Applied Measurement in Education
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653631

From Formal Embedded Assessments to Reflective Lessons: The Development of Formative Assessment Studies

Carlos C. Ayala [a]; Richard J. Shavelson [b]; Maria Araceli Ruiz-Primo [c]; Paul R. Brandon [d]; Yue Yin [ef]; Erin Marie Furtak [g]; Donald B. Young [d]; Miki K. Tomita [bh]

[a] School of Education, Sonoma State University, [b] School of Education, Stanford University, [c] University of Colorado at Denver and Health Sciences Center, [d] College of Education, University of Hawaii, [e] College of Education, University of Illinois, Chicago [f] College of Education, University of Hawaii, Manoa [g] Max Planck Institute for Human Development, [h] The Curriculum Research and Development Group, University of Hawaii,

Online Publication Date: 01 October 2008

APPLIED MEASUREMENT IN EDUCATION, 21: 315–334, 2008

ISSN: 0895-7347 print / 1532-4818 online
DOI: 10.1080/08957340802347787

Routledge
Taylor & Francis Group

From Formal Embedded Assessments to Reflective Lessons: The Development of Formative Assessment Studies

Carlos C. Ayala
School of Education
Sonoma State University

Richard J. Shavelson
School of Education
Stanford University

Maria Araceli Ruiz-Primo
University of Colorado at Denver and Health Sciences Center

Paul R. Brandon
College of Education
University of Hawaii

Yue Yin
College of Education
University of Illinois at Chicago
College of Education
University of Hawaii at Manoa

Erin Marie Furtak
Max Planck Institute for Human Development

Donald B. Young
College of Education
University of Hawaii

Correspondence should be addressed to Dr. Carlos C. Ayala, School of Education, Sonoma State University, 1801 E. Cotati Ave., Rohnert Park, CA 94928. E-mail: carlos.ayala@sonoma.edu

Miki K. Tomita
School of Education
Stanford University
The Curriculum Research and Development Group
University of Hawaii

The idea that formative assessments embedded in a curriculum could help guide teachers toward better instructional practices that lead to greater student learning has taken center stage in science assessment research. In order to embed formative assessments in a curriculum, curriculum developers and assessment specialists should collaborate to create these assessment tasks. This article describes the development of the formal embedded formative assessments and implementation plans for the collaborative research study. It describes the fundamental shift away from "summative assessment scripts" to formative assessments lesson scripts. Assessment tasks and implementation plans are described along with the rationale for why these tasks were selected and where these tasks were placed in the curriculum. Finally, we conclude about how to embed formative assessments in new or existing curricula and how to help teachers use these assessments successfully. We point out the critical importance of collaboration and professional development aimed at enabling teachers to re-conceptualize the role of assessments in their teaching, linking formative assessments to overall goals, and providing a learning trajectory as reference for teachers to locate students' ideas in the trajectory and provide feedback accordingly.

INTRODUCTION

Although some empirical evidence suggests that formative assessment leads to increased learning (Bell & Cowie, 2001; Black & Wiliam, 1998; 2004; Shephard, 2000), how these formative assessments are designed, developed, embedded, and eventually implemented by teachers is poorly understood.

In this article we report findings from a study that helps to close this gap. We describe how we went about building, refining, and embedding *formal formative assessments* into an inquiry science curriculum—*Foundational Approaches in Science Teaching* (FAST). They are termed formal because we crafted assessment tasks that would be available for teachers to use at critical times in a curriculum sequence; this contrasts with on-the-fly and planned-for formative assessments that capitalize on informal ongoing clinical observations or create teachable moments for enhancing students' understanding (see Shavelson et al., this issue). They are embedded assessments because they are inserted into a curriculum to be used at a particular time as opposed to the end of a unit. They are formative assessments because they are developed to give a snap shot to students and teachers about what students know and are able to do at a particular time such that this information can be used to close the gap in students' understanding by both teachers and students. The purpose of the article is to share the knowledge we developed during the formative-assessment construction process rather than providing the details of it.

The project went through three phases: (1) planning, designing, and developing the embedded assessments, (2) piloting the embedded assessments, and (3) refining the embedded assessments. In what follows we describe these phases. We focus the description on those aspects that we believe are transferable to any project with a similar endeavor.

PLANNING, DESIGNING, AND DEVELOPING EMBEDDED ASSESSMENTS

Five critical activities comprised this phase of embedded assessment development: (1) mapping and experiencing the curricular unit in which the formative assessments were to be embedded, (2) determining the unit goal to be assessed, (3) determining the critical points where the assessments should be embedded, (4) defining the assessment development guidelines, and (5) developing the assessments. These activities were carried out by an interdisciplinary Assessment Development Team (ADT). The ADT consisted of Stanford Education Assessment Laboratory (SEAL) assessment specialists and researchers, Curriculum Research & Development Group (CRDG) curriculum developers, FAST trainers, FAST teachers, and a scientist (see Shavelson et al., this issue). Putting together an ADT in which curriculum developers, assessment and curriculum specialists, scientists and teachers collaborate in planning, designing, and developing the embedded assessments is a critical component for the development of effective embedded assessments. The expertise that each group brings to the table is important for considering the different aspects of the assessments, from content to language to technical issues.

Mapping and Experiencing the Unit

The embedded assessments were developed for the FAST middle-school Physical Science curriculum (Pottenger & Young, 1992). The content for the 12 investigations selected as the "unit" focuses on the concept of buoyancy. FAST develops students' science understandings incrementally in a manner that parallels how science knowledge was developed in the Western world (cf. King & Brownell, 1966), and as such, students' understandings of why things sink and float are developed by building explanations of sinking and floating phenomena sequentially beginning with the concepts of mass and volume and moving to the concepts of density and relative density.

The ADT's first activities focused on experiencing the FAST investigations to provide members with a concrete idea of what the learning activities entailed, and on determining the framework that would be used to analyze the FAST investigations and to develop the specifications for the development of the embedded assessments. The curriculum developers provided abbreviated hands-on demonstrations of the investigations while other ADT members participated as students. The team discussed what FAST teachers typically used as assessments and students' corresponding answers.

The development of the embedded and end-of-unit assessments was guided by a *conceptual framework for science achievement* (see Shavelson et al., this issue, Figure 1). The framework presents science achievement as reasoning with (at least) four overlapping types of knowledge: declarative, procedural, schematic, and strategic.

SEAL research has linked certain types of assessments to this science achievement framework. Briefly put, to measure the structure of declarative knowledge, multiple-choice, short-answer and concept maps provide valid evidence (e.g., Li & Shavelson, 2001; Ruiz-Primo & Shavelson, 1996a; Shavelson & Ruiz-Primo, 1999). To measure procedural knowledge, performance assessments are appropriate (e.g., Li & Shavelson, 2001; Ruiz-Primo & Shavelson, 1996b; Shavelson, Baxter, & Pine, 1991). To measure schematic knowledge, multiple-choice, short-answer items, and performance assessments are appropriate (Li & Shavelson, 2001; Li, Ruiz-Primo, & Shavelson, 2006). Strategic knowledge is difficult to measure directly but is essential, especially with novel assessment tasks.

SEAL staff developed *storyboards* that showed the types of knowledge that were addressed in each of the investigations. Mapping the unit with the framework revealed that the curriculum did not address schematic knowledge (see Brandon et al., this issue). As a consequence, the final version of the embedded assessments (called "reflective lessons") focused primarily on schematic knowledge—explaining why things sink and float. But we get ahead of the story.

Determining the Unit Learning Goal

Once the unit was mapped and experienced the ADT defined the overarching learning goal that would be assessed at the end of the unit and that would guide the focus of the embedded assessments along the way. This task was critical. Rather than defining the goal as "students would be able to understand buoyancy," the ADT decided to focus on the conception of "*why things sink and float*." The team considered the development of this schematic knowledge to be fundamental to teaching buoyancy; "Why things sink and float" was, ultimately, the center around which the embedded assessments were designed and developed but, following our achievement framework, declarative and procedural knowledge were also prominently included in the initial set of assessments.

Determining the Critical Junctures of the Unit

A critical question in designing formal formative assessments is "What and where to embed these assessments?" Using the story boards, the ADT identified the most important concepts taught in the investigations to be used in a posttest assessment suite, and identified the points (natural joints) in the instructional sequence in which the formative assessments were to be embedded.

More specifically, the team came up with three criteria to identify the natural joints: (1) a subgoal of the end-of-unit goal is achieved, that is, there is a body of knowledge and skills sufficiently comprehensive to be assessed; (2) teachers need to know about student understanding before they proceed with further instruction; and (3) feedback to students is critical to help them improve their understanding and skills of the material already taught (Shavelson, SEAL & CRDG, 2005, p. 6). Four embedded-assessment natural joints were identified in the 12 investigation sequence (Figure 1).

The ADT then developed comprehensive assessment blueprints for the FAST investigations using the science achievement framework. The blueprints identified key junctures in the set of investigations as to where embedding should take place, and it linked types of embedded assessments to the knowledge types to be tapped. We now turn to the development of assessments.

Defining the Assessment Development Guidelines

Deciding what to assess in the embedded and in the end-of-the-unit assessments was not as straight forward as finding the natural joints. The ADT started by defining guidelines for developing assessments. Embedded assessment tasks, at

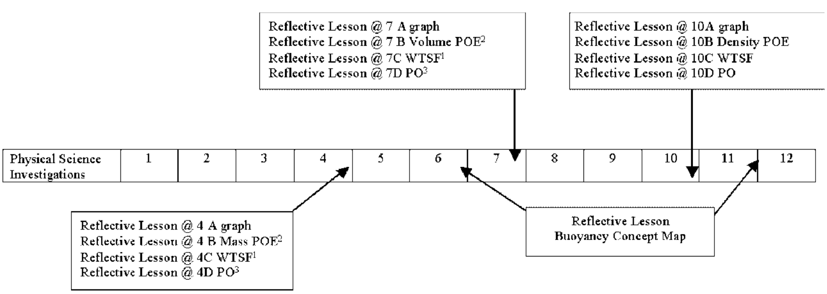


FIGURE 1 Assessment suite timeline and natural joints.

each juncture, should tap each of the three types of knowledge: declarative, procedural, and schematic using the type of assessment that best elicit each knowledge type.

With these guidelines in mind, three tasks were the focus of this activity, defining: embedded assessment content, assessment task types, and length of each embedded assessment. In the sequence of investigations that typically takes eight to ten weeks to implement in classrooms, students learn many important concepts and procedures, a typical situation in any instructional unit. Because these topics are clearly identifiable to the curriculum developers, all of these topics become equally important to them as assessment targets. It is important to note that the tension between time spent on instruction versus assessment and depth versus coverage of the assessments was a concern from the beginning and it remained so throughout the project. Achieving a balance among these dimensions ended up being more difficult than the ADT originally expected.

Developing the Assessments

Once the assessment tasks were decided on, the ADT began an iterative process of designing and refining assessments, piloting the assessments and content validating them with the rest of the team. The first *assessment suite* of embedded assessment tasks was comprised primarily of multiple-choice, concept-map and POE tasks, a suite to be administered at each of four critical joints (Figure 1).

Lessons Learned

In retrospect, another task that should have been carried out in this phase is the definition of a *learning trajectory* (see later) for the unit (moving from mass/volume to density to ratio of densities explanations). We developed this trajectory following the piloting phase, while revising the assessment suites. Having the trajectory from the beginning would have helped to determine more clearly the evidence that needed be collected about the students' level of understanding at each juncture, and to better focus the assessment tasks.

PILOTING THE EMBEDDED ASSESSMENTS

To study the quality of the embedded assessments and their implementation in teaching about buoyancy, the ADT carried out a pilot study. Three teachers were trained in the use of the embedded assessments for their classes. Teachers reviewed the curriculum, carried out the embedded assessments first with each other as students, and then as teachers working with students involved in FAST curriculum in the CRDG's summer school.

The pilot study lead to three critical findings about the teachers' implementation of the assessments: (1) teachers treated the embedded assessments just like any other test they might give; (2) feedback to students was not immediate; and (3) teachers needed increased structure on how to implement the embedded formative assessment and how to take advantage of the "teachable moments" provided by these tasks. The pilot revealed that teachers would review the material covered in the unit before the embedded assessments even though the purpose of the formative assessments was to do just that. Overall, the teachers believed these embedded assessments to be summative assessments and would revert from formative assessment pedagogy to a summative assessment "teaching script" (Shavelson, 1986).[1] The study also revealed that pilot study teachers often would provide feedback to students weeks after the assessment was administered, thus missing the teachable moments provided by the embedded assessments. It became clear that how these assessments were to be used in the classroom was very important and that teachers preconceived notions about assessments influenced their implementation. Furthermore, although teachers were able to elicit student conceptions about why things sink and float, they did not necessarily use these conceptions to further student learning. This shift from assessment activities to learning activities represented a fundamental change for teachers in the way to look at the formative assessment practices.

The study findings suggested that the embedded assessments should be: (1) reduced in number due to time, (2) short in duration and tightly focused on the key outcomes of the unit—explaining "why things sink and float" based on relative density, (3) administered in no more than two lesson periods at the critical junctures, (4) allow for immediate feedback to teacher and students, (5) provide opportunities for students to test their "why-things-sink-and-float" explanations with evidence from the assessment event and to take a stand on what they believe, and (6) set the stage for the next set of investigations.

In order to avoid the usual summative assessment teaching scripts, we changed the name of our formative assessments from "embedded assessments" to "*reflective lessons.*" The reflective lessons evolved from assessment activities to learning activities intended to provide instructional information to both the student and the teacher by: (1) building on what students already know, (2) attending to student conceptions and misconceptions, (3) making student conceptions and misconceptions public and observable, (4) priming students for future learning, and (5) reflecting on material covered. Finally, we decided to provide teachers with concrete strategies for implementing the assessment suites.

[1]A summative assessment teaching script can be conceptualized as a formalized teaching pattern consisting of a set of expectations of what events are necessary and about the temporal order of such events. A summative assessment teaching script might include studying for the tests, taking practice tests, or reviewing lecture notes with the students prior to actually giving the test.

Lessons Learned

In retrospect, the ADT did not balance three critical issues during the assessment development process: assessment-task characteristics, feedback using information collected, and training activities to help teachers better understand the critical purposes and uses embedded assessments. In the final analysis, the ADT's main focus was on the characteristics of the assessment tasks and the issues related to their administration rather than on thinking concurrently about the strategies that teachers would need in order to identify the teachable and feedback moments. Furthermore, the pilot-teacher training did not focus sufficiently on how teachers could use the information collected from the assessments to provide feedback to the students and to adjust their instruction to move students forward in their learning.

REFINING THE EMBEDDED ASSESSMENT: FROM FORMATIVE ASSESSMENTS TO REFLECTIVE LESSONS

In an attempt to avoid the usual summative assessment teaching script, we changed the name of our formative assessments from "embedded assessments" to "*reflective lessons.*" The reflective lessons also evolved from assessment activities to learning activities intended to provide instructional information to both the student and the teacher (see earlier).

The reflective lessons were composed of a carefully designed sequence of investigations (prompts) that enabled teachers to step back at key points during instruction (at natural joints) to check student understanding, and to reflect on the next steps to move forward in developing scientific explanations (Figure 1). Furthermore, these assessments were carefully designed to match the content and structure of the existing FAST investigations. The SEAL team developed drafts of the assessments and, using talk-aloud methods, refined them in tryout-refine-tryout cycles conducted with students from local schools.

Reflective Lessons: Critical Characteristics of Formal Embedded Assessments

In order to promote student understanding of why things sink and float via relative density concepts, the reflective lessons were designed to elicit and make public student sinking-and-floating conceptions, encourage communication and argumentation based on evidence from the investigations or assessment tasks, challenge students' why-things-sink-and-float conceptions, help students track their sinking and floating conceptions and help students reflect on these concepts (e.g., Duschl, 2003).

Eliciting and making students' conceptions public

The reflective lessons were intended to make students' thinking about different physical phenomena explicit (e.g., asking students, "Why things sink and float?"). In order to achieve this purpose, reflective lessons included a set of activities/investigations that brought forth students' conceptions of density and why things sink and float (e.g., having students predict whether a small-sized high density plastic block will sink or float and explain why). Moreover, they included different suggested teaching strategies, such as student group work, small and/or large group discussions, sharing activities, and questions and prompts that help make students' conceptions and thinking public.

Encouraging communication and argumentation of ideas using evidence

We designed the reflective lessons to provide opportunities for students to discuss and debate what they know and understand. By having students take a stand and make their why-things-sink-and-float conceptions public, students were provided with the opportunity to see competing student conceptions, and hear and evaluate the supporting evidence provided for the different competing views. Reflective lessons were concerned with what data counts as evidence for students (e.g., Uncle Joe's tales of the sea vs. results of an investigation), how this evidence is used to support their predictions, decisions and explanations, and how these explanations could be generalized to other similar situations—the universality of the scientific principles. The use of evidence for explanations in the reflective lessons was an extension of the pedagogical practices already found in FAST.

Furthermore, the reflective lessons provided teachers with examples and models for setting up the lessons, collecting student conceptions, establishing discussion events, and providing questions to challenge students' evidence-based why things sink and float explanations. The goal was to have student use scientifically sound evidence and test the universality of their claims. That is, the reflective lessons provided teachers with: "How do you know that?" "What evidence do you have to support your explanations?" "In what cases does your explanation apply and what cases does it not apply?"

Challenging students' conceptions

The reflective lessons were also designed to challenge students' ideas about why things sink and float. In the reflective lessons, students are presented with situations/events that seem to be buoyancy anomalies, and provide instances where everyday knowledge about sinking and floating events cannot be easily applied. For example, the reflective lessons asked students to predict whether a large-sized low-density plastic block (polypropylene) or a small-sized light-weight high-density plastic block (PVC) would sink or float. Students focused on the size and weight of the two blocks (decisions based on everyday life) rather than on the density of the plastic (decisions based on scientific evidence).

Helping students track conceptions

The reflective lessons were also designed in part to track student why-things-sink-and-float conceptions. That is, in the reflective-lesson suites there were assessment items that were used each time and that allowed teacher and students to track student understanding. With the goal of tracking students' conceptual development, each reflective lesson included a new prompt asking students to explain *Why Things Sink and Float* and the same concept terms were used in both reflective-lesson concept map administrations. By comparing earlier assessment results with later assessments results, individual and group student progress could be observed.

Reflecting with students about their conceptions

The reflective lessons were intended to help teachers pinpoint students' conceptual development at key instructional points, and help guide the development of students' understanding by identifying where students were going wrong. Through making student conceptions public and then through discussions and debate based on evidence, teachers and students would find the problems they are having as they move toward the unit goals. Furthermore, the reflective lessons provided teachers with strategies for progress that might help students achieve these goals.

Connecting assessment tasks to the curriculum activities

The reflective-lesson goals were derived directly from the FAST curriculum. The curriculum developers stated the importance of these types of learning activities (i.e., encourage argumentation about evidence, challenge student buoyancy conceptions, reflect on student's buoyancy conceptions) in their instructional materials about group work, class discussion, looking for universality of student claims. What the reflective lessons did is to make the whole reflective discussion explicit in terms of what, when, and how to do it with students.

TYPES OF REFLECTIVE LESSONS

Reflective lessons provided specific prompts expected to prove useful for eliciting students' conceptions, encouraging communication and argumentation based on evidence, and helping teachers and students reflect about their learning and instruction. These prompts varied according to where the reflective lessons were embedded within the unit. Two types were identified.

Type I Reflective Lesson

Type I reflective lessons were designed to expose students' conceptions of why things sink and float, encourage students to cite evidence to support their conceptions, and raise questions about the universality of their conceptions when applied to new situations. These lessons were embedded at three joints in the unit (Figure 1). Each Type I reflective lesson has four prompts: (1) interpret and evaluate a graph, (2) predict-observe-explain an event related to sinking and floating, (3) answer a short question, and (4) predict-observe an event related to a new idea in sinking and floating.

Graph prompt

This prompt asked students to use their knowledge of investigations, resulting data, and explanations collected in different FAST investigations as evidence to support their responses and conclusions (Figure 2). Students were familiar with the data representation used because the graphs were like those created by the students in their FAST investigations and revealed the important relationship between two variables that were just studied. The ADT chose the graphing prompts because of the importance of interpreting graphs in the curriculum in helping students use data to draw conclusions and support explanations. Graphs like these are used throughout the 12 physical science investigations in the unit.

Predict-Observe-Explain prompt

This prompt assesses schematic knowledge. It asks students to predict the outcome of a sinking-and-floating event and to justify their prediction, observe the event, and then reconcile their predictions and observations (White & Gunstone, 1992) (Figure 3). Students use what they had learned in the unit to help them make predictions and to explain and justify their ideas. In order to make students' thinking explicit, they write their predictions, observations, explanations, and reconciliations. Students were expected to use evidence in their explanations and reconciliations. Careful administration is important for this prompt type. For example an important administration consideration is when and how to collect student predictions and evidence-based explanations without affecting student reconciliations—to make ideas public and anonymous. The activities used in this assessment came from the literature on students' understanding about density, developmental psychology of dimensionality (relating two variables together) in student learning, and from the repertoire of science teachers associated with the

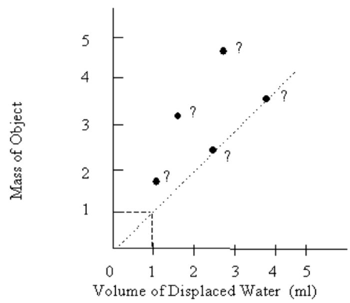
Reflective Lesson @ 7A: Volume and Mass

Name__________________ Teacher_______ Period__ Date____

Instructions: Please read along with your teacher.

Don has six objects. He measures the mass of each object. Then he measures the volume of water displaced by the objects with an overflow can. He plots the points on the graph to show the mass and displaced water volume of the objects.

He forgot to write down whether the objects were sinkers or floaters.



Tell Don whether the objects are sinkers and floaters in water. Write your answer on the graph. How did you know?

FIGURE 2 Reflective lesson graph after investigation 7.

project. These activities focus on the variables that the students had just completed in their FAST investigations.

Short-Answer prompt

This prompt is a single question that asks students to explain why things sink and float with supporting examples and evidence (Figure 4). The same prompt is used, as is, in all three Type I reflective lessons. This prompt is a direct result of the ADT work in clarifying the goal of the instructional unit. Furthermore, this prompt eventually served as a vehicle to track student understanding across FAST investigations, thus providing a link between the embedded formal formative assessments and the summative assessments (i.e., the pre- and posttests).

Predict-Observe prompt

A slight variation on the Predict-Observe-Explain prompt, this prompt asks students to predict and observe an event (Figure 5). These prompts, based on the

Reflective Lesson @ 7:Volume POE

Name______________ Teacher____________ Date____

Instructions: Read along with your teacher.
You have three bottles in front of you. Hold them in your hand. Feel their mass. Observe the bottle's volume. Record in Table 1 the mass and volume of each bottle.

Table 1

Bottle #	mass (g)	volume (ml)
Bottle 1		
Bottle 2		
Bottle 3		

Based on what you know about mass and volume, PREDICT whether each bottle will float, subsurface float or sink when your teacher places the bottles into water (Table 2). Why?

Table 2

	Prediction (Circle)	Why do you think that the bottle will float, subsurface float or sink?
Bottle 1	Float, Subsurface float, Sink	
Bottle 2	Float, Subsurface float, Sink	
Bottle 3	Float, Subsurface float, Sink	

Observe your teacher place the bottles into a container of water. Record whether the bottle floats, subsurface floats or sinks. Explain why what you observed happened.

Table 3

	Observation (Circle)	Matched Prediction?	Why did what you observed happen? Explain why you were right or wrong?
Bottle 1	Float, Subsurface Float, Sink	Yes Or No	
Bottle 2	Float, Subsurface Float, Sink	Yes Or No	
Bottle 3	Float, Subsurface Float, Sink	Yes Or No	

FIGURE 3 Type I Reflective lesson predict-observe-explain prompt after investigation 7.

Reflective Lesson @ 7A
Why things sink and float?

Name_______________ Teacher_____________ Date_______

Please answer the following question. Write as much information as you need to explain your answer. Use evidence, examples and what you have learned to support your explanations.

Why do things sink and float?

FIGURE 4 Type I Reflective lesson short-answer after Investigation 7.

FAST challenge questions, were restructured using a modified POE model and student worksheets were developed for these POs. Students are not asked to reconcile predictions and observations. POs act as a launching point for the next instructional activity of the unit in which the explanation will emerge.

Type I reflective lesson implementation

The team allowed for three class sessions for Type I reflective lesson implementation, although the team expected implementation might take only two or two-and-a-half sessions (see Furtak et al., this issue). Each session was planned to take about 45 minutes or so to complete. The implementation sequence is shown in Figure 6.

Reflective Lessons @ 7D: Predict and Observe

Name_______________ Teacher_______ Period___ Date_____

Instructions: We have one bottle. Observe the bottle's total volume. Record in Table 1 the mass and total volume of the bottle.

Bottle #	Mass (g)	Total Volume (cm^3)
Bottle 4		

Based on what you know about mass and volume, PREDICT whether the bottle will float, subsurface float or sink when I place the bottle into water (Table 2). Why?

	Prediction (Circle)	Why do you think that the bottle will float, subsurface float or sink?
Bottle 4	Sink Float or Subsurface Float	

Observe I place the bottle into a container of water. Record whether the bottle floats, subsurface floats or sinks. Explain why what you observed happened.

	Observation (Circle)	Matched Prediction?	What happened?
Bottle 4	Sink Float or Subsurface Float	Yes Or No	

FIGURE 5 Type I Reflective lesson predict-observe after investigation 7.

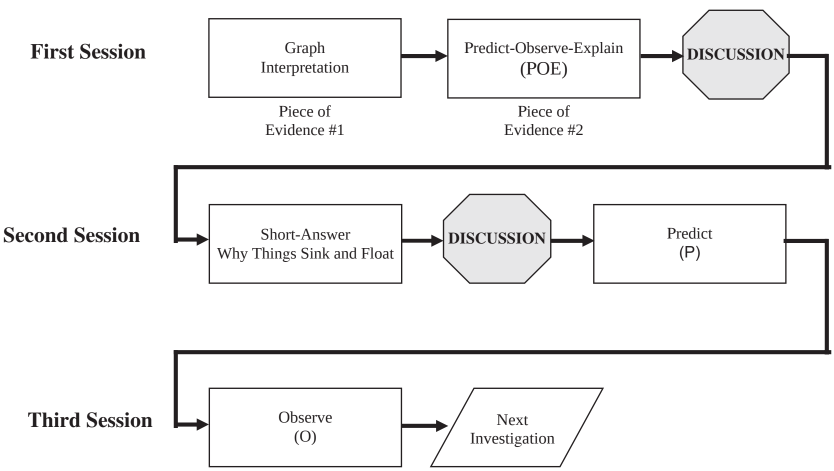


FIGURE 6 Type I Reflective Lesson Suite implementation by session (adapted from SEAL's Teacher guide to the reflective lessons, 2003).

The important components of the reflective lessons are the discussions that occur after the students have completed a few of the reflective-lesson prompts. Although the actual reflective lessons are important, they represent only part of what needs to happen to promote the goals of the reflective lessons. The discussions are key because this is where the teacher makes student conceptions and reasoning public and where the relevance and importance of supporting one's explanations with evidence and testing for universality are brought to the surface. The *Teacher's Guide to the Reflective Lessons*, developed for the project, suggests that the discussion during the first two prompts focus on the information from the graph and the Predict-Observe-Explain prompt as well as what has happened in class. In the subsequent short answer question, *Why do things sink and float?,* the discussion focuses on students' conceptions and the universality of the ideas and evidence. Students are asked to extend what they know beyond the context of the Graph or Predict-Observe-Explain prompt, although both might be used as evidence. If students rely primarily on the graph or Predict-Observe-Explain prompt as evidence, the teacher is expected to push them beyond these toward universal explanations and encourage students to use information learned in class.

The last reflective-lesson prompt, the Predict-Observe prompt, is performed in two parts. First students make a prediction based on recent class investigations, reflective lessons and discussions, and then they observe a demonstration that may challenge their ideas. The two parts may be carried out on the same day, or divided as shown in Figure 6 so that part one is carried out at the end of class session two, while part two is carried out at the beginning of class session 3.

Type II Reflective Lesson

This type of reflective lesson employs only concept maps and focuses on students' progress in conceptual understanding of the relations among key concepts (terms) across the 12 investigations. These reflective lessons are implemented in one session (Figure 7).

The first box in the implementation model indicates that students need to be trained to draw concept maps. This training takes about 30 minutes, but it needs to be done carefully and teachers need to make sure that students know how to construct concept maps in accordance with a set of rules. The second time students construct the maps they need only to be reminded of the rules and do not need the entire training.

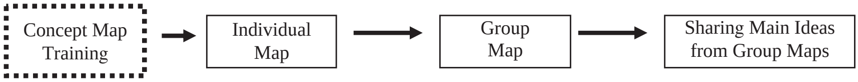


FIGURE 7 Implementation of the Type II reflective lesson in one session.

Based on pilot study observations, the most important aspect of student interaction occurs when the students construct a group concept map. Although a class concept map may be desirable, we found that only a handful of students participated in constructing a whole class concept map and consequently, the whole class concept map is not as desirable. The sharing of main ideas benefits from looking for those terms that students have not been able to place into the concept maps, looking at the terms most often used in the group maps, and having students provide a rationale for how they constructed their concept maps. For example, students are not expected to know how to relate one of the concept map terms (*density)* with the other concept terms in their why-things-sink-and-float concept map until late in the unit.

Defining a Learning Trajectory: The FAST Buoyancy Learning Trajectory

Although the curriculum was reviewed for natural joints and assessments were refined, it became clear that FAST investigations reflected somehow a learning trajectory for student understanding of why things sink and float. SEAL team formalized the trajectory to help teachers identify the expected student level of understanding at different points of the unit. The learning trajectory captured the FAST investigation sequence and exemplified students' understanding of why things sink and float was built through a series of explanations of diverse quality (Figure 8).

In the first three investigations students work with alternative conceptions of sinking and floating events (size, number of bbs, greater or lesser amounts of something) and then in investigation 4 students focus in on mass as an important variable for sinking-and-floating events. In investigations 5 and 6 students learn

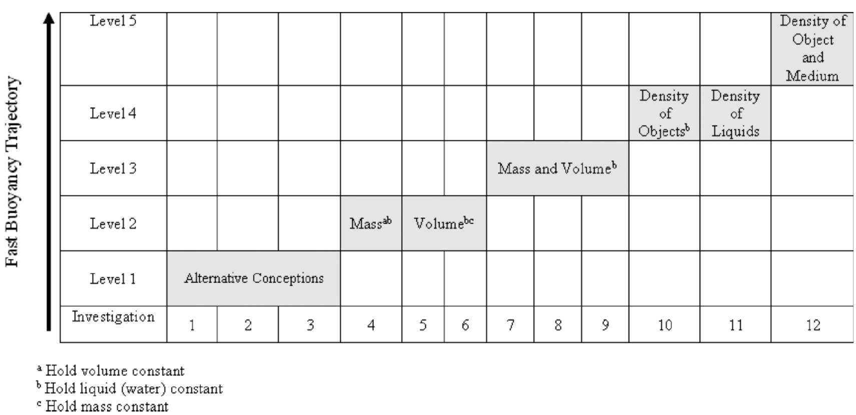


FIGURE 8 FAST buoyancy learning trajectory.

that volume, especially displaced volume, plays an important role in explaining sinking and floating events. In investigations 4–6, students hold mass and volume variables independent of each other. In investigations 7–9 students combine both volume and mass together to explain sinking-and-floating events. In investigation 10 (and part in 9), students learn about the density of an object, whereas in investigation 11 students learn that liquids have density too. Finally, in investigation 12, students learn that sinking and floating events can be explained by comparing the density of an object and the density of medium in which it is placed—relative density.

The learning trajectory is useful when a teacher is trying to understand where students are in their understanding of why things sink and float. Many student responses can readily be placed in the model and a teacher armed with this knowledge might fashion a discussion question that gets at what is important for a student to know to move forward. This learning trajectory also clarifies and simplifies the teaching task by providing a concise developmental set of categories of student conceptions.

Lessons Learned

The development of a learning trajectory was critical in focusing the assessment tasks, in focusing teachers' discussions, and in helping teachers identify where students were in that trajectory. In hindsight, the ADT admitted that this activity should come at the first phase of the project. We believe that, in future, this activity should be an outcome of the process of mapping a unit for its content. Having a learning trajectory helps not only to focus the assessment tasks, but also to have clarity about the evidence that is required from students at each juncture to determine their level of understanding. Furthermore, in retrospect, the learning trajectory could have been used as a tool for students to track their progress.

CONCLUSIONS

In order to create efficacious embedded formal formative assessments or *reflective lessons* in new or existing curriculum, several considerations may be drawn from this project. First, collaboration of the assessment specialists with curriculum developers is vital. In this assessment development, the line between assessment specialists and curriculum developers was very thin and often blurred because of the close link between curriculum and assessment. Moreover, these embedded assessments appeared seamless to students and teachers in terms of instructional approaches and in content. This helps ameliorate the summative assessment teaching script and student testing scripts as well as helps teachers and students make use of the information that they gather from these reflective lessons. If the assessments do not look like the other lessons it is hard to make the information connect to what the teachers and students are already and will be doing.

Second, professional development should be provided to teachers to reconceptualize the value of assessment, especially formative assessment. Assessment specialists or curriculum developers cannot expect teachers to use formative assessment effectively without training. The strength of teachers' summative assessment scripts may defeat the purpose of the formative embedded assessments. Student argumentation and teacher elicitation of student evidence should be emphasized in the professional development and curriculum developers and assessment developers should have a comprehensive theory of teacher feedback to students to guide this professional development.

Third, the embedded assessments must be linked to the overall goal of the curriculum and not only to the material that the students have just covered. If the embedded assessments are only linked to the material just covered, instructional time may be misused, focusing on concepts or procedures that might not be vital to the overall goal of the instructional unit (i.e., focusing on students' abilities to mass an object versus why-things-sink-and-float). Because the ADT guided the development of the embedded assessments, projects of this nature should provide all ADT members with thorough background about the curriculum and include a pilot test before field-testing teachers' and students' use of embedded assessments. Furthermore, in order to reduce the number of newly developed embedded assessments, embedded assessments should address critical content or, as in our case, content the curriculum left implicit—explanation.

Fourth, the learning trajectory proved to be a useful tool both to guide assessment development and to enable teachers to track student understanding throughout unit implementation and provide feedback.

Fifth, we lack knowledge of how teachers use embedded assessments information (but see Furtak et al., this issue). When we consider formative assessments, assessment specialists and curriculum developers must consider all five important assessment pedagogies: knowing the content, developing the assessments, implementing the assessments, collecting information from the assessments, and most importantly, using the information gained (see Ayala, 2005; Ayala & Brandon, 2008).

Finally, this project raised questions about the quantity and frequency of formative assessments in a curriculum. Having some embedded formal formative assessments or reflective lessons in a curriculum is useful because it reminds teachers to reflect back on what has been learned and hopefully guide future lessons toward unit goals. The more efficacious use of these assessments may reside in the application of formative assessment principles to all instructional activities (including reflective lessons). These formative assessments should be coherently focused on the overall goal of the unit rather than each minuscule instructional or assessment target. Furthermore, the number of formative assessment items need not be as great as that used in summative assessments to meet the reliability standards because the results of the formative assessments may be revisited through regular teacher clinical observations and review of other student work.

ACKNOWLEDGMENTS

This research was supported, in part, by a grant from the National Science Foundation (NSF/Award ESI-0095520) and, in part, by the National Center for Research on Evaluation, Standards, and Testing (CRESST/Award 0070 G CC908-A-10).

REFERENCES

Ayala, C., (2005, April). *Assessment pedagogies*. Paper presented at the annual meeting of the American Educational Research Association, Montreal Canada.

Ayala, C., & Brandon, P. R. (2008). Building evaluation recommendations for improvement: Insights from student formative assessments. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 159–166). New York: Guilford.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, *85*, 536–553.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–73.

Black, P., & Wiliam, D. (2004). Classroom assessment is not (necessarily) formative assessment (and vice-versa). In M. Wilson (Ed.), *Towards a coherence between classroom assessment and accountability* (pp. 183–188). Chicago: University of Chicago Press.

Duschl, R. A. (2003). Assessment of inquiry. In J. M. Atkin & J. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41–59). Arlington, VA: NSTA Press.

King, A. R., & Brownell, J. A. (1966). *The curriculum and the disciplines of knowledge; a theory of curriculum practice*. New York: Wiley.

Li, M., & Shavelson, R. J. (2001). *Examining the links between science achievement and assessment.* Paper presented at the American Educational Research Association Annual Meeting, Seattle, WA.

Li, M., Ruiz-Primo, M. A., & Shavelson, R. J. (2006). Towards a science achievement framework: The case of TIMSS 1999. In S. Howie & T. Plomp (Eds.), *Contexts of learning mathematics and science: Lessons learned from TIMSS* (pp. 291–312). London: Routledge.

Pottenger, F. & Young, D. (1992). *The local environment: FAST 1 foundational approaches in science teaching*. University of Hawaii Manoa: Curriculum Research and Development Group.

Ruiz-Primo, M. A., & Furtak, E. M. (2004). *Informal assessment of students' understanding of scientific inquiry.* Paper presented at the American Educational Research Association Annual Conference, San Diego, CA.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996a). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33*(6), 569–600.

Ruiz-Primo, A., & Shavelson, R. (1996b). Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching, 33*(10), 1045–1063.

Shavelson, R., Stanford Educational Assessment Laboratory (SEAL) and Curriculum Research & Development Group (CRDG). (2005). *Embedding assessments in the FAST curriculum: The romance between curriculum and assessment*. Final Report.

Shavelson, R. J., & Ruiz-Primo, M. A. (1999). On the assessment of science achievement. (English version). *Unterrichts wissenschaft*, *27*(2), 102–127.

Shavelson, R. J., Baxter, G., & Pine, J. (1991). Performance assessment in science. *Applied measurement in education*, *4*(4), 347–362.

Shavelson, R. J. (1986). Toma de decision interactiva: Algunas reflexiones sobre los procesos cognoscitivos de los profesores. (Translation: Interactive decision making: Some thoughts on teacher cognition.) In L. M. V. Angulo (Ed.), *Pensamientos de los profesores y toma de decisiones*. Seville: Universidad de Sevilla, Servicio de Publicaciones.

Shephard, L. A. (2000). *The role of classroom assessment in teaching and learning*. Santa Barbara: Center for Research on Education, Diversity & Excellence, University of California, Santa Barbara.

White, R., & Gunstone, R. (1992). *Probing understanding*. London: Falmer Press.