EDUCATIONAL ASSESSMENT, *4*(1), 37–73

Strategies and Challenges to Changing the Focus of Assessment and Instruction in Science Classrooms

Richard A. Duschl
Peabody College–Vanderbilt University

Drew H. Gitomer
Educational Testing Service
Princeton, New Jersey

The improvement of science education in accord with the current science reform agenda requires the development of sophisticated instructional strategies that are grounded in a clear recognition of student understanding. We describe a pedagogical strategy, the *assessment conversation*, that helps teachers elicit student understanding and then use elicited and diverse student understanding as the instructional basis for achieving conceptual and reasoning goals in the classroom. We then illustrate the potential and challenges of using the assessment conversation through examples that have emerged from Science Education through Portfolio Instruction and Assessment (SEPIA), a project attempting to reform practices of assessment and instruction in middle school science classrooms. We conclude with a discussion of issues facing any substantial reform of science education

A goal of the science education reform agendas (cf. American Association for the Advancement of Science, 1993, *Benchmarks for Scientific Literacy;* National Research Council & National Academy of Sciences and Engineering, 1994, *National Standards for Science Education*) is to design curricula and associated instructional strategies that will develop learners' habits of mind to reason scientifically and engage in scientific inquiry. The assumption that students can do science entails an emphasis on two complementary sets of goals. The first is that

Requests for reprints should be sent to Richard A Duschl, Box 330, Peabody College, Vanderbilt University, Nashville, TN 37203

students learn the cognitive and manipulative procedures and methods of science exploration that generate data and evidence. The second is that students learn the skills of argumentation and of theory development and evaluation that link evidence to explanations. As L. Schauble (personal communication, April 1992) put it, "What we want is kids reasoning about the things they are exploring, and exploring based on their reasoning." This view acknowledges that a principle goal of science education is the development of thinking, reasoning, and problem-solving skills to prepare students to participate in the generation and evaluation of scientific knowledge claims, explanations, models and experimental designs (cf. Klahr & Dunbar, 1988; Kuhn, 1993; Metz, 1991; Schauble, Klopfer, & Raghavan, 1991).

The position to be developed in this article is that assessment activities in classrooms can help to achieve such goals and, more important, can provide information about progress toward these goals. Although many focus on the role of on-demand performance assessments to shape instructional dynamics and educational policy and goals (Resnick, 1993; Resnick & Resnick, 1991; Shavelson, Baxter, & Pine, 1992), our focus is on the role of assessment in shaping classroom activities, diagnosing students' ideas and products, and guiding teachers' decisions. More specifically, we are concerned with establishing classroom learning environments that facilitate the acquisition of information teachers can examine and use to help students learn how to do science. It is our thesis that science instruction improves when teachers are provided with curricula and instructional strategies that allow for frequent and ongoing assessment opportunities. It is also our contention that mastering these strategies is extremely complex, introducing significant challenges to the assumptions and methods underlying the current practice of the majority of science teachers.

The achievement of new educational goals, be they conceptual understandings, cognitive outcomes, or inquiry performances, ultimately involves problems of practice. Problems of practice are largely problems of appropriate curriculum designs and instructional dynamics. The primary challenge to science education reform is to have a set of operative instructional goals and practices that are consistent with the goals and practices of science education set out by reformers. Unfortunately, the conclusion to be drawn from much classroom-based research is that teachers' assessment of information related to cognitive goals is often ignored. Instead, what receives priority is information more frequently aligned with the activity goals of the classroom (Doyle, 1983; Sanford, 1987). When attempts are made to alter the activity structure of the classrooms and the decision making of teachers so as to increase cognitive considerations of the task environment, students often feel they are being placed at risk and strive through their actions to lower or renegotiate the cognitive demands to more familiar and less challenging task situations (Doyle, 1984, 1986b).

More recently, discourse analyses of science classrooms (Carlsen, 1991, 1993; Lemke, 1990) have done little to alter this general finding. The result in science classrooms is that there is a great emphasis on the activity structure of the lesson (e.g., time, materials, students groups) with less concern about the thematic structure of the lesson (e.g., concepts, background knowledge, evidence). One outcome of this imbalance is that teachers and students generate very different perceptions about the purpose of a lab, lessons, or activity and also different interpretations about what counts as the important content to be learned (Osborne & Freyberg, 1985). Despite teachers' verbalized intentions, students perceive that the simple doing of the activity is, in fact, the essence of science.

These types of student perceptions are not at all surprising, for teaching expertise is typically associated with management routines teachers adopt and effectively employ to handle the activity structure of the classroom (Kagan, 1992). It is out of this tradition that generic instructional frameworks—for example, the Hunter Model, Models of Teaching (Joyce & Weil, 1986)—take root as the foundation for activities and tasks. We propose an alternative approach in which assessment of the qualities of student work and of learners' developing thematic structures come into balance with management of activity structures. Research has made it clear that expertise in teaching is more than just management of activity; it involves effective execution of complex cognitive tasks (Clark & Peterson, 1986; Doyle, 1983; Grossman, 1992; Leinhardt & Greeno, 1986). The role of subject matter content and the social context of learning must be emphasized as well (Pintrich, Marx, & Boyle, 1993).

In brief, effective classrooms emphasize not only the management of actions, materials, and behavior, but also stress the management of reasoning, ideas, and communication. Such a shift, however, presupposes that teachers have access to information, making it possible to manage reasoning, ideas, and communication. Access to information and the skills and strategies to process and act on that information are the critical components of assessment-driven instruction.

Viewing assessment as intrinsic to the instructional process represents a position that, though discrepant with conventional practice, is highly consistent with the first principle of assessment—to make inferences about students that support useful decisions in educational contexts. The fact that most assessment practice has been of a summative nature and has had little impact on decision making within classrooms does not diminish the potential worth of such a direction (see also Baron, 1990; Wolf, Bixby, Glenn, & Gardner, 1991).

This article considers how assessment can both support and promote a fundamental shift in teaching, while describing the challenges raised by such a shift. We describe the *assessment conversation,* which is a specially formatted instructional dialog that embeds assessment into the activity structure of the classroom. The intent of an assessment conversation is to engage students in the consideration of a diversity of ideas or representations produced by class members and then to employ evidence and age appropriate adaptations of scientific ways of knowing to foster a dialog about what does and does not fit with the emerging thematic structure of the lesson. We then describe a curricular unit—the Vessels Unit—that has been specially developed to (a) foster students' representations of ideas, (b) facilitate assessment conversations; and (c) promote a portfolio assessment process. We have called this type of science learning environment a *portfolio culture* science classroom (Duschl & Gitomer, 1991; Gitomer & Duschl, 1995).

We next turn to the experiences of two teachers using the Vessels Unit. Focusing on the accomplishments and struggles of the two teachers, we describe the intellectual challenges teachers face when the assessment of reasoning, ideas, and representations is moved to the core activities of science classrooms. Finally, we speculate about the ramifications of this research for the reform of science education and for the successful implementation of performance-based assessments.

INTRODUCTION TO PROJECT SEPIA

Science Education through Portfolio Instruction and Assessment (SEPIA)[1] is an effort attempting to improve science education in middle school classrooms by having students develop scientific explanations, models, and experiments in the course of in-depth study of restricted conceptual domains. Significant changes to the curriculum that was in place prior to the project's inception include developing more authentic problem-based curricula to serve as the context for investigations and reasoning, placing a greater emphasis on reasoning to complement student investigations, adopting conceptual change teaching strategies, and integrating assessment strategies as a primary means for making decisions about and supporting implementation of instructional activities.

Project SEPIA was designed in consideration of emergent and convergent theories and practice in cognitive science, science education, instructional science, educational assessment, and history and philosophy of science (Duschl & Gitomer, 1991). From these diverse fields, two themes dominate. First, learning and progress in science requires the active and social construction of meaning. Second, the growth of scientific knowledge requires the acquisition of science-specific process skills, the development of science reasoning skills, and the acquisition and appropriation of a rich conceptual understanding of scientific domains.

Current cognitive science research supports the idea that learning requires the active construction of meaning by individuals working within a social context (e.g., Brown, Collins, & Duguid, 1989; Pintrich et al., 1993). In science, constructing

[1]Project SEPIA is a collaboration of the University of Pittsburgh and Educational Testing Service, working with the Pittsburgh Public Schools The project was supported by the National Science Foundation (MDR905574). The opinions expressed do not necessarily reflect the positions or policies of the National Science Foundation, and no official endorsement should be inferred.

meaning requires the development and application of cognitive strategies and heuristics that make it possible to comprehend patterns of information and evidence from nature, which are used to construct and evaluate scientific knowledge claims. Novak and Gowin (1984) argued that the presence of strategies and heuristics permits students to, in effect, learn how to learn in science.

Such learning strategies are developed through modeling and explicit teaching (e.g., Palincsar & Brown, 1984; Sigel, 1978, 1993). In Sigel's model, effective learning requires cognitive "distancing" from one's immediate experience. Effective questions or challenges by teachers create discrepancies that provoke cognitive activity which, in turn, promotes more complex, abstract, and developed cognitive representations. Eventually, effective learners internalize and apply these distancing strategies as a matter of course. Therefore, constructivist teaching implies that teachers work with students so that they may develop an effective set of learning and reasoning strategies. Such teaching must include processes and tasks that help students actively construct meaning from their experiences (e.g., Resnick, 1987). Constructivist teaching, however, does not imply a laissez faire approach to teaching. Good teaching guides student construction through careful selection of learning experiences, questions, tasks, and so forth and does so in the service of established institutional, cognitive, and epistemic goals. A move to constructivist teaching represents a move from questions of "What do I give students to develop an appropriate understanding?" to questions of "How can I help students construct appropriate understanding?" (cf. Cobb, 1994; Driver, Asoko, Leach, Mortimer, & Scott, 1994).

What is an appropriate understanding in science education? There is an emerging consensus that argues for more in-depth study of topics so that students have opportunities to engage in instructional tasks that develop epistemic, strategic, and conceptual knowledge of science domains (American Association for the Advancement of Science, 1993; National Research Council, National Academy of Sciences and Engineering, 1994). Extended study is encouraged for several reasons. First, it gives students an opportunity to understand the complex interrelations of concepts within scientific domains. This approach is in reaction to a history of practice that has tended to emphasize the mastery of discrete facts. Second, developed scientific reasoning skills can occur only in the context of conceptually rich explorations. Kuhn (1993) suggested that science instruction can be conceived of as the interplay of science as exploration and science as argument. In espousing the "doing" of science, it is not sufficient to simply have students engaged in hands-on investigations. It is also necessary that students engage in the forms of reasoning that are intrinsic to scientific activity and science as a way of knowing (Duschl, 1990; Hodson, 1992). The move to hands-on science instructional approaches has led to a dominant emphasis on the investigative activity itself, typically divorced from any reasoning around the activity (Osborne & Freyberg, 1985; White & Gunstone, 1992).

Successful science education depends on students' involvement in forms of communication and reasoning that model those of the scientific community (e.g., Gee, 1994; Roseberry, Warren & Conant, 1992; Schauble, Glaser, Duschl, Schulze, & John, 1994). Scientific inquiry requires immersion into the language, culture, and tools of scientific activity, a language and culture grounded in certain logical and epistemological assumptions that make science different from other disciplines. Just as the language and culture of France is different from the language and culture of the United States, and the tools of a plumber are different from the tools of a physician, so too are the criteria for evaluating the status of knowledge claims and explanations in nonscience disciplines different from the criteria used in the sciences. Science has particular ways of considering evidence; generating, testing, and evaluating theories; and communicating ideas. A goal of science education is to help students participate in all the practices of the scientific community's culture.

The goal of Project SEPIA is to develop a classroom culture in which the previously mentioned goals of science education can be realized. We have called this educational model a *science portfolio culture* (Duschl & Gitomer, 1991; Gitomer & Duschl, 1995) because assessment, particularly classroom portfolio assessment, is a central component of this educational model. The portfolio serves as a repository of students' ideas and findings, which become the basis for classroom discourse and activity. If instruction must allow students to voice their understanding and teachers to recognize and act on this understanding in order to effect change in student's scientific conceptions, then the portfolio represents the place where students can represent their understanding. The portfolio culture classroom represents the place where teachers facilitate learners' processes of understanding through the continuous interplay of assessment and instruction. Teachers assess, and help students assess, these representations in order to recognize student conceptions, strategies, or language use, all as a basis for guiding instructional activity.

A portfolio culture is a part of assessment reform that involves more than simply designing better instruments to measure and report performance. Assessment has the potential to be the unifying concept of educational reform, leading to integrated practice in which the boundaries of curriculum, instruction, and assessment blur (Baron, 1990; LeMahieu & Foss, 1994; Resnick, 1993; Wolf et al., 1991). In conceptual change teaching, for example, the assessment of students' initial understanding suggests the instructional approaches that might be most effective. When the development of a science reasoning strategy is added to an instructional unit—for example, reasoning with arguments—then the evaluation of premises, evidence, and argument structure become an additional essential dynamic of the assessment process. Establishing educational goals and determining progress toward those goals, most especially within a complex constructivist framework, requires assessments that can provide information about students' progress on the dimensions cited previously—language, culture, and tools of scientific activity. Perhaps most significantly, the internalization of goals and monitoring of attainment of goals by students is the ultimate objective of a constructivist classroom. Thus, assessment is a process that ought to be seamlessly integrated with and pervasive in the instructional activities of the classroom.

Clearly, the actions and decisions of the teacher are paramount to the success of this enterprise. To assist in the implementation of the portfolio culture, we developed a set of science criteria for teachers and students to use during the completion of instructional activities and tasks and an instructional discourse pattern we call an assessment conversation.

The goals of the portfolio culture are embodied in scientific criteria. These criteria are an articulation of what is valued in the production and evaluation of scientific ideas. Operative criteria in many science classrooms include the recall of discrete facts and the successful execution of experimental procedures, whereas the SEPIA criteria for considering student explanations given in Table 1 focus more on student's reasoning and communication, characteristics of performance that are critical for successful engagement in the scientific enterprise. The SEPIA criteria are designed to be publicly shared and recognized in the classroom and to become, in effect, the currency by which classroom ideas are considered. These criteria also transcend particular topics or grade levels. These criteria are generalizable to all occasions when scientific ideas are to be examined.

The principles of SEPIA are best realized in several prototype curriculum units that have been developed in collaboration with project teachers. Students are presented with authentic problems and then led through a sequence of investigations, demonstrations, discussions, and reports, a process that develops both a conceptual understanding of a domain as well as specific reasoning strategies common to science as a way of knowing. For example, in the Vessels Unit, the context in which the research reported here was carried out, the problem is to design a vessel hull from a 10 in. by 10 in. square sheet of aluminum foil that maximizes load-carrying capacity. The problem requires the application of the physics of flotation and buoyancy to an engineering design problem and the development of a causal explanation. The student must relate design features (e.g., the height of vessel sides and surface area of the vessel bottom) to vessel performance and, ultimately, to buoyant forces, buoyant pressure, and water pressure.

The class works through a series of iterative cycles in which some form of exploration is conducted, either through demonstration or investigation, and students represent their understanding in some form (e.g., written, oral, graphical, or design product). Once students represent their understanding, the SEPIA model calls for an assessment conversation. These conversations are structured discussions in which student products and reasoning are made public, recognized, and used to develop questions and activities that can (a) promote conceptual growth for students and (b) provide assessment information to the teachers.

TABLE 1
SEPIA Criteria for Guiding Formation and Assessment of Students' Explanations

Relationships	What goes together? How do they go together? Is there a name we can give to the relationship? Is there anything that does not belong? How are things alike?
Clarity	Is it clear? Does it tell what you want it to tell? Will it be clear to someone else?
Consistency with evidence	Is the statement supported by observations? If so, what? Is it supported by the observations of others? If so, what? Is the statement consistent with lab data? If so, what data? Can you identify evidence from nature that supports the statement? Does your statement reflect the data?
Use of examples	Can you give an example? Is it a good example for this purpose? Is there a better example for this purpose? Can you think of an original example?
Making sense	Is this what you expected? Are there any surprises here? Is there anything that does not fit? Does your hypothesis make sense with what you know? Can you predict what will be the outcome?
Acknowledging alternative explanations	Is there another way to explain this? Is your explanation or hypothesis plausible—can it happen? What does this explanation say that the other doesn't?
Elaboration of a theme	Is this term related to something we did before? Is it familiar? If so, how? Is it related to anything you did in another class?
Accuracy	Is the statement consistent with other information on the same topic? How does the model compare with other models? How does it compare with other representations?

Note. SEPIA = Science Education through Portfolio Instruction and Assessment

The assessment conversation is an idealized model of teaching practice. It requires a set of teaching strategies and assumptions that are quite different from those of traditional practice. Gitomer and Duschl (1995) described some of the challenges to successful implementation of the assessment conversation. One of the goals of this effort is to examine how the actual implementation of assessment conversations changes as teachers become more experienced with an educational innovation.

The assessment conversation has three general stages, presented in Table 2. The first step is to *receive* student ideas. This requires that students be allowed to represent their understanding as fully as possible. To this end, SEPIA instruction incorporates detailed writing, drawing of annotated pictures, linkages between drawings and writings, construction of storyboards, and many other techniques that allow students to "show what they know."

Once students have represented their understanding, it is the responsibility of the teacher to *recognize* the ideas in the classroom in relation to unit or lesson goals. Inevitably, there will be a diversity of ideas. In traditional classrooms, this diversity is quickly constrained through an appeal to find the "best answer." In a portfolio culture assessment conversation, diversity is made public and resolved through a discussion that is governed by scientific criteria related to language, culture, and tools. In recognizing the diversity of student work, teachers need to select work that differs on dimensions relevant to the conceptual, cognitive, and epistemic domain(s) being explored. They then must lead the class through a discussion in which the critical differences in student representations and reasoning are highlighted.

Once the diversity is public, the teacher can *use* the diversity of ideas as a basis for achieving a consensus view in the classroom. The teacher does not simply dismiss student ideas through appeals to some authority (e.g., the teacher or text), but uses classroom discussion to determine which group of students' representation or reasoning satisfies the criteria more than other students' (e.g.. is more consistent with the evidence). The teacher's role is to pose questions and facilitate discussion that results in a consensus view acceptable to the classroom. A final use of student

TABLE 2
Stages of the Assessment Conversation

Stage	*Description*
Stage 1—Receiving Information	Individual or group efforts on specialized tasks that by design bring about among students a diversity of responses and range of representations or ideas Teacher and students make explicit and publicly display via posters, presentations, charts, overheads, and so forth the diversity of students' efforts, representations of meanings and understandings, and performances on the tasks
Stage 2—Recognizing Information	Teacher examines critically and makes an appraisal of the diversity of student efforts, meanings and understandings, and performances and selects according to conceptual goals and employing criteria Teachers and students work toward a synthesis of what comes to count as or stand for appropriate efforts. meanings and understandings, and performances employing SEPIA criteria
Stage 3—Using Information	Applying what has been learned to an evaluation of previous efforts, meanings and understandings, and performances or to the design of an investigation for advancing efforts. meanings and understandings, and performances in the present domain of inquiry

understanding is to entertain how the accepted view generalizes to new and different situations.

THE VESSELS UNIT—A PROTOTYPE TO EXAMINE ASSESSMENT CONVERSATIONS

The goal of any SEPIA unit is to develop learners' conceptual understanding of the science domain under study and to enhance students' capacity to use and evaluate the cognitive and metacognitive skills needed to reason in the context presented—for example, resolve an engineering problem and construct a causal explanation. Frequent feedback on meaning making and reasoning is a vital component of SEPIA units. So too is allowing students multiple ways to express or represent knowledge claims. Creating effective communication between students and between teachers and students is critical to an assessment-driven learning environment. Given this expanded instructional agenda, a decision has been made to constrain the conceptual domain, that is, number of concepts, presented in SEPIA units. Let us now turn to a description of one of these prototype units—the Vessels Unit.

In order to involve students in the doing of science that is meaningful and motivating, it is necessary that curriculum units focus on problems and questions of some consequence for the students. If students engage in work without purpose, there is virtually no likelihood that thoughtful consideration of scientific ideas and reasoning will result. A goal of the Vessels Unit, then, is to engage students in (a) a consideration of both the syntactic and semantic structures of scientific knowledge claims, (b) the use of cognitive and metacognitive strategies relevant to the selected problem space and to thinking scientifically, and (c) the accurate presentation and representation of scientific knowledge claims and forms of discourse.

The instructional sequence for the Vessels Unit is an intermingling of investigations, experiments, demonstrations, and assessment conversations as well as presentations of students' ideas and products. The unit is partitioned into four parts. In general, Part 1 begins by acquainting the students with an authentic problem they consider worth solving. Part 2 allows the students to test their initial ideas regarding the problem and build a conceptual framework of knowledge necessary to solve the problem. Part 3 is designed to help the students consider their results and evidence and perform additional experimentation. In Part 4, the students carry out the final tests and formally make their findings public. An outline of the entire instructional sequence for the Vessels Unit is presented in the appendix.

The conceptual goal of the unit has two parts or objectives. One learner outcome is the development of a causal explanation for flotation. The other learner objective is the construction of a reasoned design for maximizing the carrying capacity of a square sheet of aluminum foil, that is, stating how the design features of the vessel relate to the vessel's performance. A representation of this conceptual domain is presented in Figure 1. Again, the four parts of the Vessels Unit have been prepared to facilitate for teachers and for students the appropriation of the concepts, reasoning skills, and evidence to achieve the conceptual goal of the unit.

Part 1 introduces the students to an authentic problem through a letter from the City of Pittsburgh, presented in Figure 2, stating the need to build a fleet of vessels to haul construction materials. The students are specifically asked to design vessels with features that maximize each vessel's capacity to carry a load. The letter outlines the problem and also the expectations of student work. See, for example, the four expectations set out at the end of the letter. The letter clarifies that the goal is not simply to build a prototype, using a specified amount of aluminum foil, with the largest capacity, but also to display a full understanding of the principles that effect the load-carrying capacity of a vessel and be able to communicate the results or position to others.

Students complete the first portfolio entry by restating the problem as they understand it and listing the basic information needed to solve the problem. This information is then used by the teacher to conduct an assessment conversation. The instructional dialog helps the students to consider and focus on what is being asked of them and helps the teacher to receive information about the students' levels of comprehension.

The purpose of Part 2 of the Vessels Unit is to provide students with the opportunity to test their initial ideas and conceptions about what makes a vessel float and on the building of a vessel with features that maximize its carrying capacity. Items placed in the portfolio include vessel designs and data records of vessel performance. These items help the students to collect not only information regarding the vessel and its performance, but also critical evidence that will be used later to assist them in understanding the differential pressures explanation for flotation. For example, students are asked to draw the appearance of the vessel with respect to the water level as it takes on greater load. This item can then be used to help the students understand that buoyant forces for floating objects change with depth. Such evidence assists students in understanding that water pressure increases with depth and that high vessel sides are one important variable in maximizing the vessel's load-carrying capacity.

The purpose of Part 3 is to have the students apply the knowledge and evidence from Part 2 to come up with a solution to the original problem—design a vessel with features that maximize its load-carrying capacity.

Students begin Part 3 by reviewing the purpose of the original task and the knowledge and evidence they have accumulated up to this point. Then, the students are asked to think about ways they can test via experiments the features of a vessel that maximize its load-carrying capacity. The students test their hypotheses through controlled experimentation. Results are recorded in the "Student Report of Inves-

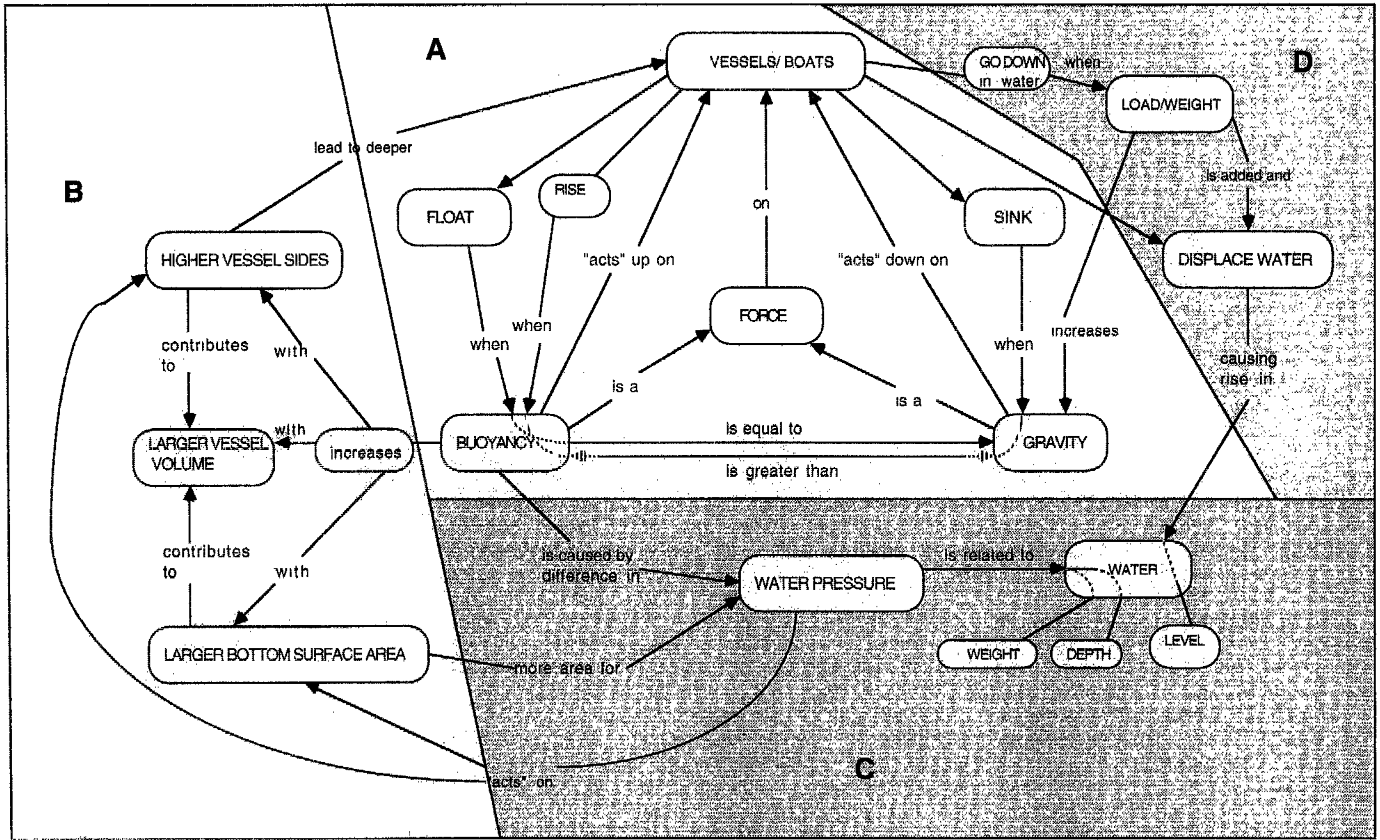


FIGURE 1 Conceptual goals of the Vessels Unit. The letters A, B, C, and D correspond to specific domains of concept development within the unit but do not imply an order of presentation.

tigation." The purpose of these investigation reports is to help students realize that there is a trade-off in maximizing the volume of the vessel (i.e., either higher sides and smaller bottom surface area or lower sides and larger bottom surface area). The ideal vessel is one which makes a compromise between the two variables such that the volume is maximized.

Part 4 is the culmination of the inquiry process. It begins by allowing the students to construct their final vessels using their acquired knowledge and understanding from Parts 1, 2, and 3. After constructing the second set of vessels and testing them, the students prepare their formal presentations following the guidelines outlined in the original letter. The purpose of these presentations is to allow the students to communicate their results and display their mastery of unit content and processes. The presentations also create an atmosphere of success for students. They allow the students to feel that even though they worked hard, the work itself was interesting; they were involved in the educational process; and they now have a product that can be presented to their classmates, teachers, parents, and other involved individuals. The Vessels Unit takes approximately 4 weeks to complete. Note in the appendix the frequency with which assessment conversations occur throughout and the amount of student work that is not only placed in the portfolio, but is used as the basis for further development and demonstration of understanding.

Instructional activities and tasks for the Vessels Unit have been prepared so that students' ideas can be made public, explored, and refined. Student activities—portfolio entries—are designed so as to facilitate teachers' receiving information about students' ideas. The significant challenge posed to teachers is that assessments need to occur on multiple fronts. Students need feedback on their developing understanding of the core science concepts, the characteristics of the emerging science explanation, the reasoning they employ when considering evidence and relating it to explanations, and the ways in which they choose to represent and report scientific information and knowledge claims.

The assessment conversation is a critical pedagogical tool to facilitate this complex assessment process. Following the completion of an instructional activity like reading the letter, students are given a portfolio item (PI) to complete. The PIs ask students to draw or write (or both) an interpretation or representation. A good example of a PI and how it is used to stimulate an assessment conversation is provided in Figure 3. In this PI, students are asked to describe, after they have carried out the task, the forces acting on a cup as it is slowly pressed into a tub of water. First they are to draw and then prepare a written explanation. Two student samples are provided. In the first, labeled A, we can see how the student has represented the strength of the buoyant force decreasing with depth by using smaller and smaller arrows and the force of gravity increasing with larger and larger arrows. We can also see that the arrows are labeled G and B with a greater than sign (>) between them and that the written explanation is minimal. In the second student

City of Pittsburgh

Sofia Maslow, Mayor

Department of City Projects
Land Use Planning Program
1600 W. Rosedale St.
Pittsburgh, PA 15219

Office of Asst. Planning Director - Peter Remraf

RE: Request for Design Plans

Dear Applicant:

The purpose of this letter is to provide you with information for submitting a bid to the City. I am pleased to learn that you and your staff will be submitting hull design plans to be used for building a fleet of river-going vessels. In order to assist you with the development of your design plans, let me tell you how we intend to use the vessels.

The City intends to build office complexes, apartments, shopping centers, marinas, and playgrounds on Herrs Island. Herrs Island is an island on the Allegheny River up the river from the 16th Street Bridge. In fact, the 31st Street Bridge directly crosses Herrs Island. We think the total project will take 10 years to complete. The best way to deliver construction materials - sand, cement, lumber, bricks, cinder blocks, pipe, etc. - is using the river. The bridges that go to the island will not hold up after 10 years of traffic from heavy truckloads of materials. Therefore, we feel it is in the City's best interest to build its own fleet of vessels.

The contract for supplying the construction materials has been awarded to Best Construction Materials Supply Co. which is located on the Monongahela River. The main function of the vessels will be to take materials between the Mon River and the Allegheny River. Attached is a print of a map of the Point which shows the locations of Best Construction and Herrs Island.

As you can tell, the ability to carry materials is important. The successful plan will be one that explains how a vessel hull should be designed so it stays afloat while carrying the most load. In order to make the competition fair, we are asking all bidders to design hull models using aluminum foil that is the same size.

After completing your investigation, the packet of information you submit to the City should contain the information and materials in the items listed below. Only complete packets will be considered. We want to hire the firm that can design the best hull. But the City must have confidence that the designers understand and can explain why a vessel will float and carry a load. Without this explanation, the City can't be certain the design model you submit will work.

Design Packet Items

1. *A sketch of the vessel hull.*
The sketch should be neat and have the height, length and width of the vessel labeled.

2. *A scale model of the vessel.*
The scale model should be made of aluminum foil. It will represent the hull of the vessel. It should be made as best as you can to look like the sketch you submit.

3. *Sketches of the vessel hull in water with and without a load.*
These two sketches should be side by side on the same piece of paper. Using arrows, science terms and the names of forces, label the sketches to explain the forces that keep the vessel afloat. Please mark the water line.
These sketches are a very important part of the design packet. We want to hire the firm that understands and can best explain why vessels float.

4. *A report of tests and results.*
Please list the tests, experiments, and investigations you performed. Then provide the a report of results. For example, what is the mass in grams (g) that it took to sink your vessel. Include in your packet any tables, graphs, or test design sketches you think will demonstrate you have thought through the problem carefully.

Good luck!

Sincerely,

Peter Remraf, Project Manager
Herr's Island Development Proposal

FIGURE 2 Introductory letter for the Vessels Unit.

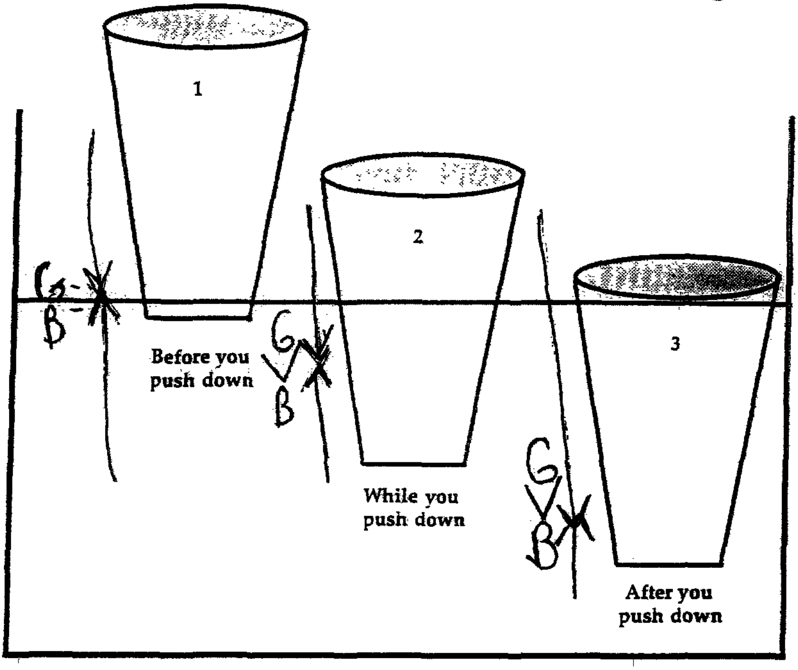
Portfolio Item 6

Nam

Datc

Teac

Drawing A



1. Draw arrows on the drawing above to describe what you feel.
2. Now write in your own words what you felt while pushing the cup down into the water.

It was hard to push down

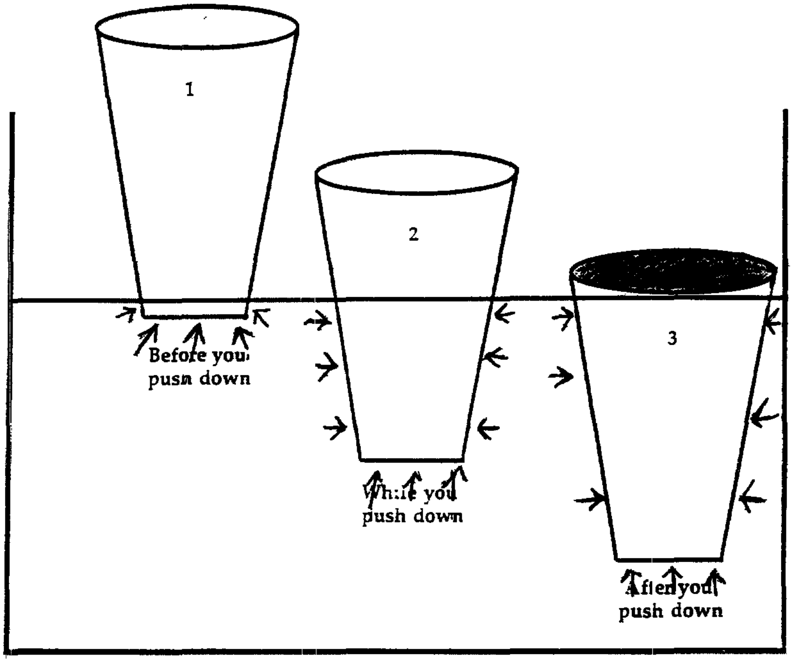
Portfolio Item 6

Name

Date

Teach

Drawing B



1. Draw arrows on the drawing above to describe what you feel.
2. Now write in your own words what you felt while pushing the cup down into the water.

Number 1 push up a littl.
Number 2 push a littl mor than
number 1. and 3 was very hard
to push down.

Project SEPIA - Fall 1992

FIGURE 3 Sample student work on Portfolio Item 6 sample—B—we see a different representation. In particular, we see that the student has put arrows on the side of the cup and kept the number of arrows at the bottom the same number and size.

From these two drawings alone, one could address the notational conventions of representing forces with arrows and the magnitude of the forces by the length or number of arrows. Drawing A does this but Drawing B does not. Drawing B provides side arrows, and it can be used to explore with students one of the key concepts of the unit—water pressure increases with depth. This is what makes the height of the sides in the design of the vessel an important variable to consider. One could ask if the side arrows represent the same thing as the arrows drawn at the bottom of the cup. If the arrows are to represent the buoyant forces acting on the cup, all arrows should be pointing upward. If the arrows are to represent the water pressure acting on the cup, then the arrows at the bottom should be larger than those at the top and some indication should be made that the pressure acts in all directions. Hence, an asterisk (*) notational form might be more meaningful. Similarly, the arrows at the bottom of the cup should reflect the "very hard to push down" in the written statement. Back to Drawing A, we could ask why, if it is harder to push down at Position 3, the B-arrow in this position is the shortest. Finally, we could begin to explore with students their understanding of what causes the buoyant pressure and buoyant force to increase as a floating object in water achieves ever lower depths. It is differences like this that spawn assessment conversations that address critical issues of science learning.

Sharing with students the multiple ways they have presented and represented scientific evidence or ideas makes it possible to provide feedback on the quality of evidence and ideas put forth by class members. In addition, it is important to note that it also becomes possible to provide feedback on the presentation and representation of ideas themselves. Out of the conversations around student work among students and teachers, information about how students are reasoning, using evidence and experiences, and constructing explanations and arguments becomes visible and tangible.

Incorporating this assessment-based instructional strategy is not a simple matter, however. The challenges of teaching and of managing a classroom learning environment are significantly altered when one is asked to receive, recognize, and then use student-generated information for the purpose of conducting assessments on a frequent basis. In the next section, we present three examples of two different teachers teaching Vessels Unit lessons. The first example demonstrates the successful implementation of the first stage of the assessment conversation and the challenges teachers face in realizing the final two stages. The focus of the second example is on the successful implementation of all three stages of the assessment conversation. The third example comes not from an assessment conversation, but is a different discussion between teacher and researchers that highlights some of the tensions that surface through this kind of interaction. We argue that a central strategy and significant teaching challenge to the successful implementation of assessment-driven instruction is the need to engage and sustain a learning environment that emerges out of students' personal efforts, products, and ideas.

THE CHALLENGE OF ASSESSMENT CONVERSATIONS

Project SEPIA's three stage assessment conversation has been a significant challenge for Project SEPIA teachers. Curriculum units like the Vessels Unit have made it possible for teachers to receive students products and ideas (Stage 1). The teachers have been quite inventive with ways to display publicly the diversity of representations. Implementation of the Vessels Unit as designed represents a dynamic shift from the hands-on, teacher-directed curriculum used in most middle school science classrooms. A fundamental difference is the shift from instructional activities and tasks that had all students produce the same response or answer to activities and tasks that encourage students to produce a diversity of representations and responses as answers.

Although some aspects of a portfolio culture classroom have been implemented and refined with success, other critical elements have been sources of concern for both teachers and researchers. In this section we use examples from classrooms to describe these successes and challenges. In particular, we want to focus on two issues that have been the object of this project's attention and that represent perhaps the most significant challenges to successful transformation of classrooms. The first issue concerns the use of assessment conversations as an instructional vehicle to facilitate students' meaning making and reasoning. The second involves the shifting of responsibility and ownership for conceptual and cognitive development from teachers to students.

The experiences of two teachers (pseudonyms George and Martha) are presented as contrasting illustrations of how these two issues play out in classrooms. In Example 1, the focus is on an opening unit lesson by George in which he successfully conducts the first stage of an assessment conversation but then goes no further. In Example 2, we present Martha's success at navigating students through the design and execution of experiments (PI 7 in Part 3) that are intended to determine which vessel design variables—height of sides, surface area of the bottom, shape, thickness of foil (folding)—affect the load-carrying capacity of the vessel. As a contrasting case to Examples 1 and 2 for the purpose of highlighting important strategies and challenges in portfolio culture classrooms, we present Example 3, in which George relates his unsuccessful execution of PI 7. Data for the preparation of the examples are taken from classroom videotapes, field notes, transcriptions of classroom discourse, and students' work.

Example 1—Receiving but Not Recognizing Information

A key feature of assessment-driven instruction is making public the efforts and ideas of students. The use of student work in assessment-driven instruction should function at two levels. First, the discussion of student work should provide opportunities for assessment conversations that examine, develop, and evaluate students' understanding of science concepts and processes. Second, the discussion of student work should provide opportunities for assessment conversations that examine, develop, and evaluate students' reasoning and explanatory skills. Although SEPIA teachers have been uniformly successful at eliciting student representations and understandings of individual or small networks of concepts, the use of student work to explore students' representations and understandings of the unit's conceptual goals (e.g., constructing a causal explanation and a reasoned design) has been less consistent.

Consider the following classroom situation as an example. George is starting the Vessels Unit on buoyancy and flotation with his sixth-grade students. As a first activity prior to the reading of the letter, he has the students represent what they know about the flotation of boats via a drawing accompanied by a brief written explanation. The specific directions are to draw a boat and then write a sentence that explains why it floats. As the students engage in the task, George begins to circulate through the room with a clipboard. He stops from time to time to examine student work, make supportive comments, and on occasion jot down some information on his clipboard.

At this juncture of the lesson, George is exhibiting, for him, a very new pedagogical practice designed to capture students' diverse ideas. He is addressing successfully the first stage of an assessment conversation by engaging students in an activity that develops a diversity of ideas. His intent, which he announces to the students during the introduction of the lesson, is to have some of them come to the front of the classroom and display their drawing and explanation on the overhead projector. From the lack of student questions about what to do, one can infer that this practice is not new to the students and it is treated as a normal instructional practice. George has been incorporating this come-to-the-overhead form of publicly displaying student work for the last year or so.

The first student called to the overhead projector draws a tugboat pulling a barge and writes that it floats because of the engine. The second student draws a canoe and writes that it floats because of its shape. The third, and final student, draws a submarine and states it floats because of the air inside. Finally, George himself draws a pontoon boat with a canvas cover on it and asks the students to come up with an explanation for why it floats. One student suggests that it floats because of the air pushing up on the cover, sort of like on a sail. George provides the explanation that there is air in the pontoons.

George has surveyed the diversity of students' responses to the task and selected three diverse representations of boats. Furthermore, we see that he has established as routine an instructional practice that is student-centered and that provides for the public display of student work. George has successfully implemented the first stage of an assessment conversation by having students engage in a task that leads to a diversity of responses and then by making the diversity of student ideas explicit and public.

But he did not seize on the opportunity to extend the lesson that day, or subsequently, into the ensuing stages of an assessment conversation that ask students to judge the adequacy of the explanations given. In turn, then, George has not provided students with the opportunity to engage in either the cognitive or argumentation processes of science. The opportunity to address the epistemic and cognitive orientations of learning science are bypassed. Though George has engaged students in the exploration of different boats and explanations for why they float, he has not engaged students in the argument of which explanations are better explanations for boats floating. Thus, although the students are comfortable with the classroom practice of presenting and describing their knowledge claims, they are not being introduced to the strategies and rules for judging the adequacy of knowledge claims. There is neither reference to the student-generated concepts of flotation nor any examination of the diversity of meanings used by students in the class. Missing is the second stage (and ultimately the third) of the assessment conversation.

What might have taken place if there had been a public recognition that in addition to the three diverse vessels drawn, there were also three equally diverse representations for why the vessels float—engine, shape, and air? Though George himself recognized the explanatory diversity, he did not attempt to have students reflect on the differences they presented about what they think made boats float. Let's consider the first and second students' explanations. The first student says the boat floats because of the engine, whereas the second, drawing a canoe, states it is the shape. But the second drawing is also of a boat without an engine. It is not hard to imagine having students compare these two responses in paired groups and determining from their own conversations whether or not an engine is needed.

When we consider the explanation of the third student's drawing, a more compelling and intriguing notion of flotation presents itself. The drawing is of a submarine, and it, according to the student explanation, is floating in the water. The two other explanations involve floating on the water. Neither during this lesson nor in those to follow does George relate back to these competing student explanations. So, in addition to the absence of the later stages of the assessment conversation, George has also missed the opportunity to explore with students, using their own work, a critical conceptual goal for that unit—namely, that the pressure of water changes with depth. It is precisely these kinds of "floating in–floating on" type conceptual conflicts that emerge from the display of student work and that teachers need to learn how to (a) recognize and (b) attend to and manage in a science classroom. Perhaps even more important is the need for the students to be given opportunities to develop, and have modeled for them, the habits of mind used to examine and evaluate knowledge claims.

Example 2—Making Students' Efforts Central to the Lesson

A criticism of school science is that students often engage in laboratory activities without any experimental rationale. A principle use of the assessment conversation is to define and then explore with students the rationale for particular experiments. Example 2 focuses on Steps 7 and 8 in Part 3 of the Vessels Unit (see the appendix). The lesson sequence is designed to entertain the different hypotheses proposed after an examination of data obtained from the first testing of the vessel's carrying capacity. Next, students are asked to suggest and design experiments to address each of several hypotheses in a clear and "fair" way. The students then carry out the experiments. One group is testing the effect of the height of the sides. The students build two vessels that have the same shape and the same size bottoms but different heights of sides. Implementing similar control of variable techniques, other groups are testing the shape, size of bottom, and the effect of folding or not folding the foil.

Martha approached this segment of the Vessels Unit with an innovation that, in the end, established a strong motivation for focusing students efforts on both the task of writing the experimental design and then on the analysis of the results of the experiment. The innovation involved having students prepare and record the steps of the experiment for another group to follow and execute. This may appear to be trivial, but in fact it represents a significant contribution to the creation of a portfolio culture. Through this single instructional decision, students are now being asked to reason through and then communicate clearly to others the steps that need to be taken to do the experiment:

M: What we did on Friday, Karen, you weren't here. We switched, Group 1 did Group 4's experiment and Group 4 did Group 1's experiment. Group 2 did Group 3's experiment and Group 3 did Group 2's experiment .... You wrote an experiment for folding and not folding, but your group did the experiment for the shape of the boat .... Group 2 followed your directions and did your experiment that you guys wrote. Okay. Alright. Let's talk a bit about this. Ah, I would like someone from Group 1, Group 1 tested the height of the sides. What did you find out, Jess? What did you learn on Friday?

Moving group by group, Martha reviews with each group what it is they did and what it is they found out. She then moves into the next segment of the lesson.

> M: Okay. I would like to talk about all of this information for a few minutes .... So what I would like to look at first would be the "height of the sides" group. That is what they were testing, that is exactly right. Why don't we look at their four boats, okay. They made four boats, now Group 4 didn't start out by having to make four boats. We decided as we started to do this experiment to make four boats—okay two folded and two unfolded. We will put their folded on one side and we will put their unfolded on the other side.

Martha takes the models each group made to carry out their experiments and tapes them to the front board. Written inside each of the model vessels on a piece of tape is the number of metal washers the vessel was able to hold. This excerpt indicates that, 3 weeks into the Vessels Unit, Martha is implementing instructional steps that build out of students' efforts and products. It is clear that she is sustaining the critically important first stage of the assessment conversation by receiving information from students. Then, and only then, does she set out to have students reflect on the content and accuracy of their responses.

> M: Okay, Jess, could you tell us which one of these boats of these two held the most?
> S1: This one and this one.
> M: This one and this one? Why do you think, let's look at these two boats, why do you think that this one held more stuff than this one.
> S1: It had more space.
>
> ~~~~~~~~~~~~~~~~~~[represents break in discourse pattern]
>
> M: What about these two? Which one of these held more, Jesse? This one or this one?
> S1: The big one.
> M: The bigger bottomed one. Okay. The same reason?
> S1: Hmm-uhmm (yes).
> M: Now, let me ask you about this? Which one of these do you think held more, this one or this one? [Pointing to the next set of two vessels.]
> S2: The bigger one.
> M: Why do you think the bigger one
> S2: Because there is more space inside.
> T: Because there is more space inside, okay, alright. So does the height of the sides matter?

S2: No.
M: They don't matter!
S3: Yes they do but not that much
M: Hey, you just told me two different things. I don't understand. Can you explain to me what you mean? I am not quite understanding what you mean.
S3: Okay, if you just put a piece of foil on the water, it will sink but if you have little side on it like that [pointing to the model on the board] and then you have like a little sides it will hold more.
M: Okay, so the sides matter.
S3: Yes.
M: You just told me no, they didn't.
S3: They do but not that much.
M: They do but not that much. What do you think Monica? What do you think about size? You said no she was wrong. Do you think that the height of the sides matters a lot? Why?

We see from this excerpt that the focus of the lesson is on the efforts and products of the students. The discussion of the science concepts emerges out of this student-centered context, and it can be seen that students do not think that the height of the sides contributes much to the performance of the vessel. What is emerging from the assessment conversation is a picture that the space in the vessel has more to do with the size of the bottom and less to do with the height of the sides. ("So, does the height of the sides matter?" "No.") Martha has received information from the students that the students are considering only one of the two variables that contribute to determining the volume of the vessel. In her own word, Martha is "weaving" together the ideas of the unit but she is recognizing that the reasoning of the students is incomplete. Martha has received information and then recognized a problem. Next she conducts a demonstration to put the students' ideas about height of sides to a test.

M: So, sides I think, are we saying that they are something to consider? Should we consider sides when we make a boat?
S1: No.
M: You shouldn't consider sides when you make a boat?
S1: [inaudible]
M: So, I shouldn't consider sides when I make a boat?

~~~~~~~~~~~~~~~~~

M: So, if I take a flat piece of foil and put it on the water like that, what is going to happen?
S1: It is going to float.
~~~~~~~~~~~~~~~~~

M: It is going to float but what is going to happen when we start putting washers on it?

S1: It is going to sink when we put the first one on.

M: Well, maybe not the first one. Let's try it and see what happens? Do you want to try it? [Teacher moves to the side of the classroom to get a tub.]

S1: I think that it is going to sink.

M: Well, let's see because Jesse's made the [comment about sides] he doesn't think that they are important. [Students move to the front of the room around the tub and conduct the experiment with a flat piece of foil placed on the water. It holds twelve washers and then sinks.]

~~~~~~~~~~~~~~~~~~

M: Alright, so Jesse, let's talk about this Jess and everybody else. Alright, Jesse [teacher moves to the front board where the two folded and two unfolded vessels are displayed], this one held eighty right? This one sixty, this one held forty, this one held twenty-five, and that one held twelve [pointing to the flat vessel just used in the demonstration]. Now Jess, I want you to think about this, do you want to change your mind, do you think that sides matter? [Student nods yes.] Yeah, I think that sides matter.

Nested experiments require two significant alterations of classroom practice. First, students are designing the experiment, not simply carrying out a prespecified design. Second, the time given to exploring a topic area is significantly increased, for students in this proposed environment are not only learning a given conceptual area in greater depth, they are also spending more time in developing an understanding of experimentation. Taking the time to revisit issues to better understand a conceptual terrain is not standard practice in most middle school science classrooms. Martha's success on this section of the unit is due in large part to the fact that she kept alive the contributions of students efforts and products to the instructional sequence. A contrast to her success is George's struggle with this same section of the Vessels Unit.

Example 3—Tensions in Changing Practice

As luck would have it, on the day two researchers (L and R) showed up to videotape the beginning of George's teaching the nested-experiments section of the Vessels Unit, George was alone in his classroom during planning time. "Didn't you get my message? I called to tell you not to bother coming out. I'm not gonna do this anymore." What had come to be an insurmountable problem, in his mind, was the students' inability to work through the steps of the nested experiments—that is, to
~~~~~~~~~~~~~~~~~~

design a test of one variable of the vessel to report to other members of the class. George wanted the students to begin the process of experiment design by first having them learn how to correctly write a scientific hypothesis. It is significant to note that George wrote and designed this segment of the Vessels Unit. Here is some of what he had to say:

G: I am giving up on it. I know I'm tired of it.
L: Yeah
G: Okay, and I know the kids are tired of it. We're tired of these boats. We've been talking about these boats for a month now.
L: ???
G: So then I thought .... I was ready to give up. Really, early and that's one thing I hate to do is give up on anything. And, uh, then I went back to the lady who is an aide in here with me and she was in here and I said to her "What do you think?" and she says ... she thinks again it is too much for them to grasp, if we could have broken it down piece by piece. What I have on the board, okay. See we came up with this. Then I thought here's what we did we came up with a list of steps. Oh, I was so just darn frustrated.

The level of George's frustration is demonstrated by the fact that he invested 2½ to 3 weeks of instruction to work through the first two parts of the curriculum and then he only allocated 2 days for the last two parts (see the appendix). Though George designed and wanted to pursue the nested experiments, George perceived constraints that caused him to abandon this part of the unit. It is clear from some of the statements that George felt the students were lacking the skills and thinking processes to perform the nested experiments. The conflict resulted from, in part, his reluctance to allow the students to take ownership of the problem conceptualization—to do it themselves. We see in the two passages below his level of frustration and his sense of the problem.

G: Right, that's what I'm saying—you know they had a big problem coming up with a hypothesis on this. They had a problem, I could sit there and tell you but when you
L: Kind of had to see it to believe it.
G: When you see what's going on and you sit there and you look at these kids I'm frustrated, they're frustrated—you know it's just like it is really bad, it really was bad . And I thought there is no way I want to put my kids through this. You know um there I go sounding possessive of my kids.
L: Hey listen, they are your kids.
G: But um you know it was very, I mean it was really frustrating to go through that. I just realize, I don't know I realize what we are doing the idea, the whole idea is okay but have to realize

L: It's real hard

G: and I think even as myself I, this is a hard thing for me to accept but I, I think we really have to raise that possibility that maybe they can't. At this age, maybe they can't. I mean if Piaget is right, maybe they can't.

~~~~~~~~~~~~~~~~~~

G: ... And there you go again now, we're asking and now I think that is where the problem is. We are taking it very much from concrete to abstract and from abstract to abstract

R: Well that's. You are absolutely right that's part of the reason why we are engaging in this experiment. Because we are actually trying to find out if we want to push to this direction—we want to find out how the kids are handling it, how the classroom is handling it and also how you are handling ....

G: I think, I think too as I got to this and I saw they were stumbling I thought this would really be a good thing for them to learn how to do. You know, that's what kind of course, that's our ultimate goal but until you lay this in front of them you know. Our ultimate goal is to be able to get the kid to think well you said that that science process, that scientific thinking, you know here you have a problem now how do you test for that. You know I thought this is perfect I mean there it is right there. You know but, we haven't built to that. What we have done is we've done labs to make it [inaudible] carefully [inaudible] concept to that process and we have and that is not in any way shape or form been gearing kids for this kind of thing. You know, how would you attack the problem?

L: Well I'm betting that one of the hard parts for them is to conceptualize the problem before they begin to attack it. And I think that there are ways you could lead a conversation that could get them to conceptualize the problem and there may be better or worse things that they could come up with than actually set down the step.

In his concern with getting students to produce a statement of hypothesis and complete the task and activities, George shied away from involving his students in the all-important task of giving them the opportunity to make sense and construct meaning out of what they are learning. (It is important to note that George had made a decision to try to finish the unit prior to the winter holiday break. His decision to end the Vessels Unit occurred on December 18.) He justified his abandoning this part of the unit by appealing to an interpretation of Piaget ("But I, I think we really have to raise that possibility that maybe they can't. At this age, maybe they can't. I mean if Piaget is right, maybe they can't.")

On the day before this, George had students attempt to complete three separate SEPIA tasks. Observations confirm that in his concern for satisfying activity
~~~~~~~~~~~~~~~~~~

demands, not a single one of these tasks was used as a source of information for an assessment conversation. In addition, observations of a focus group of students completing the three tasks revealed they did not complete all three tasks and suggested the students did not understand the purpose of the tasks.

Thus, in contrast to the description of his class in the preceding section in which he took the time to have three students come to the overhead to present their ideas and in which he completed the first stage of an assessment conversation, here we find that he has not made any headway, even on the first stage. The challenge of asking students to take control of charting an experimental plan and the management of concepts is judged to be overwhelming. In the preceding part of the unit, students shared their understanding, but it was clear that the teacher was still in control of all conceptual orchestration. This inference is substantiated by the classroom observation on the next, and what would be the last, day of instruction for the Vessels Unit. Once again George displays "get through it" teaching behaviors. He has decided to show the students, through a series of demonstrations, what features of a vessel are most important in building their second vessels. He has abandoned the students' doing the work and thus the possibility of even soliciting diverse views through individual or group efforts on a task.

He takes three objects—a large (24 oz) styrofoam cup, a blue plastic rectangular tray, and a wooden box—and, one at a time, he tests each "vessel" to determine its carrying capacity. The cup, which has the highest sides, holds 38 washers; the wood box, which has the thickest sides, holds 34 washers; and the plastic tray, which has the largest overall volume, holds 59 washers. "Does the height of the sides have anything to do with it?" he asks, and in unison the students respond "No!" "Okay," George responds, "I'm going to show you what matters." Now there are two things to emphasize here. One is that the curriculum has sought all along to stress the fact that what matters is the height of the sides in combination with the bottom surface area. Together they determine the volume of the vessel, but, as stated earlier, maximizing the height and bottom area of the vessels is a problem to be solved. Here George is dismissing one of the most important concepts for designing a vessel that maximizes carrying capacity.

The next thing to emphasize is George's use of the word *show* in the statement "I'm going to show you what matters." He is in control, and it is his ideas about vessels and not those of the students that are now the currency of the classroom. The authority is not the evidence any longer but the teacher.

CONCLUSIONS

The goal of Project SEPIA is to develop teaching practices that are consistent with visions of science education suggested by current reform efforts. Central practices include:

- Acknowledgment of student conceptions through assessment strategies and the tailoring of instruction to restructure those conceptions in accord with scientific principles.
- Evaluation of knowledge claims through application of scientifically legitimate criteria.
- Emphasis on explanations, models, and experimentation as critical forms of scientific reasoning.
- Communication as a requisite skill in all science activities.
- A dual commitment to exploration and reasoning about exploration.

The assessment conversation is an instructional strategy that is designed to incorporate these practices. Our experience to date strongly suggests that full implementation of assessment conversations, and other strategies that share common goals, is a significant, but achievable, challenge for current practitioners. We can point to several issues that need to be addressed in order that the practice we envision become commonplace.

First, teachers' view of science and their concomitant view of teaching science is, as noted in the introduction, dominated by tasks and activities rather than conceptual structures and scientific reasoning. Thus, steps of the assessment conversation that focus on activity (e.g., drawing and presenting an explanation) are more readily mastered than those that focus on conceptual structures and reasoning (e.g., relating evidence and applying criteria to student explanations).

Teachers' views of science make it difficult to move to an instructional structure that is governed by scientific criteria rather than by topic coverage. The SEPIA criteria are designed to function as an underlying, consistent theme that cuts across whatever activities the class may pursue. The idea of thematic unification runs counter to traditional practice and thus represents a significant challenge to the proposed model of instruction.

Second, and not surprisingly, formal curricula do not support the current initiative. As already noted, curricula focus on the activities of exploration and not the thinking about exploration. In addition, the piecemeal nature of most formal curricula does not entertain the possibility of extended pursuit of scientific understanding. George's comment that the curriculum has done little to prepare students to design tests to determine which features of the vessel are important is but one example of the disparity between the goals of a project like SEPIA and the goals that guide most middle school science practice.

Third, reconceptualizing the relation between assessment and instruction is a major hurdle. Teachers are not used to using student information to guide and revise instructional decision making. Missing are the processes of science that address argumentation and the social dynamics of the classroom that stress the management and assessment of information and ideas. Instead, assessment is seen through a summative lens—it is something that is done at the end of an instructional sequence in order to account for student learning. Assessment is also viewed from a deficit model perspective. The traditional approach has been to question what of the target material has been learned and what has not been mastered. The current approach advocates assessment that makes public what students do understand and then uses that information to suggest appropriate instructional next steps.

Successful assessment starts with a set of very clear outcome goals for students, in this case embodied in the SEPIA criteria. At present our working criteria reflect a commitment to two important elements: (a) criteria that emphasize the development of reasoning skills and (b) criteria that stress meaning making and sense making of scientific knowledge claims. It is a working list because the criteria should change over time as the students develop the capacity to engage in higher and higher levels of cognitive processes or as the class decides to examine other contexts of science that then require other criteria (i.e., statistical significance).

Fourth, successful facilitation of assessment conversations requires a reasonable grasp of the subject matter being explored, understanding that is often lacking in middle school science teachers. Teachers need to develop a clear sense of the conceptual terrain they are exploring and also need to have a pedagogical sense of the likely understandings that students will bring to a domain. With sufficient content and pedagogical knowledge, teachers can respond to student work in a productive fashion. However, content understanding alone is not enough, for the inability of teachers to engage students in meaning making and reasoning has as much to do with confusion surrounding how to manage the flow of information, knowledge claims, and ideas produced by students as it has to do with teachers' lack of knowledge about scientific principles and concepts. This is, as Doyle (1986a) suggested, a problem of classroom organization and management.

Over the course of this project, as teachers become more familiar with particular content domains, we are noticing more facile managing of conversations about student work (Gitomer, Zohar, Chang, & Duschl, 1994). Obviously, a move towards fewer topics will also enhance the possibility that teachers can at least become competent in a few domains. This will enable an emphasis on reasoning and exploration within domains of knowledge, which we believe is a closer approximation to legitimate science activity than that which is currently practiced in most schools.

These four challenges to the assessment conversation are the intellectual demands created by working from and with students' ideas and students' representations of scientific knowledge. The intellectual challenge is one that demands teachers have knowledge of subtleties about the social, epistemic, and cognitive dynamics of the classroom and of the ways in which each of these three types of dynamics develops. For example, when an explanation is given in class, it will be analyzed, employing the SEPIA criteria, according to the social criteria that it embraces as well as epistemic claims it makes. It is legitimate then to ask students questions like "Is the explanation clear?" or "Is the explanation supported by evidence from any investigations or demonstrations done in class?" Such questions cut across tasks and embrace a more elaborate and legitimate sense of what it means to be doing science. These questions begin to provide channels of feedback that assesses the information generated by students. The feedback, in turn, ought to embrace the dual management problem of handling the sense making of science concepts and processes and the social dynamics that support students' communicating, reasoning, and reflecting on what they know and how they have come to know it.

Gradually, teachers and researchers have been developing strategies to meet these challenges. Employing a teacher-as-researcher model, we are identifying important directions to explore in order to fully realize the potential of assessment conversations and other related instructional techniques. Early in the project, teachers only felt comfortable exploring the early steps of the assessment conversation, but now they willingly pursue projects targeted at the latter, more cognitively demanding steps of the conversation. Although encouraged by the progress being made, we are also impressed by the amount of work that will be needed to significantly alter science practice on any large scale. The forms of practice that are being developed are not readily articulated in a procedural text; they require deep understanding of science, students, teaching, and assesssment—understanding that will require concerted effort by all of us in the science education community.

REFERENCES

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

Baron, J (1990). Performance assessment: Blurring the edges among assessment, curriculum, and instruction. In A. B. Champagne, B E. Lovitts, & B J. Calinger (Eds), *Assessment in the service of instruction* (pp 127–148). Washington, DC American Association for the Advancement of Science

Brown, J. S , Collins, A , & Duguid, P (1989). Situated cognition and the culture of learning *Educational Researcher, 18,* 32–41

Carlsen, W (1991) Saying what you know in the biology laboratory *Teaching Education, 3,* 17–29

Carlsen, W (1993) Teacher knowledge and discourse control Qualitative evidence from novice biology teachers' classrooms. *Journal of Research in Science Teaching, 30,* 471–482

Clark, C., & Peterson, P (1986) Teachers' thought processes. In M Wittrock (Ed), *Handbook of research on teaching* (pp 255–296) New York: Macmillan

Cobb, P (1994). Constructivism in mathematics and science education *Educational Researcher, 23,* 4

Doyle, W (1983). Academic work, *Review of Educational Research, 53,* 159–199.

Doyle, W. (1984). How order is achieved in classrooms: An interim report. *Journal of Curriculum Studies, 16,* 259–277

Doyle, W (1986a). Classroom organization and management In M. Wittrock (Ed), *Handbook of research on teaching* (pp. 392–431). New York Macmillan.

Doyle, W (1986b). Content representation in teachers' definitions of academic work. *Journal of Curriculum Studies, 18,* 365–379

Driver, R., Asoko, H., Leach, J , Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*, 4, 5–12.

Duschl, R. (1990). *Restructuring science education.* New York Teachers College Press.

Duschl, R., & Gitomer, D. (1991). Epistemological perspectives on conceptual change: Implications for educational practice. *Journal of Research in Science Teaching, 28*, 839–858.

Gee, J. (1994, April). *Science talk How do you start to do what you don't know how to do?* Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Gitomer, D. H., & Duschl, R. A. (1995). Moving towards a portfolio culture in science education. In S. Glynn & R. Duit (Eds.), *Learning science in the schools· Research reforming practice* (pp. 299–326). Washington DC: American Association for the Advancement of Science.

Gitomer, D. H., Zohar, A., Chang, M., & Duschl, R. A. (1994, April). *The impact of portfolio culture practices on classroom discourse.* Paper presented at the meeting of the American Educational Research Association, New Orleans, LA

Grossman, P. (1992). Why models matter: An alternate view on professional growth in teaching. *Review of Educational Research, 62*, 171–180.

Hodson, D. (1992). Assessment of practical work· Some considerations in philosophy of science. *Science & Education. 1*, 115–144.

Joyce, B., & Weil, M. (1986). *Models of teaching* (3rd. ed.). Englewood Cliffs, NJ: Prentice Hall.

Kagan, D. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research, 62*, 129–170.

Klahr, D., & Dunbar, K. (1988) Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48.

Kuhn, D. (1993). Science as argument. *Science Education, 77*, 319–338.

Leinhardt, G , & Greeno, J. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*, 75–95.

LeMahieu, P. G., & Foss, H K (1994, May). Standards at the base of school reform: What are the implications for policy and practice? *The School Administrator,* 16–22.

Lemke, J. (1990). *Talking science: Language, learning, and values.* Norwood, NJ. Ablex

Metz, K. (1991). Development of explanation: Incremental and fundamental change in children's physics knowledge. *Journal of Research in Science Teaching, 28*, 785–798.

National Research Council, National Academy of Sciences and Engineering. (1994). *National Standards for Science Education* Washington. DC. National Academy of Sciences Press.

Novak, J., & Gowin, R. (1984). *Learning how to learn* New York: Cambridge University Press.

Osborne, R., & Freyberg, P (1985). *Learning in Science· The implications of children's science.* London: Heinemann.

Palincsar, A., & Brown, A (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction, 1*, 117–175

Pintrich, P. R., Marx, R. W , & Boyle, R. A. (1993) Beyond cold conceptual change. The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*, 167–199.

Resnick, L. (1987). Learning in school and out. *Educational Researcher, 16*, 13–20.

Resnick, L. (1993). Standards, Assessment, and Educational Quality. *Stanford Law and Policy Review, 4*, 53–59.

Resnick, L. B., & Resnick, D P (1991) Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments· Alternative views of aptitude, achievement and instruction* (pp. 37–75). Boston Kluwer.

Roseberry, A. S., Warren, B., & Conant, F. R. (1992). Appropriating scientific discourse Findings from language minority classrooms *The Journal of the Learning Sciences, 2*, 61–94

Sanford. J (1987). Management of science classroom tasks and effects on students' learning opportunities Journal of Research in Science Teaching, 3, 249–266.
Schauble, L., Glaser, R., Duschl, R., Schulze, S., & John, J (1994). Experimentation in the science classroom *The Journal of the Learning Sciences, 4,* 131–166
Schauble, L., Klopfer, L, & Raghavan, K (1991) Students' transitions from an engineering model to a science model of experimentation *Journal of Research in Science Teaching, 28,* 859–882.
Shavelson, R., Baxter, G., & Pine, J. (1992). Performance assessment: Political rhetoric and measurement reality *Educational Researcher, 21,* 22–27
Sigel, I. (1978). Constructivism and teacher education. *The Elementary School Journal, 78,* 3, 333–338.
Sigel, I. (1993). The centrality of a distancing model for the development of representation competence In R. Cocking & K. A. Renninger (Eds.), *The development and meaning of psychological distance* (pp. 141–158). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
White, R., & Gunstone, R. (1992). *Probing Understanding.* London: Falmer.
Wolf, D. P., Bixby, J, Glenn, J., & Gardner, H. (1991). To use their minds well Investigating new forms of student assessment In G. Grant (Ed), *Review of Research in Education* (Vol. 17, pp 31–74) Washington, DC. American Educational Research Association

APPENDIX
VESSELS UNIT OUTLINE

Part 1

Step 1. Engaging Authentic Problem or Question

Letter–Reading the letter:

- Emphasize the goals: to build a model that helps in the design of a vessel; to explain why and how the design works
- Emphasize the function of the model—to maximize how much a vessel can carry
- Emphasize the performance variable—interactions with water, what matters in the letter—what doesn't matter in the letter

Capture prior knowledge about vessels:

- Diversity of vessels
- Design of uses
- Flotation questions
  Why do things float?
  Why do things stay afloat when a load is added?
  Why do things sink?

The development of lists of important concepts from the discussion of the letter should be captured and displayed publicly as word banks, concept map, cards.

Step 2. Assessment Conversation Related to Step 1

Models

Student work (PI):

- Sketch of a vessel
- Label or otherwise explain:
  - Why a vessel floats
  - Why a vessel sinks

Teacher-led SEPIA criteria discussion of student work:

- Performance criteria—that is, clarity and precision
- Subject matter content focus

Step 3. Perform the Task—First Effort

Individually students sketch–plan–do:

- Students build first vessel
- Sketch vessel (PI)—relate to goals in letter

Step 4. Assessment Conversation Related to Step 3

SEPIA criteria discussion gives rise to:

- Performance predictions (Which vessels will work best? Why?)
- Initial conversation about contrast features
- Need to capture details about vessel design—acquire bottom surface area and height of sides
  - Do all the boats weigh the same?

Teachers can pursue this question as either a warm-up activity or as a demonstration. Take one students's vessel. Ask if anyone thinks his or her vessel will weigh a significantly different amount (+ 2 g). If a student volunteers, then take that vessel and place it on a double pan balance with the first vessel. Compare and point out they weigh the same. Continue this procedure until you have convinced the students that all of the vessels regardless of shape are in the same narrow weight range.

Part 2

Step 5. Test or Solve

Students reminded to "keep an eye on things"—boat down, water up, why my boat sinks, how my boat sinks

Students reminded to "keep a record"—surface area value, weight it took to sink the vessel, design features of the vessel

Group students so that there is a distribution of vessels according to size. This will facilitate completion of the vessel testing within one class period. It will also facilitate the acquisition of evidence for the ensuing assessment conversation.

Step 6. Look for Contrasts and Patterns, Assessment Conversation related to Activities 3, 4, and 5

Review performance predictions and explanations during warm-up:

- Graph display of vessels: Student work (PI), visual representation of graph
- Locate examples of contrasts and patterns
  Same performance–different design (within same category)
  Different performance–same design (bottom area)
  Different performance–different design (extreme categories)
- Summarize contrasts and patterns
- Return to subject matter focus—why do things float and sink?

Apply SEPIA criteria to:

- Review and critique of performance, strategy, plan
- Student work (PI): Provide sketch and explanation of performance, strategy, plan
- Capture diversity of ideas and knowledge claims

- Acquire evidence that support ideas and knowledge claims
  Interaction with water
  Name the forces buoyant force—gravity force
  Pressure increases with depth

Demonstrations can be used to assist in establishing and reviewing the concepts and evidence involved in flotation and buoyancy:

- Level of water
- Pressing cups or tubs into a trough or sink or aquarium of water
- Coffee can with holes (the taller the object the better)
- Manometer (thistle tube with rubber diaphragm attached to glass u-tube)

Student work:

- Compare and relate cup pressing in water with adding weight to vessel
- Sketch, draw or other wise explain how the demonstration with the cup is related to the performance of the vessel (PI)

Part 3

Step 7. Nested Unit on Models, Experimentation, or Explanation

Class discussion of criteria for plan and a fair test
Groups of students design individual plans
Class discussion of exemplary plans; that is, those that address SEPIA and fair test criteria
Implement the plan
Report the results
Post the results
Experiments on contrasts (to include, but not limited to):

- Shape of vessel
- Bottom size of vessel
- Height of sides of vessel
- Distribution of weight in the vessel
- Measurement of change in depth of water

Step 8. Assessment Conversation Related to 7

Return to contrasts and patterns—what counts and what doesn't count

Apply SEPIA criteria to guide dialog:

- Relationships
- Alternative explanations
- Evidence for explanations

The purpose of this assessment conversation is to highlight the elements of vessel design that help to meet the goal of the project—design a model that maximizes the load a vessel can carry and provide an explanation of why it works.

Step 9. Perform the Task—Second Effort

Review goals and SEPIA criteria
Plan of action by groups of students
Sketch of vessel design with performance explanation (PI)
Construct vessel—each student makes a vessel (PI)
Performance packet (PI)

The test of the vessels can be done as a large group activity with each vessel being tested at the front of the class. The vessel that has the best results will be the one submitted to the seventh-grade competition. Stress that the effort was a group effort—whole class effort.

Part 4

Step 10. Assessment Conversation Related to Step 9

Submission of final plans and packet
  Assemble portfolio of work