the class, overheads, posters). Recognizing students' thinking requires the teacher to make a judgment about the differences among students' responses, explanations, or mental models so that the critical dimensions relevant for their learning can be made explicit (e.g., teacher compares students' responses according to the evidence provided or responds to students by asking which explanation is more scientifically accepted based on the information provided). Using information from assessment conversations implies mainly helping students to achieve a

consensus based on scientific reasoning (e.g., universality of the explanations). The most appropriate and scientifically-based consensus is not necessarily achieved in every assessment conversation. The range of student conceptions at different points of a unit determines the nature of the conversation. Therefore, more than one iteration of the informal formative assessment cycle may be needed to reach a consensus that reflects the most complete and appropriate understanding. The most known use of the information in formative assessment would be, of course, to provide helpful feedback that clearly helps students to improve their learning/performance.

Assessment conversations require teachers to be facilitators and mediators of learning, more than providing students with the correct and acceptable answer. In sum, successful classrooms emphasize not only the management of actions, materials, and behavior, but also stress the management of reasoning, ideas, and communication (Duschl & Gitomer, 1997).

Assessment conversations in the context of science inquiry should focus on three integrated domains (Duschl, 2003): *epistemic frameworks* for developing and evaluating scientific reasoning; *conceptual structures* used when reasoning scientifically; and *social processes* that focus on how knowledge is communicated, represented, and argued. Although Duschl (2003) considers *cognitive processes* as part of the first domain, we believe that *cognitive processes* are involved in all of the three domains. Epistemic structures are the knowledge frameworks that involve the rules and criteria used to develop and/or judge what counts as scientific (e.g., experiments, hypotheses, or explanations). Epistemic frameworks emphasize not only the abilities involved in the processes of science (e.g., observing, hypothesizing, experimenting, using evidence, logic, and knowledge to construct explanations), but also the development of the criteria to make judgments about the products of inquiry (e.g., explanations or any other scientific information). When students are asked to develop or evaluate an experiment or an explanation, we expect students to use epistemic frameworks. 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 that allows students to use that knowledge in an effective manner in appropriate situations. Social processes refer to the frameworks involved in students' scientific

communications needed while engaging in scientific inquiry. They can be oral, written, or pictorial. It involves the syntactic and semantic structures of scientific knowledge claims, its accurate presentation and representation, and the use of diverse forms of discourse and argumentation (e.g., read and write the various genres in science such as lab reports; Lemke, 1990; Marks & Mousley, 1990; Martin, 1989, 1993).

Is sum, a key characteristic of effective informal formative assessment is to promote activities in which frequent assessment conversations may allow teachers to *listen to inquiry* (Duschl, 2003). Listening to inquiry should focus on helping students "examine how scientists have come to know what they believe to be scientific knowledge and why they believe this knowledge over other competing knowledge claims" (Duschl, 2003, p 53). Therefore, informal formative assessments that facilitate listening to inquiry should: (1) involve discussions in which students share their thinking, beliefs, ideas, and products (*eliciting*); (2) allow teachers to acknowledge student participation (*recognizing*), and; (3) allow teachers to use students' participation as the springboard to develop questions and/or activities that can promote their learning (*using*).

Looking closely at informal assessment strategies. How can assessment conversations be identified? What operational criteria can be used for recognizing and or guiding assessment of students in the context of science inquiry? Following the characteristics of informal formative assessment and the three dimensions previously described, we developed some strategies for guiding the identification of the different aspects of assessment conversations. The strategies are intended to represent what we and others (Duschl, 2003; Duschl & Gitomer, 1997; National Research Council, 2001) have considered relevant aspects of informal formative assessment practices in the context of inquiry. Table 1 provides the strategies for recognizing and guiding assessment conversations. Strategies are organized by the informal formative assessment characteristics (i.e., eliciting, recognizing, and using information) and the three domains (i.e., epistemic frameworks, conceptual structures, and social processes). The strategies reflect the questions that teachers may ask students to elicit information and the teacher actions that may reflect the recognition and use of information. Due to the integral nature of characteristics and domains, some strategies are repeated across the two dimensions of the table.

Table 1.

Strategies for Recognizing and Guiding Assessment Conversations by Dimension and Informal Formative Assessment Components.

	Eliciting	Recognizing	Using
Epistemic Frameworks	*Teacher asks students to:* Apply procedures involved in science Provide responses not based on observations Share/Provide observations Make predictions/ Provide hypotheses Interpret information, data, patterns Provide evidence and examples Relate evidence and explanations Formulate explanations Evaluate quality of evidence Suggest hypothetical procedures or experimental plans Elaborate their responses (why, how) Compare/contrasts others' ideas Share students' thinking/classroom presentations Share everyday experiences related to current discussions	*Teacher* - Clarifies/Elaborates based on students' responses - Takes votes to acknowledge different students' ideas - Compares/contrasts students' responses to acknowledges and discuss alternative explanations conceptions - Repeats/paraphrases students words - Revoices students' words (incorporates students' contributions into the class conversation, summarizes what student said, acknowledge student contribution) - Promotes debating and discussion among students - Explores students ideas - Captures/displays students' responses/explanations	*Teacher* - Promotes argumentation/Helps students to achieve consensus - Helps relate evidence to explanations - Provides descriptive or helpful feedback - Promotes making sense - Promotes debate/discussion among ideas - Promotes exploration of students' own ideas
Conceptual Structures	*Teacher asks students to:* Provide potential or actual definitions Apply, compare/ contrast concepts Elaborate their responses Share students' thinking/classroom Share everyday experiences related to current discussions Compare/contrasts others' ideas Check their comprehensions	Same as Above	Same as Above
Social Processes	*Teacher asks students to*: Share everyday experiences related to current discussions Share responses not based on observations (e.g., from homework) Share students' thinking/classroom presentations Share/Provide observations Provide evidence and examples Relate evidence and explanations Make predictions/ Provide hypotheses Interpret information, data, patterns Formulate explanations Evaluate quality of evidence Compare/contrasts others' ideas	Same as Above	Same as Above

Appendix A provides a simpler list of the criteria classified according to teacher questions and teacher actions. Notice that the criteria focus mainly on students' explanations, reasoning, and communication—critical characteristics for successful engagement in the scientific enterprise. Appendix B provides definitions for each of the criteria. In the next section we explain the context in which the study reported is embedded.

The Context of the Study

The present paper is an exploratory study that is part of a larger project that focuses on the effects of formal embedded assessments on students' learning (Shavelson & Young, 2000). The study is a small randomized experiment carried out over a 6 month period with 6 "experimental" and 6 "control" teachers measuring student achievement and motivation as well as teachers' beliefs about assessment and learning. The study grew out of 18 months of development and pilot studies with 3 teachers and their students. The study is being conducted within the context of the "Foundational Approaches in Science Teaching" (FAST) middle-school science curriculum (Pottenger & Young, 1992).

FAST 1 is a middle-school science education program developed by the University of Hawaii's Curriculum Research & Development Group (CRDG) and is aligned with the National Science Education Standards (Rogg & Kahle, 1997). It is based on a constructivist philosophy of learning, in which students construct their own knowledge and understanding from their experiences incrementally. This knowledge is developed and clarified through interactions with others (Pottenger & Young, 1992). Investigations are carefully sequenced and connected to previous experiences both in and out of school. Students often work in small groups to share data, ideas and experiences; conduct investigations; summarize; and draw conclusions. Class discussion follows each investigation to identify and clarify generalizations. The FAST program consists of three texts: *FAST 1, The Local Environment*; *FAST 2, Matter and Energy in the Biosphere*; and *FAST 3, Change over Time*.

This study focuses on the introductory Physical Science 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. In

doing so, they work with different states of matter and use their knowledge to explain everyday phenomena. Within this strand, the study focuses on the first twelve investigations (PS1 to PS12). Table 2 provides a quick summary of the characteristics of the twelve investigations (or units).

Table 2.

Summary of the characteristics of the twelve investigations

PS	Title of Investigation	Major Activities	Major Learning Goals
1	Liquids and Vials	Observing a buoyancy anomaly	Making scientific observations; testing predictions
2	Sinking a Straw	Adding mass to a straw and measuring its depth of sinking	Predicting the number of BB's required to sink a straw to a chosen depth
3	Graphing the Sinking-Straw Data	Creating a graph of mass versus depth of sinking	Representing data in graphs
4	Mass and the Sinking Straw	Sinking straws to depth based upon total mass	Increasing mass means a straw will sink more
5	Sinking Cartons	Sinking cartons of different sizes with equal mass	Predicting the depth to which a carton will sink
6	Volume and the Sinking Cartons	Calculating the volume of cartons	Calculating the displaced volume of a carton
7	Floating and Sinking Objects	Calculating the mass and volume of objects	Predicting the displaced water of floating and sinking objects
8	Introduction to the Cartesian Diver	Experimenting with a Cartesian diver	Discovering how a Cartesian diver works
9	Density and the Cartesian Diver	Finding the density of a Cartesian diver	Finding the density of a diver at different depths
10	Density of Objects	Calculating the density of objects	Finding the density of floating and sinking objects
11	Density of Liquids	Finding the density of liquids other than water	Finding the density of liquids
12	Buoyancy of Liquids	Finding the relationship between buoyancy and density	Understanding relative density

In the larger study, six matched pairs of FAST teachers were randomly assigned to experimental and control groups.[2] The study involves a pre- and post-

[2] Ethnicity, free lunch, and student proficiency level were used to match the pairs as best as possible.

test design on student achievement (multiple choice, short answer, performance assessment and Predict-Observe-Explain), motivation (e.g., task and ego goals), and teacher beliefs (SEAL, 2003). The experimental teachers participated in a five-day training program focusing mainly on the implementation of formal embedded assessments that use assessment information to provide immediate feedback to students around the fundamental question that underlies the 12 investigations: Why do Things Sink and Float?We named the assessments Reflective Lessons rather than embedded assessments to make evident to the teachers that their purpose was not to grade students. In the pilot study, we found that use of the words "embedded assessment" somehow contributed to teachers implementing the assessments in a more formal environment, assigning grades in most cases. Calling the embedded assessments Reflective Lessons mirrored their design to enable teachers to step back at key points during instruction of the 12 investigations, check student understanding, and *reflect* upon the next steps they needed to take to move forward. Reflective Lessons are a unique setting involving specific prompts designed for eliciting students' conceptions, encouraging communication and argumentation, and helping students and teachers reflect about learning and instruction. The prompts vary according to where the Reflective Lessons are embedded within the unit. It is important to mention that the training program focused only on the implementation of the Reflective Lessons and not on the implementation of any of the Physical Investigations. Table 3 provides a schematic representation of where they are embedded within PS 1-12 (SEAL, 2003).

Table 3.
Reflective Lessons Embedded in Across the Twelve Investigations.

First 12 Investigations of the FAST1 Physical Science Unit by Section																
Section A Mass					Section B Volume					Section C Density						
1	2	3	4	RL	5	6	RL	7	RL	8	9	10	RL	11	RL	12

The exploratory study reported in this paper is part of what we have called the Implementation Study that focuses on obtaining information about the fidelity of the implementation of the FAST investigations and the Reflective Lessons. Implementation of FAST investigations, including the Reflective Lessons for the experimental teachers, are being monitored daily with teacher logs and videotaping

for all twelve teachers involved in the study. Each classroom is being visited once during the course of the implementation over a two or three-day period. Although most visits have been conducted during Investigation 7, some classrooms have been visited at a later time due to external factors (e.g., snow on the East Coast).

This paper reports an exploratory study that focuses on a strategy developed to analyze the informal assessment practices teachers use during the implementation of the Physical Investigations. The study focuses on the experimental teachers for two reasons. First, they participated in the same training program for the Reflective Lessons; therefore, if the program had any effect on their teaching and assessment practices, it should be expected to occur on the same direction. Second, the effect of diverse informal practices can be observed on the Reflective Lessons as a measure of students' learning. In the next section, we describe in detail the instruments and means used to collect the information.

Method

Participants

The participants in the study were three experimental teachers trained in the use of the FAST curriculum. These teachers already used a FAST program in their science teaching before the data for this study was collected. The teachers' characteristics are briefly described in Table 4.

We conducted a series of ANOVAS to test whether teachers' students differed in achievement and motivation at the beginning of the school year. To measure achievement we used a 40-item multiple-choice test and to measure motivation, a 46-item questionnaire. The multiple-choice test tapped declarative, procedural, and schematic knowledge and focused on questions on density, mass, and volume (alpha coefficient = 0.86; Yin, 2004). The five-scale Likert motivation items (strongly agree to strongly disagree) focused on: goal orientation, perceived task goal, self-efficacy, interest, self-reflection, ego approach, ego avoidance, and perceived ability (alpha coefficients within each set of items ranged from .70 to .89; Yin, 2004).

Table 4.
Teacher General Characteristics

Characteristics	Teacher 1	Teacher 2	Teacher 3
Gender	Male	Female	Male
Ethnicity	White (not Hispanic origin)	White (not Hispanic origin)	White (not Hispanic origin)
Highest Degree Earned	BA	MA	MA
Major in Science	Yes	No	Yes
Minor in Science	Yes	No	Yes
Teacher Credential	State in Science, Diverse Areas	Pre K-6	Residency Certification K-8th, - Science & English
Years of Teaching	14	3	2
Years of Teaching Science	14	1	2
Years teaching 6th/7th grade	3	1	2
Grade Level Taught	7th	6th	7th
Science Sessions Length	55 minutes	40 minutes	55 minutes
Number of Students Taught*	25	25	26

* Students with complete data used in this study.

Results indicated no significant differences in achievement ($F_{(2, 75)} = .24$, $p = .79$), and in four of the motivational beliefs measured; i.e., task goal orientation ($F_{(2, 78)} = .84$, $p = .43$), self-efficacy ($F_{(2, 78)} = .23$, $p = .80$), interest ($F_{(2, 78)} = .33$, $p = .72$), ego approach ($F_{(2, 78)} = 1.99$, $p = .14$), ego avoidance ($F_{(2, 78)} = .83$, $p = .44$), or perceived ability ($F_{(2, 78)} = 1.54$, $p = .22$). However, students from Teacher 3 scored significantly lower than students from the other two teachers in perceived task goal ($F_{(2, 78)} = 3.42$, $p = .03$) and self-reflection ($F_{(2, 78)} = 4.53$, $p = .01$). We concluded that, in general, students among the three groups were similar at the beginning of the study.

Instruments

We developed an Informal Assessment Practices Coding System for examining the informal assessment practices and the Reflective Lessons as a source of information to assess students' learning.

Informal assessment practices coding system. Based on the strategies developed for guiding the identification of the different aspects of assessment conversations, we developed an *Informal Assessment Practices Coding System*. The

coding system focuses on the teacher's actions during assessment conversations. It is organized according to codes used to track informal assessment practices and diverse instructional episodes. The rationale behind the criteria selected for the coding system was previously explained. The purpose of organizing the codes around instructional episodes was to capture the diverse opportunities for discourse and dialogue during instruction.[3] We identified different types of instructional episodes that followed the organization and sequence of the FAST curriculum: Review of Previous Work (REV), Introduction of New Work (INT), Conducting an Investigation (CON), Graphing (GRA), Discussing and Summarizing Investigation Results (RES), Discussing Challenging Questions (CHA), Formal Assessment of Student Learning (ASM), and Classroom/Small Groups Discussion (DIS).

The codes are numbered and organized into the categories of Teacher Questions to the students (eliciting information) and Teacher Actions (recognizing and using information). Another set of codes that will not be discussed but were included in the coding system are: Consistency Practices in the Context of Formative Assessments, Ineffective Practices, and Other Relevant Codes. Consistency Practices involves: Making learning goals explicit, Making classroom expectations/standards explicit, Providing review criteria, Making connections to previous learning, and Monitoring students. Ineffective practices include the following codes: Providing evaluative responses (no explanation provided), Interrupting flow of discussion or students responses, Asking for yes/no answers or fill-in-the-blank answers, and Asking repaired questions (questions are posed but students are not provided with the opportunity to respond). Other relevant codes included: Modeling scientific communication, Modeling process skills, Referring to the nature of science, and Connecting topics with the real world. We do not discuss these codes in this paper because we are focusing on the three components of the informal formative assessment cycle.

All the videotapes for every session taught from PS1 to PS4 were transcribed. Transcriptions were segmented following the *speaking turns* between the teacher and the students. Each teacher speaking turn was numbered and segmented in what we named verbal units (VU). Content of the teacher speaking turns determined the

[3] An *episode* is defined as a series of related and coherent events that delineate/describe an activity packet within each lesson. These related events form a coherent sequence within itself.

segment boundaries. Transcriptions were analyzed twice. The first analyses lead to the identification of the assessment conversations. To identify the assessment conversations we considered the following: The conversation concerned a concept from or aspect of one of the FAST investigations (e.g., did not concern classroom procedures such as checking out books or upcoming school assemblies), the teacher was not the only speaker during the episode (i.e., students also had 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). All the assessment conversations identified were coded by two coders. Coding was done by reading the transcriptions and watching the videos. Twenty-six transcripts were obtained across the three teachers.[4] Figure 2 provides an example of the code format used.

Episode Codes

REV	Reviewing	**CON**	Conducting an Investigation	**RES**	Results/Summarizing	**ASM**	Formal Assessment of Ss Learning
INT	Introducing new work	**GRA**	Graphing	**CHA**	Challenge Question/Extending	**DIS**	Class/Group Discussion of Results

Coding By Verbal Unit

E	VU	Codes				E	VU	Codes				E	VU	Codes			
	1						38						75				
	2						39						76				
	3						40						77				
	4						41						78				
	5						42						79				
	6						43						80				

Figure 2. Informal assessment practices coding form.

Student assessments. To assess student learning we used the information collected in the Reflective Lesson administered after PS4. In what follows we describe each of the formal formative assessments. Reflective Lessons involved four types of assessments: Graphing (3 items), Predict-Observe-Explain (POE; 12 items), Open-Ended Question (1 item), and Prediction Question (2 items). These assessments (Reflective Lessons) are implemented in three sessions, although they

[4] Teacher 2's transcript was lost due to low battery.

may take only two or two-and-a-half sessions. Each session takes about 40 minutes to complete. In what follows we describe each of the questions in more detail.[5]

Graph. This type of prompt asks students to use the data they have collected in different FAST investigations as evidence for their conceptions. Here the question provides a representation of data that is familiar to the students: a scatterplot that shows the relationship between two variables, mass and depth of sinking. This prompt requires students to judge the quality of the graph (completeness and accuracy) and to interpret a graph that focus on the variables involved in sinking and floating. There is also one prompt that asks students for a self-evaluation about how well they did judge the quality of the graph presented.

Graph - Interpretation	***Code***	***Score***
No paper attached/No response/ Illegible/Off Task		
Describes an incorrect relationship – *less amount of weight makes things sink deeper or float higher*		0
Describes one or more plots on the graph or *gral* trend – *you need 0.8 g of BBs to sink the straw 4 cm.…*	1	1
Uses mass, weight (heavy/light) AND/OR depth of sinking BUT no relationship stated	2	1
Provides examples of the relationship – *if you add weight to a can it will sink if you take it away it will float*	3	2
Correctly identifies the relationship using		
BBs and depth of sinking – the more BBs, the deeper the sinking	4	3
Weight/heaviness and depth of sinking – the more weight/heavy the more sinking	5	3
Mass and depth of sinking – the more mass the more depth of sinking	6	4
Mass/ Weight/ BBs and depth of sinking BUT relationship is backwards	7	

Figure 3. Portion of the scoring form used in Question 1, Graph–Interpretation.

Scoring of this question focused on three aspects of the students' responses: correctly identifying the graph components that need to be fixed, correctly fixing those problems, and correctly interpreting the graph. Another score was assigned for the self-evaluation. The interpretation of the graph involved not only a score based on the quality of the interpretation, but also a code that could distinguish the quality of the scientific communication. Therefore, students' responses with the same score (e.g., 3) could be coded differently based on the language used to state

[5] We did not include in our analysis the information on the open-ended question (a single open-ended question that asks students to explain Why things sink and float with supporting examples and evidence) because the information gathered was similar to the interpretation of the graph. The information will be analyzed at a later point.

the relationship (e.g., BBs vs. weight contributing to depth of sinking). Figure 3 provides an example of the scoring approach used across the questions.

Predict-Observe-Explain (POE). This type of prompt focuses on one event and engages students in three tasks. First students *predict* the outcome of some event related to sinking and floating and *justify* their prediction. Then students observe the teacher carry out the activity and describe the event that they see. And finally, students *reconcile* any conflict between prediction and observation.

Scoring of the Prediction part was based on the following aspects: Correctness of prediction and appropriateness of the justification provided for the prediction. Scoring of the Observation and Explanation part was based on: correctness of the observation, correctness of the decision about the matching between the prediction and observation, and quality of the reconciliation (explanation) provided.

Predict-Observe. A slight variation on the POE, this prompt asks students to predict and observe (PO) an event. Students are only asked to predict and explain their predictions, and they are not asked to reconcile their predictions with what was observed. POs act as a *transition* to the next instructional activity of the unit in which the third piece of the prompt, the explanation, will emerge. POs can be thought of as a springboard that motivates students to start thinking about the next investigation.

Scoring of the Prediction part was based on correctness of prediction and appropriateness of the explanation provided for the prediction. Observation was not scored since it was not critical to assess the students' performance.[6]

Procedure

Teachers were asked to videotape their classrooms in every science session they taught beginning with their first FAST lesson. Each of the teachers was provided with a digital video camera, a microphone, and videotapes. All were trained on how to videotape their classes. Again, no further instruction was provided on how to conduct the Physical Investigations from FAST. Teachers were asked to submit the tapes every week in stamped envelopes.

[6] It is important to remember that this question of the Reflective Lesson was used as a springboard for the next investigation.

Results

First, we provide information about the teacher informal assessment practices, then on the students' observed performance on the Reflective Lesson questions. Finally, we link the students' performance to the informal assessment practices observed across teachers.

Teachers' Informal Assessment Practices

In this section we first provide a general characterization about the teaching sessions across the four investigations (PS1 to PS4). We then provide information about the pattern of informal assessment practices. We end the section with some examples of the conversations across the three teachers.

Characterizing teaching sessions. Figure 4 shows that the number of sessions taken by these teachers to teach the four investigations was similar (Teacher 1 = 10, Teacher 2 = 9, Teacher 3 = 8). The difference arose in how these sessions were distributed. For example, whereas Teacher 3 took three sessions in PS2, Teacher 1 and 2 took only one session. Teacher 1 took two sessions to go over again PS2 and PS3, whereas the other two teachers did not. We concluded that teachers implemented the four investigations in a similar amount of time.

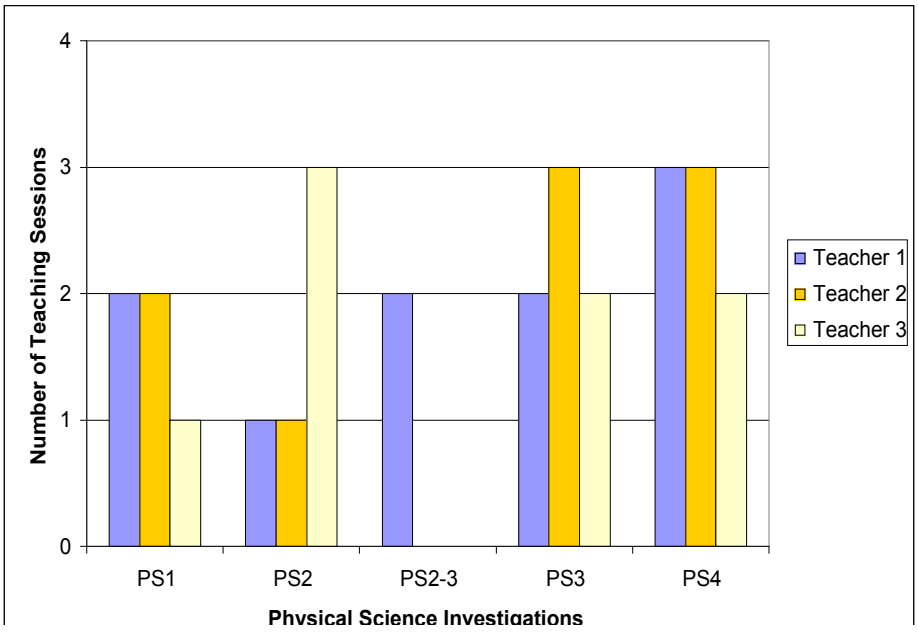

Figure 4. Number of teaching sessions by Teacher and Psychical Science Investigation (PS).

To display the general characteristics of the assessment conversations across teachers, we provide information on: (1) the number of conversations over the teaching sessions across the four investigations, (2) the number of teacher speaking turns by conversation, and (3) the episodes in which these conversations took place. We identified assessment conversations in 19 of the 26 transcripts.

Figure 5 provides information about the assessment conversations identified across the four investigations. The number of assessment conversations over the four investigations may provide information about how frequently teachers practiced this aspect of informal formative assessment. More assessment conversations were identified for Teacher 2 (10) than for the other two teachers (Teacher 1 = 5; Teacher 3 = 8). However, the number of assessment conversations is not enough. Some conversations involved more interactions between teachers and students than others.

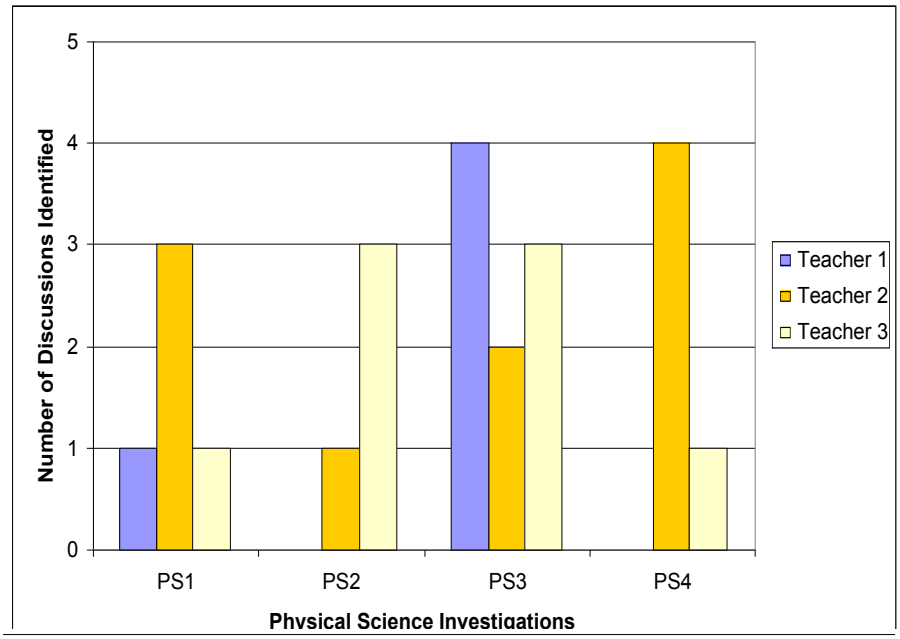

Figure 5. Number of assessment conversations by Teacher and Physical Science Investigation (PS)

Table 5 provides the number of speaking turns as an indicator of this interaction as well as information on the instructional episodes in which the assessment conversations took place. Notice that Teacher 2 held assessment conversations on every type of instructional episode that was suitable at that point

in the FAST curriculum (PS4).[7] Teacher 2's assessment conversations involved more speaking turns between teacher and students (388) than the assessment conversations of the other two teachers (Teacher 1 = 184; Teacher 3 = 342). In what follows, we characterize the assessment conversations in greater detail.

Table 5.
Number of Speaking Turns by Episode for the Four Investigations (PS1 to PS4)*

		1	2	3	4	5	6	7	8
PS	Teacher	REV	INT	CON	GRA	RES	CHA	ASM	DIS
1	1		76						
	2		55	67					18
	3			52					13
2	1								
	2	18	13				22		13
	3		76						35
3	1								23 22
	2				42	38			
	3				77 17				
2&3	1					64			
	2								
	3								
4	1				4				18
	2				44 39				32
	3								80 69

* More than one number within the same episode indicates multiple assessment conversations for the same PS.

Characterizing assessment conversations. We portray conversations based on the strategies used across three characteristics of informal formative assessment discussed previously: eliciting, recognizing, and using information. If the informal formative assessment cycle is being completed we should expect teacher to use strategies across the three characteristics. Furthermore, these strategies should somehow be linked to each other.

[7] The assessment episode came after PS4 and those videos were not coded.

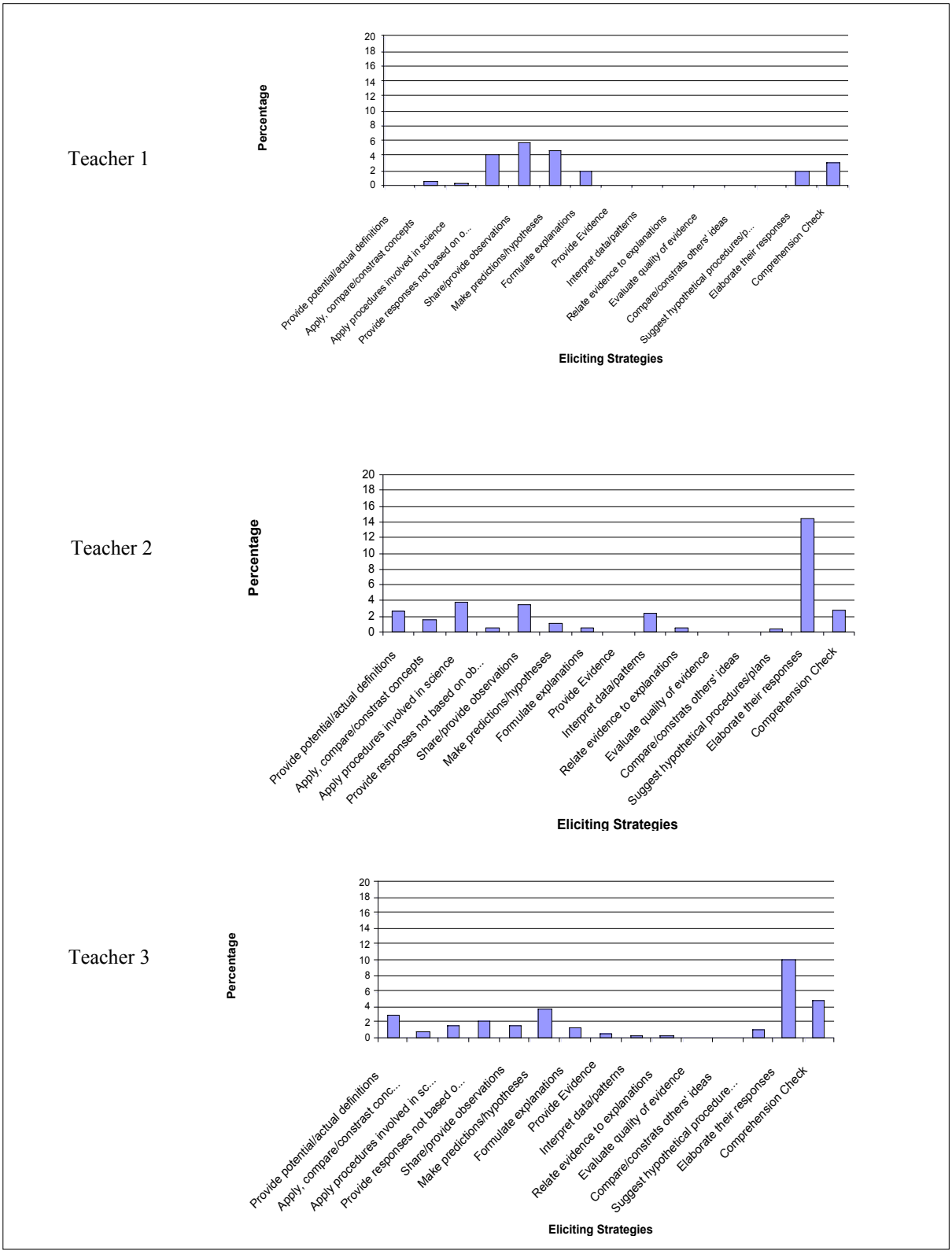

Figure 6. Percentage of eliciting questions by type across the three teachers.

Frequency of codes for all the assessment conversations was transformed to percentages based on the total number of verbal units coded across the transcripts. For simplification, we only present these percentages instead of by transcript. Figure 6 provides the eliciting strategies across the three teachers. Notice that percentages do not add to 100 because other codes, not discussed here, are included in the calculations.

The most outstanding finding on the eliciting practices across teachers is related to the strategy, "Ask Students to Elaborate Their Responses." This type of question involves mainly asking students why and how (e.g., How do you know that? What do you mean by how much density the liquid has? Why will it start to sink?). This type of question is fundamental in the context of assessment conversation because it influences student thinking. It helps students to become aware of their own thinking in the context of a discussion (van Zee, Iwasyk, Kurose, Simpson, & Wild, 2001). Furthermore, this question can reveal students' mental models. This type of elicitation question helps the teacher to assess students' epistemic frameworks (e.g., what constitutes data) as well as to access students' understanding and use of science concepts (Duschl, 2003). Furthermore, by making students' thinking visible through oral, written, or pictorial means, teachers can recognize more easily the communication frameworks students use in the context of science inquiry and what changes are required in those frameworks. Relative to other percentages, those for elaboration were the highest but the absolute value was not very high. Teacher 2 used this strategy the most (14 percent), followed by Teacher 3 (10 percent) and Teacher 1 (2 percent).

Effective science teachers continually assess students' conceptual understanding (Duschl, 2003). We considered that the questions, "Ask Students to Provide Potential/Actual Definitions, Apply/Contrast/Compare Concepts, and Comprehension Checking" could be considered appropriate for eliciting information about students' conceptual understanding. Although percentages are low, the three teachers do practice this type of elicitation question (Teacher 1 = 0.7, 0.3, and 3.1 percent respectively; Teacher 2 = 2.6, 1.6, and 2.8 percent, respectively; and Teacher 3 = 2.9, 0.9, and 4.9 percent, respectively). Notice that Teacher 1 rarely asked students for potential or actual definitions; he tended to define the concepts for them (4.1 percent).

Epistemic frameworks are the knowledge structures (e.g., data, evidence, principles) and the rules and criteria used in scientific inquiry for determining what counts, for example, as a good piece of evidence, graph, explanation, argument, or experiment (Duschl, 2003; Duschl & Gitomer, 1997). None of the teachers elicited this type of information from students (e.g., Ask Students to Evaluate the Quality of Evidence, Ask Students to Compare/Contrast Others' Ideas). But, for example, all of them asked for predictions at some point (Teacher 1 = 4.8 percent, Teacher 2 = 1.1 percent, Teacher 3 = 3.7 percent) and two of them asked students to relate evidence to explanations (Teacher 1 = 0 percent, Teacher 2 = 0.6 percent, Teacher 3 = 0.3 percent). The issue here is whether teachers ask students, for example, to elaborate how they arrived to such predictions or how they connected evidence to explanations.

Eliciting information from students, overall, is easier than doing something with the information provided by students. Figure 7 provides information about the strategies teachers use to recognize the information students are providing. Across the three teachers, Repeating and Revoicing are the two strategies used more frequently. Although repeating is a common way to acknowledging that students have contributed to the classroom conversation, revoicing has been considered a better strategy for engaging students in the classroom conversations (O'Connor & Michaels, 1993). When teachers practice revoicing, "they work student's answers into the fabric of an unfolding exchange, and as these answers modify the topic or affect the course of discussion in some way, these teachers certify these contributions and modifications" (Nystrand & Gamoran, 1991, p. 272). Revoicing, then, is not only a recognition of what the student is saying, but constitutes, in a way, an evaluation strategy because the teacher acknowledges and builds on the substance of what the student says. Furthermore, it has been found that this type of engagement in the classroom has positive effects on achievement (Nystrand & Gamoran, 1991). The highest percentage of this code was found for Teacher 2 (10 percent), followed by Teacher 3 (5.75 percent) and Teacher 1 (5.4 percent).

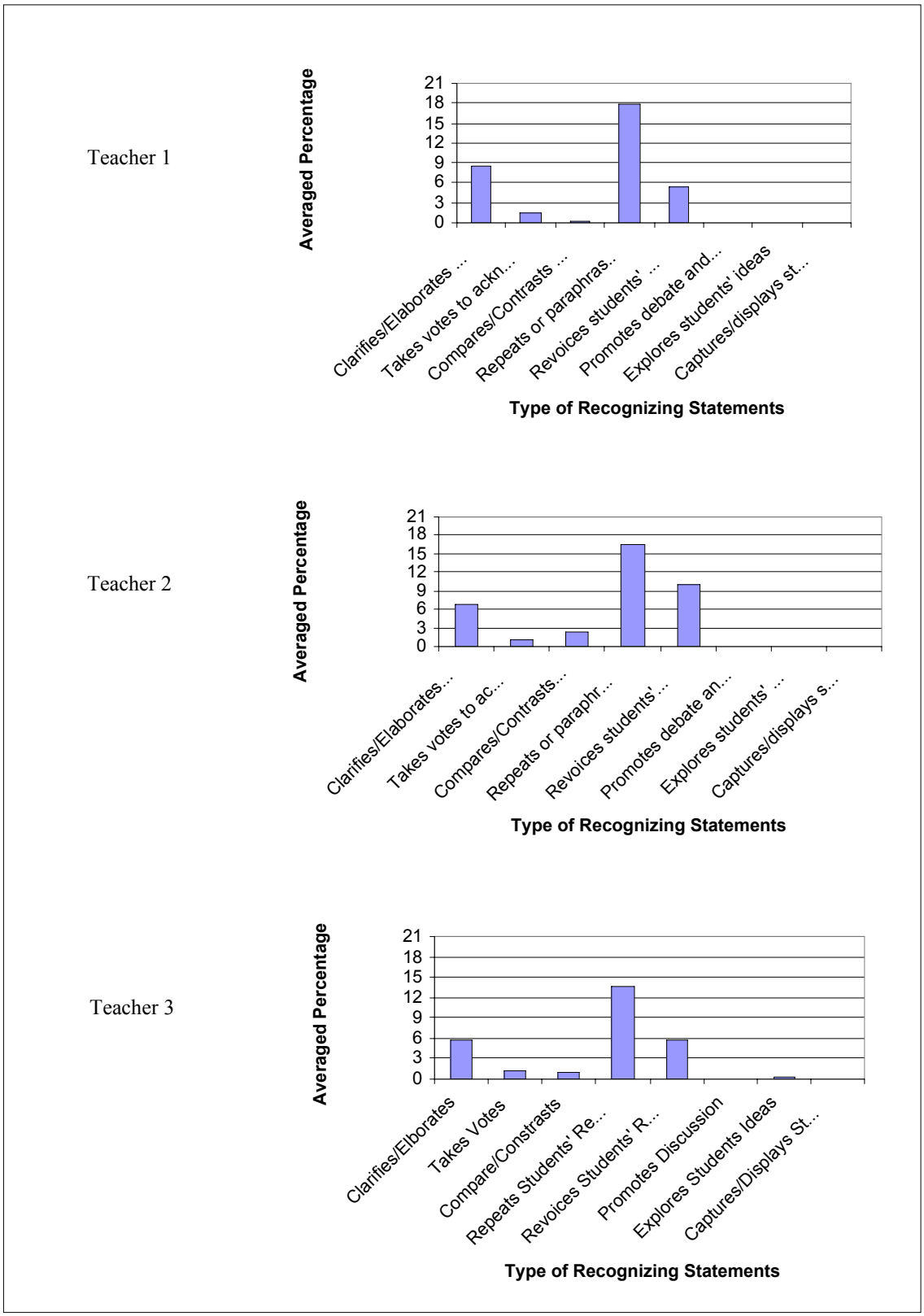

Figure 7. Percentage of recognizing information by type across the three teachers.

Another strategy that has been considered essential in engaging students in assessment conversations is Comparing and Contrasting Students' Responses (Duschl, 2003). This strategy is essential in examining students' beliefs and decision making concerning the transformation of data to evidence, evidence to patterns, and patterns to explanations.

We found that Comparing and Contrasting Students' Responses is not a strategy frequently used by these teachers. The highest percentage was observed with Teacher 2 (1.6 percent), followed by Teacher 3 (0.9 percent) and Teacher 1 (0.3). The critical issue in this strategy is whether teachers point out the critical differences in students' reasoning, explanations, or points of view. Furthermore, for the cycle (eliciting, recognizing, using) to be completed, the next necessary step would be helping students to achieve a consensus based on scientific reasoning. This can be done by promoting discussion and argumentation. Figure 8 provides information about the strategies for "using the information."

Clearly, Promoting Argumentation (whole class) was not a strategy used by these teachers. Teacher 2 used it only once, even though she compared and contrasted students' responses on thirteen occasions. It seems, then, that the third and most important step to close the cycle was missed. The opportunity to move students forward in their understanding of conceptual structures and/or epistemic frameworks was lost.

Helpful feedback, or comments that highlight what has been done well and what needs further work, was also a strategy rarely used. Only 1.6 percent of the verbal units were coded under this strategy for Teacher 2, and 0.1 for Teacher 3. Both Teachers 2 and 3 promoted making sense, but again quite infrequently (Teacher 2 = 0.4 percent, Teacher 3 = 0.6 percent).

The most dramatic finding is that Teacher 1 did not use any of the strategies considered under this component of the informal formative assessments practices. We concluded that, overall, Teacher 2 showed better strategies of informal assessment than Teachers 1 or 3. Did the quality of teacher informal assessment practices affect student performance on the Reflective Lesson question? In the next section we provide information about the students' performance on the questions used to measure their learning.

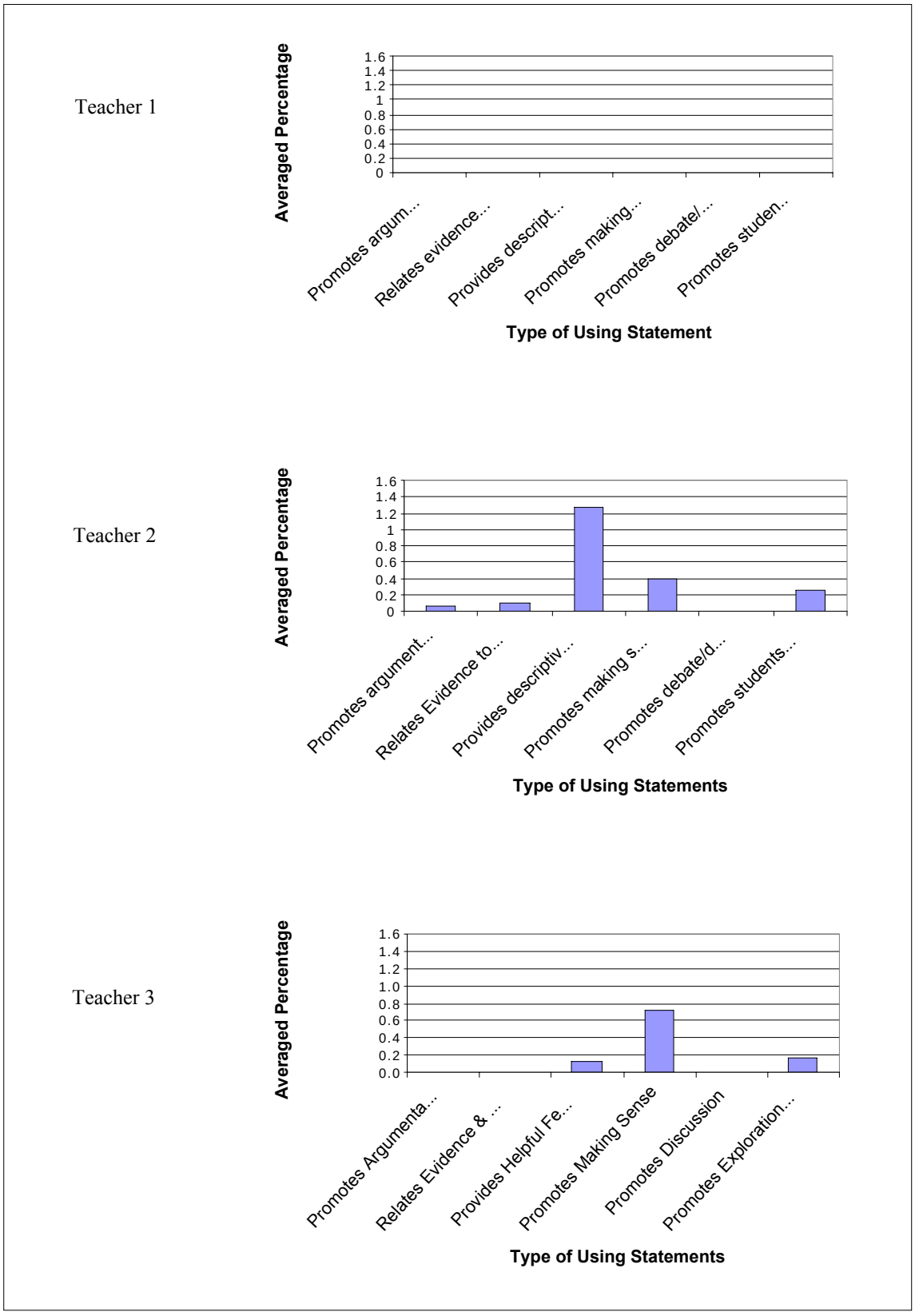

Figure 8. Percentage of using information by type across the three teachers.

Assessment conversation examples. In what follows we describe some examples of the assessment conversations we coded. The examples focus on the same topic, graphing. We selected graphing at PS3, Graphing the Sinking Straw Data, because one the Reflective Lesson questions deals with this topic. The excerpts deal with the scaling of a graph. While the complete text of the conversations has been excluded for the sake of brevity, the examples presented can provide a sample of the conversations that took place around graphing across the three classrooms.

Teacher 1. During this "conversation" the teacher provides some guidance to the students on the task they will do and models on an overhead how to graph. Notice how the teacher repeats students' responses (e.g., speaking turn 10, 11, 12), but does not revoice or ask students to elaborate their responses (e.g., speaking turn 7). Another interesting characteristic of this teacher is the length of speaking turns as compared to Teachers 2 or 3.

5. Well, we're going to have to come up with some sort of a scale, some way that we're going to be able to use our X and Y, horizontal and vertical axes, and so when we do that, let's try to figure out, how many tall is this? Three boxes tall. Focus that. It's blurry to me. Is it blurry to you? How many tall? I've only counted 20 boxes high, and what's our maximum number over here on the sheet? It looks like 39, so we could probably go by twos, okay, we'll start out with twos. Let's go with zero here, and everything, every full line is going to be two. Two, four, six, eight, ten, twelve, fourteen, sixteen, eighteen – you guys can count, I don't have to count for you. I can almost get the whole thing on there for you to look at. Okay. Now, what did you say, three here? If we look back over to PS3, [inaudible]. It doesn't tell us, so we can put PS3 on [inaudible], that's fine. You'll know what we're talking about. So, PS3 and then put individual or mine or something like that that will tell you that this is your group's, and then we'll go to the next one and it will be PS3 and then we'll put the class; that way, you'll know one's class data and one's yours.

[Inaudible.]

6. Okay. Now, going across the bottom here, let's move a little bit quicker on this, going across the bottom, our numbers are going to be, what, we've got 4 centimeters, 6, 5, 6, 7, 8, so we've got 10, so if we could go by four – how many [inaudible]?

Sixteen.

7. Sixteen across the bottom. Let's say each one could be, or each two could be one, but that's probably going to push the issue. [Inaudible.] Okay, so let's go with each box is worth one, and that way we fit everything, we may get a little cramped but we'll fit everything on the page.

[Inaudible.]

8. I just stopped at nine because that's all we have up here. Now, we've been preaching about you've got to have units and variables, so what were we measuring down here, length?

Of the straw.

9. And how did we measure it?

Centimeters.

10. Centimeters, okay? So that's supposed to be a parentheses [inaudible] length in centimeters…

In speaking turn 5, Teacher 1 provides the students with direction on how to make a graph as he models it for the class. Although he asks questions in the course of his speaking turn, he does not wait for students to respond (repaired questions). Although students do respond to some of his questions later, the conversation elicits a minimal amount of information, and the students' responses are repeated but not revoiced.

Teacher 2. As with Teacher 1, during this "conversation" Teacher 2 provides some guidance to the students on the task and she also models on an overhead how to graph. Notice how she orients students, and then starts with checking students (e.g., speaking turn 11), how, after repeating a student's response, she clarifies/elaborates (e.g., speaking turn 15), and how she asks students to elaborate student's responses (e.g., speaking turn 14). Also, in turn 15 she praised a student for using "a very scientific word."

11…So the first three things you want to do, very important things, you want to label your vertical axis, you want to label your horizontal axis, and then you want to give the whole graph a title. And we've done that. So, taking a look at this, am I ready to go? Can I start plotting my points?

No.

12. Why not?

You didn't…

13. What do we need to do? Eric?

Put the scale

14. The scales. What do you mean by scales?

The numbers.

15. "The numbers." Good. Excellent. I liked that you used the word "scales," it's a very scientific word. So, yes, we need to figure out what the scales are, what we should number the different axes. So, let's take it a step at a time; let's think about the different lengths, what we did in centimeters first, what were the different lengths of depth that we wanted it to sink to in centimeters? What were they? Just by looking at this or looking at your own table and paper, what were the different depths that we wanted it to sink to? Brooke?

In contrast to Teacher 1, Teacher 2 in this example starts her conversation with student input, and builds upon it through revoicing, asking students to elaborate,

and clarifying and elaborating upon student comments. This pattern occurred many times in Teacher 2's transcripts.

Teacher 3. This teacher also models graphing on an overhead and checks students (e.g., speaking turn 35). However, similar to Teacher 1, the conversation is more focused on delivering guidance on how to make a graph rather than relying upon student input. Questions are asked of the students but the responses sought are not provided, so the teacher provides the information to the students.

> 13…Jane, what do we call that when we number the X and Y axis?
>
> *Labeling?*
>
> 14. Right, but what is it after we label the X and Y axis? What does that become?
>
> *The [inaudible].*
>
> 15. When I say it, you guys will remember it. It's called the scale. Right, we have to put a scale, our scale, it tells us what it is we're measuring, right? Actually, it tells, the title tells us what it is we're measuring, but the scale gives us increments to use to measure it with. So, looking on our data, where is my data sheet, looking on our data up here, how far do we have to go, how big of a number do we [inaudible]? Someone who hasn't talked yet today. If you look on here, we want to know the number of bb's; right? We need to be able to fit on our graph all the bb's that were counted off in this data table. What's the biggest number that we have? Someone who hasn't talked yet. Jessica.
>
> *Forty.*
>
> 34. Forty. Is 40 your biggest one? Over here? Oh, you know what, that is our biggest one. We can use these numbers. We don't have all of them but we can go ahead and use them. So, 40. You want to make sure that we can get 40 on the graph. So, you know what, guys, I take that back, I take that back. I only want to graph this data to the left of that line; does that make sense? Do you guys want to graph these numbers over here?
>
> *No.*
>
> 35. The only reason I say that is because I'm looking at my graph, and we're going to have our data here all kind of clumped in one spot, and then these numbers are going to be huge because we [inaudible] centimeters. So, I think this, for this graph, it might look, it might make more sense to [inaudible] this is our first graph, I don't want to confuse you, because it's line of best fit I find often confuses people. So I'm going to make this first one a little less confusing. So, I only want to graph this part of the [inaudible]. Okay? So, if you're in group 1, you're going to do that first, group 1 first. But, anyway, you have that [inaudible]. Does that make sense? Did that confuse anybody? No? Okay.

While Teacher 3 shows concern for the level of his students' understanding, reflecting the fact that he has previously elicited and recognized their responses, line 35 shows that he asks students if his statements have made sense rather than promoting consensus or providing feedback.

Students' Performance

Before presenting the students' performance results we first provide information about the interscorer reliability of the Reflective Lesson questions.

Reliability and agreement across scorers. Twenty randomly selected student responses to the three assessments were independently scored by two scorers to evaluate interscorer reliability for each question in the Reflective Lesson after PS 4. Table 6 provides the reliability coefficients of the students' responses.

Table 6.
Interscorer Reliability Across Questions

Question	Interrater Reliability
1 Graph	.97
Self Evaluation of the Graph	1.00
2 Predict-Observe-Explain	.98
3 Predict-Observe	.94
AVG. TOTAL	.97

The magnitude of the interscorer reliability coefficients varied according to the question at hand. However, coefficients were never below .94. Furthermore, the averaged reliability across questions was .97. Cohen's kappa for students' responses codes for Question 1, Graph Interpretation, was .82. We concluded that scorers were consistent in scoring and coding students' responses to the different questions involved in the Reflective Lesson after PS4. Based on this information, the rest of the students' responses, 56, were only scored by one of two scorers. For the 20 students scored for assessing the interscorer reliability, the averaged score across the two raters was used on the rest of the analyses conducted.

Students' observed performance. In this section we focus on the student performance as observed in their responses to the Reflective Lesson after PS4. Mean scores and standard deviations across questions are provided in Table 7. Students' performance varied across questions and teachers. The highest students' performance across the three items was observed for Teacher 2, followed by Teacher 3 and Teacher 1. Only in Question 3, Predict and Observe, the student performance of Teacher 3 was lower than the observed performance on students from Teacher 2.

We carried out a series of ANOVAS for each question to evaluate the differences across Teachers.[8]

Significant differences were observed in Question 1, Graph ($F_{(2,69)} = 5.56$; $p = .006$). Tukey HSD indicated that Teacher 2's students performed significantly better than students of Teacher 1 and Teacher 3. No significant differences were observed between Teachers 1 and 3. The same pattern was observed for the Self-Evaluation of the Graph ($F_{(2,65)} = 8.09$; $p = .001$). Students of Teacher 2 were able to self-evaluate themselves significantly more accurately than those students of Teachers 1 and 3.

Table 7.

Mean Scores and Standard Deviation Across Questions

Question	Max	Teacher 1			Teacher 2			Teacher 3		
		n	Mean	S.D.	*n*	Mean	S.D.	*n*	Mean	S.D.
1 Graph	14	23	6.17	2.87	23	8.76	2.73	26	6.44	3.09
Self Evaluation of the Graph	2	22	0.59	0.91	23	1.56	.66	23	0.95	0.88
2 Predict-Observe-Explain	20	25	6.32*	2.89*	25	14.84	4.49	22	10.78	4.34
3 Predict - Observe	2	22	1.27	0.98	22	1.54	1.01	10	0.40	0.52

* Students' responses for the Prediction part were not available.

Unfortunately, not all students' responses were available for Teacher 1 in Question 2, Predict-Observe-Explain. Only the Observation-Explanation part could be scored for this teacher. Significant differences observed ($F_{(2,70)} = 28.93$; $p = .000$) were not considered valid. Therefore, we carried out a second analysis with Teacher 2 and Teacher 3 only. Results indicated that Teacher 2's students' performance was significantly higher than Teacher 3's students ($F_{(1,46)} = 10.08$; $p = .003$). A third ANOVA for the Observe-Explain part including the three teachers showed significant difference among the three groups ($F_{(2,63)} = 14.31$; $p = .000$). Tukey HSD

[8] Assumption of homogeneity of variances was met in all ANOVAs, except for Question 3, Predict-Observe. Significance of the Levene statistic ranged from .24 to .71.

indicated, again, that Teacher 2's students performed significantly better than students of Teacher 2 and Teacher 3. No significant difference was observed between Teacher 1 and 3.

Significant differences were also observed in Question 3, Predict-Observe ($F_{(2,51)} = 5.26$; $p = .008$). Dunnett C (equal variances not assumed) indicated that students' performance for Teacher 3 was significantly lower than Teacher 1 and 2. No significant differences were observed between Teachers 1 and 2.

In what follows we present information on the quality of the students' responses in Question 1, Graph-Interpretation. Table 8 provides the percentage of students' responses by code. Codes are presented based on the quality of the communication—Code 1 being the least appropriate and Code 6 being the most appropriate. Code 7 is a special case, backwards relationship, of a correct relationship (e.g., *This graph tells that the more centemeters (sic) you have the more BeBes (sic) it is going to take to sink the straw*). Results indicated that more than 50 percent of the students' graph interpretations from Teacher 2 were of high quality. Communications were clear and students used the most appropriate language. In contrast, only 12 percent of students' responses for Teacher 1 and 27 percent for Teacher 3 were coded as 6. We concluded that students' communication from Teacher 2 were more appropriate than the communications observed in the other two classrooms.

Linking teacher practices to student performance. The results described previously indicated that students of Teacher 2 were those whose observed performance was higher across all the questions. The above analysis also indicates that the students of Teacher 2 performed significantly higher that those students from the other two teachers on the Reflective Lessons. The strategies reflected in the assessment conversations analyzed in the previous section indicate that Teacher 2 was the one teacher whose strategies during the assessment conversations were closer to what one should expect in the context of an inquiry class than those of the other two teachers. We concluded, then, that better informal assessment practices lead to better student performance—at least in the sample we used for this study.

Table 8.

Percent of Students' Interpretation of the Graph Question by Quality Code

Code	Explanation	Teacher 1	Teacher 2	Teacher 3
		(*n* = 25)	(*n* = 25)	(*n* = 26)
0	Non Valid Code (Incorrect Response)	16.0	8.0	3.8
1	Description of general trend or plots	8.0	16.0	11.5
2	Uses mass, weight AND/OR depth of sinking BUT No relationship provided	16.0	0.0	11.5
3	Provides examples of the relationship	4.0	0.0	11.5
4	States relationship using BBs & depth of sinking	8.0	4.0	7.7
5	States relationship using weight & depth of sinking	12.0	0.0	26.9
6	States relationship using mass & depth of sinking	12.0	52.0	7.7
7	Stated relationship is backwards	4.0	8.0	0.0
77	Off task response	8.0	0.0	0.0
88	Illegible	4.0	0.0	0.0
99	No Response	0.0	0.0	15.4
100	No paper attached	8.0	8.0	3.8

Conclusions

This paper provides information on an exploratory study about informal formative assessment practices in three science classrooms. We provide a framework for examining these practices based on three components of formative assessment (eliciting, recognizing and using information) and three domains linked to science inquiry (epistemic frameworks, conceptual structures, and social processes).

We developed a coding system to track strategies teachers used across the three informal formative assessments components. A piece of information related to the validity of the coding system was provided. The coding system could capture differences in assessment practices across the three teachers. Furthermore, based on three questions used for assessing students' performance we were able to link

students' level to the quality of the teachers' assessment practices. We concluded that the strategies provided important information about the teachers' informal assessment practices. Unfortunately, due to time constraints we could not provide information about the interscorer agreement of the coding system proposed. However, we believe that coders can be trained to consistently code teacher practices, a belief that needs to be empirically tested.

Our findings revealed that for teachers it is more common to elicit information from students than to recognize and use the information to improve student learning. Unless the assessment cycle is completed, the opportunities for helping students to improve their learning cannot be fulfilled. Yet our findings indicate that certain eliciting practices, such asking students to elaborate their responses, seem to be more beneficial than asking simple questions (fill-in-the-blank questions).

Our findings also suggest areas for further inquiry as data analysis is extended to all 12 teachers involved in the larger study. For example, Teacher 2, found to exhibit the most informal formative assessment practices and higher student performance than Teachers 1 and 3, does not hold a bachelor's degree in a science field, and holds a teaching credential in K-6 rather than in science. Furthermore, extending the present analysis across all 12 Physical Science investigations may yield more useful information on the longitudinal use of informal formative assessment across an instructional unit.

While the present study is only exploratory, our results indicate that there is room for professional development in this area. Teachers need to have more assessment conversations that engage students more substantially in their own learning.

REFERENCES

Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht, Netherlands: Kluwer Acedemic Publishers.

Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science Education*, *21*, 49-97.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Duschl, R. A. (2003). Assessment of inquiry. In J. M. Atkin & J. E. Coffey (Eds.) *Everyday assessment in the science classroom* (pp.41-59). Washington, DC.: 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*(1), 37-73.

Lemke, J.L. (1990). *Talking science. Language, learning and values*. Norwood, NJ: ABLEX Publishing Corporation.

Marks, G., & Mousley, J. (1990). Mathematics education and genre: Dare we make the process writing mistake again? *Language and Education*, *4*(2), 117-135.

Martin, J.R. (1989). *Factual writing: Exploring and challenging social reality*. Oxford: Oxford University Press

Martin, J.R. (1993). Literacy in science: Learning to handle text as technology. In M.A.K. Halliday & J.R. Martin (Eds.). *Writing science: Literacy and discursive power* (pp.166-202). Pittsburgh, PA: University of Pittsburgh Press.

National Research Council (1999). *The assessment of science meets the science of assessment*. Board on Testing and Assessment Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press

National Research Council (2001). *Inquiry and the National Science Education Standards*. Washington, DC: National Academy Press.

Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English*, *25*(3), 261-290.

O'Connor, M. C., & Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly*, *24*(4), 318-335.

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

Pottenger, F. & Young, D. (1992). *The Local Environment: FAST 1 Foundational Approaches in Science Teaching.* University of Hawaii, Manoa: Curriculum Research and Development Group.

Rogg, S., & Kahle, J. B. (1997). *Middle level standards-based inventory*. Oxford, OH: Miami, University of Ohio.

Shavelson, R. J., & Young, D. (2000). *Embedding assessments in the FAST curriculum: On the beginning the romance among curriculum, teaching and assessment*. Proposal submitted at the Elementary, Secondary and Informal Education Division at the National Science Foundation.

Shavelson, R. J., Black, P., Wiliam, D., & Coffey, J. (2003). *On aligning summative and formative functions in the design of large-scale assessment systems*. Paper submitted for publication.

SEAL (2003). On The Integration Of Formative Assessment In Teaching And Learning with Implications for Teacher Education. *Paper presented at Biannual EARLI Conference. Padova, Italy*.

van Zee, E. H., Iwasyk, M., Kurose, A., Simpson, D., & Wild, J. (2001). Student and teacher questioning during conversations about science. *Journal of Research in Science Teaching*, *38*(2), 159-190.

Yin, Y. (2004). *Formative assessment influence on students' science learning and motivation*. Proposal to the American Educational Research Association.

Zellenmayer, M. (1989). The study of teachers' written feedback to students' writing: Changes in theoretical considerations and the expansion of research contexts. *Instructional Science*, *18*, 145-165.

APPENDIX A

Criteria for Recognizing and Guiding Assessment Conversations

Teacher asks students to:	Teacher:
♣ Provide potential/actual definitions	♣ Defines concepts
♣ Use, apply, compare, and contrast concept(s)	♣ Clarifies/Elaborates
♣ Use & apply known procedures involved in science	♣ Takes votes
♣ Provide responses not based on observations (e.g., homework responses)	♣ Compares/contrasts students' responses/explanations
♣ Share/Provide observations	♣ Promotes argumentation/consensus among students based on scientific reasoning
♣ Share experiences outside the classroom	♣ Relates evidence and explanations
♣ Make/Provide predictions/ hypotheses	♣ Repeats/Paraphrases student words
♣ Formulate explanations	♣ Revoices student words by. incorporating student's contribution into the class conversation, OR acknowledging student's contributions
♣ Provide evidence	♣ Does comprehension checking
♣ Interpret data, patterns	♣ Provides feedback: descriptive or explicit improvements
♣ Relate evidence to explanations	♣ Promotes making sense
♣ Evaluate quality of evidence	♣ Makes learning goals explicit
♣ Compare/contrast others' ideas or explanations	♣ Makes explicit classroom expectations/standards
♣ Suggest hypothetical procedures or plans	♣ Provides/reviews criteria
♣ Elaborate/explain response (why, how)	♣ Captures/displays student responses/explanations
♣ Shares experiences from outside the classroom	♣ Makes connections to previous learning
	♣ Refers to nature of science
	♣ Connects topic with the real-world
	♣ Promotes sharing of students thinking (presenting to the class)
	♣ Promotes debating/discussing among students
	♣ Models how to communicate scientific knowledge
	♣ Models process skills

APPENDIX B

Definitions of the Strategies and the Codes Used

Ask Students	***Students are asked to***:
Define concept(s)	Provide actual or potential definitions for concepts, e.g. 'what is mass?'
Use, apply, and compare concept(s). Use and apply known procedures	Make use of or apply previously defined concepts, or to compare/contrast concepts with each other, e.g., 'Is this an anomaly?' Or, 'What is the difference between mass and volume?'
Share experiences outside the classroom Share experiences inside the classroom	Provide experiences they have had outside the context of the present science classroom to content related to the current discussion.
Provide responses not based on observations (e.g., hw)	Read or display responses from homework or some other format that were not initially gathered as observations
Observations (use ● for written observations)	Share their observations. These could be either oral or written.
Provide predictions/ hypothesis (use ● for written explanations)	Share their predictions and hypotheses. These could be either oral or written.
Explanations (use ● for written explanations)	Share their explanations. These could be either oral or written.
Provide evidence and examples Interpreting data/results, graphs	Provide evidence supporting evidence and examples collected in a scientific manner.
Relate evidence to explanations	Connect evidence collected during class to an explanation.
Evaluate quality of evidence (promotes)	Examine the quality of evidence provided for a claim.
Compare/contrast ideas	Compare and contrast the ideas of other students.
Suggest hypothetical procedure/experimental plan	Propose a hypothetical procedure or experiment that could be performed to investigate a problem.
Teacher	
Defines	Provides a definition to students
Clarifies/Elaborates	Provides more information to further tune or make clearer a previous statement made by a student or the teacher.
Takes Votes	Counts how many students agree with a conception, a statement, a response, prediction, etc.
Compares/contrasts students' responses/explanations	Openly compares or contrasts different points of view expressed by the students
Driving students to achieve consensus	Compares and contrasts the responses of students, in discussion or visually (board/overhead).
Relates evidence and explanations (WTSF)	Connects evidence collected in class to scientific explanations.
Repeats/Paraphrases student words	Repeats verbatim with the omission or adjustment of only a few words the statement of a student. The paraphrasing or repeating does not change or contribute additional meaning to the student's statement.
Revoices student words by. incorporating student contribution into the class conversation, OR 2. acknowledging student contributions. Summarizes what student said.	Incorporates students' words into the flow of a classroom conversation by using selected students' words in an ongoing train of thought. This involves the teachers modifying or building upon the students' original meaning to further a point or to "toss" a comment back to a student with new wording to see if the teacher has grasped the student's original meaning.
Does comprehension checking (factual, open-ended questions)	Asks students questions of a factual nature, but not in a yes/no or fill-in-the-blank manner.

Coding Categories (Continued)

Teacher	
Does comprehension checking (monitoring students)	Halts the course of a discussion to see if students are "with it," e.g. Are you with me? Is this making sense?
Provides evaluative responses (correct/incorrect)	Provides students with evaluative responses that indicate or suggest the student is correct or incorrect. This may include responses that do not involve evaluative language, but that are contrary to what a student said with corrective intent, e.g. yes! Good! Or S: weight? T: mass.
Provides feedback: descriptive or explicit improvements	Provides students with feedback that is either descriptive of the student's response in a positive or negative way, or states explicitly how student performance can be improved
Promotes making sense	Promotes connections - Is this what you expected? Is there anything that does not fit? Does you hypothesis make sense with what you know?
Models how to communicate scientific knowledge	Enacts a method for scientific communication for students
Models process skill	Enacts a procedure for students
Waits for student to respond (wait-time)	Allows a noticeable amount of time to elapse after asking a question and before students respond, or before deconstructing or elaborating on the question.
Makes learning goals explicit	States or refers to the learning goals for the lesson, unit, day, or year.
Orienting Students/Setting a task	Sets a task to students or orients how to do a task
Makes explicit classroom expectations/standards	Makes explicit expectations for the classroom, such as behavior or management routines, or standards, such as how lab reports are to be written, etc.
Provides/reviews criteria	Makes explicit the criteria or standards by which students' performance will be evaluated.
Captures/displays student responses/explanations	Uses some manner of capturing and displaying students' responses, such as by writing them on the board or overhead, or having students write their responses on posters, which are then displayed to the class.
Makes connections to previous learning	Makes a connection between present learning to learning that occurred previously in the classroom context
Refers to nature of science	Makes a reference to the nature of science or the actions or real scientists, such as, for example, scientists make observations, it's important to test only one variable, etc.
Connecting topic with the real-world	Connects the present topic with a real-world application or example.
Sharing of students thinking (presenting to the class)	Promotes sharing students' thinking by making presentations of their work to the whole class.
Debating/discussing among students	Promotes students speaking to each other independently of the teacher in a whole-class setting.
Promotes students exploring their own ideas	Encourages students to address or look for information regarding issues of their own concern.
Explores students' ideas	Performs/demonstrates in response to a student's idea.
Interrupts flow of discussion or student responses	Cuts off a student making a statement, or interrupts or disturbs the course of conversation.
Asks for yes/no or fill-in-the-blank answers	Asks closed-ended, recitation-style questions of a yes/no nature, or where a fill-in-the-blank, clearly expected response is suggested.
Repaired questions-no opportunity for the student to respond	Asks a question that is rhetorical, or which the teacher does not provide students an opportunity to respond by asking another question, moving on, or answering the question
Other codes	
On-task BUT Off-Interest	
Inaudible Teacher	
Inaudible Student(s)	