Development and Validation of an Inquiry Science Student Achievement and Attitudinal Suite

Carlos C. Ayala Ph.D.
Sonoma State University

A paper presented at the
Annual Meeting of the American Evaluation Association
October 29th, 2005, Toronto Canada

Introduction

The purpose of this paper is to examine the validity of a student assessment suite. These student outcome measures have been designed to examine student learning in the Foundational Approaches to Science Teaching (FAST; Pottenger & Young, 1992). classroom. Moreover, the assessments designed here are intended to be sensitive to possible differences in program implementation that may be linked to variations in teacher implementation of the FAST Program that in turn may be linked to variations in FAST professional development. That is, teachers are assigned to a traditional FAST ten-day professional development or assigned to a new FAST five-day professional development with a long term follow up combined with an electronic resource (i.e., a collection of video clip demonstrating FAST lesson implementation). We hypothesize that teachers who participate in the different FAST professional development trainings may manifest differences in student outcomes measures. That is, the purpose of these student outcome measures is to be sensitive to the fidelity and quality of implementation of the FAST curriculum. The purpose of this paper is to present the development of the measures as well as to argue to their validity.

In a nutshell, the primary targets of the FAST 1 student assessment suite (FSAS) are the science content knowledge and science inquiry covered in FAST 1. The primary science content of FAST 1 is physical science (e.g., matter, buoyancy, states of matter and energy) linked through relational studies (e.g., studies of the water cycle and air pollution) to concepts of ecology (e.g., plant and animal relationships to the environment). We hypothesized that increases in the fidelity and quality of instruction would manifest in greater student science content learning and increased science inquiry performance. Secondary assessment targets of the FSAS are attitudinal targets (i.e., student self-efficacy towards science investigations, motivation towards science and student views of the nature of science). We hypothesized that students' relationship to science would change as they progressed through the FAST curriculum where the students are expected to learn science by participating in science. If students are asked to be scientists and actually carry out their own investigations then how they view science may change.

Content Knowledge-Type and Science Inquiry Frameworks

Knowledge Type Framework. The content and inquiry assessment blueprints for the FSAS were based on a content knowledge-type framework (de Jong & Ferguson-Hesser, 1996; Li & Shavelson, 2001), scientific inquiry (Duschl, 2003) and the FAST curriculum (Pottenger & Young, 1992). The knowledge-type framework provides for a broad definition of science achievement based on the types of knowledge students are expected to learn in science. The content knowledge-type framework includes declarative (knowing that; *facts and concepts*), procedural (knowing how to; *measuring and experimenting*), schematic (knowing why; *explaining models*) and strategic knowledge (knowing when and how knowledge applies; *applying a procedure from one domain to another*).

Shavelson and the Stanford Assessment Lab posit effective and efficient assessment methods that correspond to the different knowledge types. For example, while the extent of declarative knowledge can be easily assessed by multiple-choice items, the structure of declarative knowledge can be assessed via concept maps. Procedural knowledge is best assessed by performance assessments or laboratory practicals. Schematic knowledge can be assessed with multiple choice items and constructed response items (e.g., Why do things sink and float?). Strategic knowledge may best be assessed via performance tasks. Once the specific content to be covered in a curriculum has been identified and the content classified into the different knowledge types, then assessment methods can be identified for the different content pieces.

Science Inquiry. Duschl's conceptualization of the assessment of scientific inquiry points towards the attainment and evaluation of data and evidence and how it is used to create models and explanations in three integrated domains: 1) conceptual (scientific knowledge and reasoning), 2) epistemic frameworks used when developing and evaluating scientific knowledge and the 3) social processes that shape how knowledge is communicated. He argues that, "The assessment of inquiry is best thought of as a set of elements that place emphasis on examining the *processes* of engaging in scientific knowing and learning as opposed to the *products* or outcomes of scientific knowing and learning" (2003, p. 44). Assessing inquiry requires the designing of a task to promote inquiry activities and to capture students' reporting and sharing of information and ideas.

Duschl argues that the core of the inquiry process is about collecting data, transforming that data into evidence, the evidence into models, and finally then into explanations that are then used to developing new questions. That is, that assessment of student inquiry should occur at three transformation points along the Evidence-Explanation (E-E) continuum (Figure 1). In our development of the materials covered in FAST curriculum, we extend the continuum beyond the three transformations and include reformulation. That is, as Duschl suggests deciding on what data is needed and what questions to ask (c.f., the E-E-R continuum). The first transformation is selecting and evaluating data to become evidence, the second analyzing evidence to create models and finding patterns, and third, determining scientific explanations that account for the models and patterns and the reformulation is where students suggest new questions and decide on the data and collection methods they will need. Students share their thinking at each of the transformations and reformulation by engaging in "argument, representation and communication, and modeling and theorizing" (p. 45) and provide us with the opportunity to evaluate their inquiry processes. Critical to the success of assessing along the EER continuum is capturing the reasoning found in student's judgments and explanations. Therefore it is critical to ask students to present their supporting evidence for their explanations (e.g., "What evidence do you have to support your conclusion that objects sink based on their density?"). And if student are working in groups as they do in the FAST 1 curriculum, then students should work in groups. Finally, students could report their findings and conclusions. Furthermore, we see these transformations and reformulations as cyclical and not linear and as such there is no starting point (Figure1). Therefore, students may be asked to respond and provide supporting evidence for different locations on the EER continuum at different times in an assessment task. These transformations generally represent procedural knowledge, however, we believe that schematic knowledge of the whole inquiry method may be assessable

.

<<Insert Figure 1 about here>>

In the case of this study, we are interested in an individual summative assessment of student inquiry, therefore, we engage students in the social aspects of the inquiry processes by starting off the tasks with the students in groups, but then we move them to individual work. Furthermore, this group setting matches the student's FAST group work (a validity match to the curriculum). While the students are engaged in individual work, we can assess students' ability to decide about data, create explanations from evidence with justifications and measure their science communication. Students may be asked to evaluate data, ask questions at different times in the processes. We realize that not all groups will be the same so we will look for a group effect in our analysis of the data.

Linking the knowledge-type framework and Scientific Inquiry. Shavelson's knowledge framework and Duschl's domains and transformations are not independent. Shavelson suggests procedural knowledge and reasoning may represent these transformations.

Content knowledge and science inquiry outcome measures development

Content Knowledge. Following the successful assessment development process utilized by the Stanford Education Assessment Laboratory (SEAL) and the University of Hawaii's Curriculum Research and Development Group CRDG on a previous grant (Embedding Assessments in the FAST Curriculum: The Romance between Curriculum and Assessment, 2005), the Sonoma State University team (SSU) went through the student materials, instructor's guide, evaluation guide, and training guide and began the initial identification of the assessment targets for the FSAS using the knowledge framework. That is, the key concepts, procedures and schema/ models covered in FAST 1 were identified and then classified into the three knowledge types (Table 1). Ultimately we identified 75 major potential science knowledge assessment targets. The topics that were identified were the primary foci of the lessons and materials reviewed.

Second, to reduce the list to a smaller number of valid targets, the initial lists were presented to the CRDG team members (curriculum developers, FAST trainers). In small groups, the CRDG team members then reviewed the initial lists, reviewed the instructional and training materials and began a selection process, as to what were the most important targets covered. Initially, in the previous Embedded Assessment study, we found this task to be quite difficult because the curriculum developers found that all the targets were important. Therefore, the CRDG attended to the bigger ideas and content that bridge larger groups of lessons as opposed to each individual lesson. Each small group team then created reduced lists of targets reflecting the most important targets. These reduced lists served as the main assessment targets.

Furthermore, the CRGD team then described the different elements of inquiry that were important in the FAST curriculum. The main targets identified by Duschl fit the FAST model and those were ultimately chosen. Additionally, SSU and CRDG decided that the pre-test FSAS should take no longer than two instruction periods and that the post-test FSAS should take no longer than three instructional periods. The amount of time spent on the assessing student knowledge for the study was and continues to be a major issue in the project.

Once the assessment targets were chosen and classified into the knowledge-type framework, specific assessment methods were identified for each target. Then, the SSU team began collecting known multiple-choice and constructed-response items from national and state level tests. While we were interested in developing the measures that were closely linked to the FAST curriculum, we also believed that high fidelity implementation would lead to students overall general science achievement. In order to make this link, of the assessment items used in the FSAS, 8 of those are TIMSS, NAEP or state achievement tests. In order to achieve a balance of items, we selected those measures that were for the most part close to the curriculum (FAST 1 content) and a few that were distal to the content (related but not explicitly part of the FAST 1 content).

Table 1:
Partial list of assessment targets identified during first review of FAST 1 materials by potential assessment method.

Targets	Knowledge Type	MC	CR	PA	POE	Concept Map
Absorption (Hydrology)	declarative	*				*
Accumulation (Hydrology)	declarative	*				*
Acid (Ph)	declarative	*				
Atmosphere	declarative	*				
Buoyancy	schematic		*		*	
Calorie	declarative	*				
Climate	declarative	*				
Collecting/Organizing	procedural			*		
Communicating	procedural		*			
Condensation (Hydrology)	declarative	*				*
Constructing Data Table	procedural		*			
Density	declarative	*				
Density Of Gases	declarative	*				
Density Of Liquids	declarative	*				
Designing Experiments	procedural		*	*		
Displacement	declarative	*				
Evaporation (Hydrology)	declarative	*				*
Extrapolating	procedural	*				
Gases	declarative	*				
Graphing	procedural	*	*			
Ground Water	declarative	*				*
Heat Exchange	declarative	*				
Mass	declarative	*				
Matter	schematic		*			
Mixture (Solutions)	declarative	*				
Movement Of Gases	declarative	*				
Percolation (Hydrology)	declarative	*				*

For the multiple-choice and constructed-response test, three different test versions were created and administered to 200 middle school students whose curriculum closely matched the FAST curriculum (Physical Science). In order to validate the link between what each item was intended to target and how students interpreted each question, a test administrator asked students about each question on the different versions and reported problematic and discrepant items. These items were dropped or revised. We then carried out analysis of the items against total score and additional problematic items were reviewed and then revised or dropped from the different test versions. Once the items had been reviewed in this way, one long version of the test was created. This version of the Student Science Achievement Test was then administered to the University of Hawaii Lab School FAST 1 students ($\alpha = .86_{\text{multiple-choice items}}$). Curriculum developers and FAST 1 trainers reviewed and validated (with respect to content) the final version of the FSAS (Table 2).

Table 2
FAST Student Assessment Suite Content Knowledge Test.

Item	Type[a]	Source	Knowledge Type	Description
S1	Mc	Original	Procedural	What happens to two glasses (one hot) overnight?
S2	Mc	TIMSS	Declarative	Air is made of gases?
S3	Mc	Multiple	Declarative	Primary source earth water cycle energy?
S4	Mc	Multiple	Declarative	Snowball internal temperature?
S5	Mc	TIMSS	Schematic	Coastal and inland climate reasons
S6	Mc	Original	Procedural	Variables in how much water to lettuce need study?
S7	Mc	Original	Declarative	Specific heat of object and heat transfer.
S8	Mc	Romance	Schematic	Block floats w/o hole, what happens with a hole?
S9	Mc	Romance	Declarative	What happens to floating object in larger container?
S10	Mc	Original	Schematic	Which graph represents temp of heating water to boiling?
S11	Mc	Romance	Declarative	Ball of equal mass and volume, one hallow, do they both sink?
S12	Mc	Original	Procedural	Which question is the question she wanted to answer?
S13	Mc	MOD	Procedural	Which is did not contribute to different weather readings?
S14	Mc	TIMSS	Schematic	During the day, organisms use up or give off?
S15	Cr	TIMSS	Procedural	Machine X and Y, which is more efficient?
S16	Mc	Romance	Schematic	What happens to density when block of wood is cut?
S17	Cr	Romance	Schematic	Which object can be used to determine density of second liquid?
S18	Mc	NAEP	Declarative	Temperature of freezing of different amount of water.
S19	Mc	Multiple	Declarative	What happens to salt when water evaporates?
S20	Mc	Romance	Schematic	What factor has the greatest effect on sinking or floating?
S21	Mc	Original	Procedural	How much energy to heat water?
S22	Mc	TIMSS	Declarative	What happens to atoms after animal dies?
S23	Mc	Original	Declarative	What happens to water vapor as air temp increases?
S24	Mc	Mod	Declarative	Prediction of mass of melted ice in can?
S25	Mc	Original	Declarative	What is the boiling point of a mixture?
S26	Mc	Original	Procedural	Why do scientist measure something several times?
S27	Mc	Multi	Declarative	What is the best reason why hot air balloons rise?
S28	Mc	Romance	Schematic	Estimate the density of plastic block in two liquids?
S29	Cr	Multi	Schematic	Describe the water cycle
S30	Cr	Mod	Procedural	State the relationship between Insect A and Insect B

[a] Mc— Multiple-choice item, Cr— Constructed-response

Science Inquiry. Once the science inquiry targets are identified and since we found these targets to be procedural knowledge and reasoning, we selected a performance assessment format as the appropriate assessment type. A science performance assessment is a "lab practical" where students carry out an investigation to solve some problem (e.g., find the density of this block using a balance and a beaker). These assessments are valued for their congruence with what happens in the science classroom as well what happens in the science lab. “A science performance assessment comes as close as possible to putting a student in a laboratory, posing a problem and watching as the student devises procedures for carrying out an investigation, analyzing data, drawing inferences from the data and his prior knowledge…” (Shavelson, 1995, p. 59).

A science performance assessment can be characterized by three components called “the triple”: the task (a hands on activity or problem that students are asked to solve), the response format (the nature of the response the student is expected to provide—student notebook) and scoring system (the method used to evaluate student performance) (Ruiz-Primo & Shavelson, 1996b; Shavelson et al., 1991; Shavelson et al., 1998). The *task* invites the student to solve a problem. It requires the use of concrete materials that react and provide feedback to the student, and is related to the curriculum covered. A *response format* provides a place where students record their findings allows students to decide how to present their findings and asks students to justify their answers. The prompt only nudges the student towards the procedures, but does not spell it for him or her. The *scoring system* reflects both the goals of the task and assessment targets (i.e., science inquiry) and captures the scientifically justifiable procedures and allows for insight into student's problem solving abilities.

In order to create a performance assessment that captures FAST 1 content and science inquiry captures the EER continuum we chose a relational study *task*. A FAST1 relational study is an investigation where students explore an ecological situation using physical science principles. Since students in FAST 1 carry out investigations related to pollution in the environment, we decided to emulate this with a performance assessment. We chose the idea of factories polluting a river in which to embedded our assessment items for FAST science inquiry and the EER continuum (Figure 2). In this task students sample water from different locations on Rocky River and test the samples for high levels of pollution. Students must set up comparisons using controls, pollution indicators and limited testing sites.

Fish Deaths in Rocky River

Instructions: You are going to find out if either of two factories (or none) are polluting Rocky River enough to kill fish. You will start working in a group and then finish by yourself.

On one side of Rocky River there is a plastic cup factory. On the other side of the river there is a plastic plate factory. Both factories put pollution into the river, but usually not enough to kill fish. One day, students found dead fish downriver below the factories on Middle Island. Your job is to find out if either of the factories (or none) are putting enough toxic levels of pollution into the river to kill the fish.

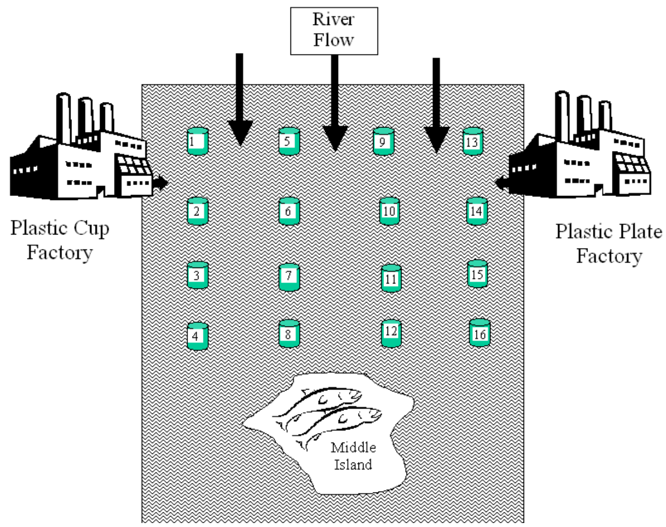

Figure 2 Fish Deaths in Manoa Stream

Students then record their finding on the response format (*notebook*). In the notebook, students are prompted to record their findings and explain why they have reach their conclusions. It is in the *notebook,* students are given the items related to FAST science inquiry and EER continuum. For example, students are asked to decide which factory is polluting the stream (Patterns to Explanations), provide evidence for their conclusions (Patterns to Explanations) and then decide if the data that they are using make sense or not (Data to Evidence) (Figure 3).

Part 3: Conclusions—Working by yourself.

Now using your own data table and working by yourself:

1. Which factory is polluting the stream enough to kill the fish (none, one or both)?

2. Using the evidence you collected, explain how you know which factories if any are polluting enough to kill the fish.

3. Do the data that you collected make sense? How do you know?

Figure 3 Fish Death on Manoa Stream Notebook

The scoring system (rubric) links the student responses to the assessment targets and assigns values to student responses. In order to capture the transformations and reformulations, students are asked to first carry out the investigation with in a group and then asked to repeat the investigation with new sites by themselves. There are multiple opportunities to capture student thinking in the EER in the Conceptual, Epistemic and Social domains. In order assess students in the conceptual domain, we focus on the soundness of their responses. To get at the epistemic domain we tease out whether students understand how science knowledge is developed and evaluated. Finally to get at the Social domain we attend to extent of the students' science communication with in each particular transformation and reformulation. We cross the transformation and reformulation across the three domains and embedded these crosses in the scoring rubric (Table 3).

Table 3
The evidence explanation continuum transformations and reformulation and the Conceptual, Epistemic and Social Process Domains

	Conceptual	Epistemic	Social
Data to Evidence	Is student data correct?	Do students understand the purpose of standards?	To what extent are student's science communication clear, focused with minor technical errors.
Evidence to Patterns	Does the student use the appropriate evidence to describe patterns?	Do students know how to present data in an organized way to make sense to others?	To what extent are student's is the presentation of evidence relevant, clear, focused with minor technical errors. Are data tables clearly labeled?
Patterns and models to explanations	Does the student select both factories (one more than the other) as killing the fish?	Are student explanations supported by evidence?	To what extent are student's explanations clear, focused with minor technical errors.
Deciding what new questions are needed	Does the student select new sites that would be meaningful?	Do student rationales for new sites express the reason for why the new information might be valuable?	To what extent are student's rationales for choosing new sites clear, focused with minor technical errors.

Validity. The major validity threats of the claim that this performance assessment measures an individual's science inquiry knowledge are the group work (Is there an interaction effect between group partners and an individuals performance?) and is this performance assessment actual measuring what we think that it is measuring. We have carried out think alouds on four versions of the performance assessment and found that in most cases students performance on the group tasks superior to the individual work. That is, working in the group, as FAST students should, students perform better in their work, and that working individually students revert to earlier procedural knowledge and reasoning to solve the problems. The think alouds were also used to review how students carry out the investigations and to see if the wording of the performance assessment made sense to students. Finally, we had experts, science graduate students and science teachers (trained biologists) carry out the performance assessment in order to set the standards for the response for the rubric. A total of 272 students completed the Rocky River Performance Assessment.

Student Attitudinal Measures

Not all student outcomes are content related. Attitudinal measures are important when considering the effectiveness of a program especially one that is intended to engage students in science because students' perceptions and attitudes towards science may influence their learning. *Motivation* (Pintrich, 1999, 1993; Haydel & Roser, 2002), *Self Efficacy* (Bandura, 1986; Pajares, 1995,1996) and in the case of science education, views of *the nature of Science* (NOS) (Lederman, 1992) have been found to related to student achievement. Furthermore, based on the Bandura's Social Cognitive Theory (1986, 1977), we incorporate additional motivations constructs including *Science Anxiety* (Britner & Pajares, 2001; Pajares and Urdan, 1996) and *science value* (Britner & Pajares, 2001; Meece. Wigfield, & Eccles, 1990). An 81 item-survey was created from multiple sources to get at students Motivation, Self-efficacy and NOS views. This survey was adapted from multiple sources described below.

Self-efficacy and student lab performance

Bandura (1986) argued that *self-efficacy* is the most influential factor in human functioning. He defined self-efficacy as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (1986, p. 391). Self-efficacy mediates the effects of prior achievement, knowledge, and skills on subsequent achievement. Thus, it is often a better predictor of success than actual abilities. This may help explain why people with similar abilities may have different levels of achievement. Self-efficacy affects behavior by influencing people's behavioral choices, the amount of effort they expend, and the persistence they exhibit in the face of failure.

Most research on science self-efficacy has focused on science teaching self-efficacy (Cannon & Sharmann, 1996) and science self-efficacy as a predictor of career choices (Gwilliam & Betz, 2001; Lusso, Hasper, Albert, Bibby, & Martinelli, 1999). There are few investigations of confidence in science as a predictor of subsequent science achievement, and fewer investigations focusing specifically on laboratory skills or learning through science investigations and studies focusing on the effects of a particular science curriculum. Britner (2002) investigated middle school science students' self-efficacy with respect to science and science lab grades. She found that student science grade self-efficacy was positively associated with the grades. Furthermore, girls' grades were also associated positively with science self-concept and negatively with value of science. For reasons resulting from problematic instructional practices (lab grades might be associated with attendance rather than performance), lab skills self-efficacy was not associated with lab grades.

We hypothesized that since in the FAST 1 curriculum, students are expected to carry out and learn from their science investigations, their judgments about their capabilities to carry out science investigations and learn from these science investigations should increase. And that, in pedagogically strong classrooms their judgments about their capabilities would change more as they learn from their investigations throughout the year.

Instrument Development. A student survey was created to elicit student self-efficacy based on Britner (2000). The Science Investigation Self-Efficacy was assessed with the Lab Skills Self-Efficacy Scale (Britner, 2002). This scale consists of 12 items asking students how sure they are that they can perform specific science process skills commonly used in laboratory activities (Chiapetta & Koballa, 2002; National Research Council, 1996). Britner's items were adapted to match the FAST 1 language. Students estimated their confidence that they could perform each skill on a scale from 0 (no chance) to 100 (completely certain) (Table 4). This was administered to the students in a pre test suite and the post-test suite.

Table 4

Science Investigation Self-Efficacy Scale (Adapted from Britner, 2002).

On a scale of 1 to 100 rank how sure are you that you can…

1. correctly follow directions to complete a science investigation?
2. make appropriate predictions (hypotheses) about what will happen during a science investigation?
3. use laboratory equipment correctly?
4. make accurate measurements during a science investigation?
5. make appropriate observations during a science investigation?

The measures here will be correlated with a direct measure of student laboratory performance—the Rocky River Performance Assessment. Previously, Britner (2002) found that the Lab Skills Self-Efficacy Scale scores were not significantly correlated with subsequent lab grades for these undergraduates. Evidently, the lab grades that students are receiving are not related with the measure of their confidence in their abilities in using the skills that are believed to be the criteria for determining the lab grades. This is unusual because grades generally are correlated with the measure of confidence in the skills needed to get good grades. Britner suspects that the lab grades in their studies may be based more on attendance than on competence in the lesson.

Science Anxiety and Value

In addition to making comparisons between science investigation self-efficacy and science performance assessments, it is important to look at the relationship between self-efficacy and science anxiety (sample item: Just thinking about science makes me nervous) and science value (sample item: I like doing science investigations). Following the lead of Britner & Pajares (2001), we explore the relationship between science value, science anxiety and science self-efficacy as important predictors of science achievement and may also be influenced by the fidelity of FAST implementation.

Motivation and science education

Snow (1994) hypothesized that individual differences in achievement can be seen as a “moment to moment” transaction between characteristics of the person and the situation itself. Snow believed that individuals bring to a task certain cognitive and motivational aptitudes that shape their performance. In order to look at the relationship between students’ achievement and FAST instruction, we decided to explore the relationship between motivation and achievement in FAST. We draw on Dweck and her colleagues’ theory of the organization of achievement-related goals and competence-related beliefs that are linked to academic performance. Through empirical work and logical analysis of why some students engage with and perform better on particular task, she found three sets of motivational processes that predict differences in achievement outcomes. The sets include a student’s beliefs about the malleability of their intelligence, the intellectual confidence and their achievement goals. They propose three motivational types: 1) mastery-oriented students, 2) ego-oriented students and 3) helpless orientation students.

Mastery oriented students are students who believe that intelligence is malleable and can grow over time. They pursue goals in which to develop their intelligence. Ego-oriented students are defined as students who believe that intelligence is fixed and as such students adopt goals in which proving their fixed ability or hiding their inability. If those students have high confidence in their abilities then students view tasks as an opportunity to reinforce their sense of superior ability. Helpless orientation students like the ego-oriented students believe that intelligence is fixed however, they have low confidence in their abilities and as such, these students are thought to be preoccupied with the goal of hiding their inability from others. There is some evidence to suggest that students who are members of traditionally considered inferior intellectually (females in science) may be more likely to adopt this helpless orientation (Dweck, Davidson, Nelson & Enna, 1978; Steel, 1997).

We hypothesize that in a complete and high-quality FAST implementation that students motivational group patterns will be different than in incomplete or low-quality FAST implementation classrooms. That is, as the students are exposed to investigations where they discover their own knowledge, where students are exposed to Socratic inquiry and are asked to think for themselves and come to their own conclusions, their beliefs about what they can learn will be different than in classes where they are not asked to think for themselves and not come up with their own conclusions but rather repeat conclusions already in place. Furthermore, motivational types can be used to further explain differences in achievement.

In order to identify students in the different motivation pattern types, measures of students epistemic beliefs, self-confidence in science ability and goal orientation are collected (Table 5). Then following Roser and Haydel's (2002) method (adapted from Dweck and Henderson, 1989) students will be classified into one of the motivational types. Comparisons between pre and post measure proportions as well as measured differences with respect to fidelity of instruction will be carried out.

Table 5

Motivation targets and sample items

Aptitude Motivation Target	Sample Item.
Epistemic Belief	How well I do in science depends on how smart I was when I was born.
Epistemic Belief	You are born smart in science.
Epistemic Belief	I have to be really smart to do well in science.
Ego Avoidance Goal	It is very important to me that I do not look stupid in my science class.
Ego Avoidance Goal	One of my main goals in science class is to avoid looking like I can't do my work.
Ego Mastery Goal	I like the work in my science class best when it really makes me think.
Ego Mastery Goal	An important reason I do my science work is to master challenging concepts.
Perceived Ability Goal	Our teacher points out those students who get good grades as an example to all of us.
Perceived Ability Goal	Our teacher lets us know which students get the highest scores on tests.
Perceived Task Goal	Our teacher wants us to really understand the concepts, not just to memorize facts.
Perceived Task Goal	Our teacher gives us time to really explore and understand new ideas.

Overall the number of items and reliabilities for items in each group is as follows.

All Motivation Constructs	Number of Items
Self Confidence in Science Ability	6
Epistemic Beliefs	4
Inquiry Epistemic Beliefs	4
Pee Epistemic Beliefs	4
Ego Avoidance Goal	5
Ego Mastery Goal	5
Ego Performance Goal	5
Perceived Ability Goal	5
Perceived Task Goal	7

Student nature of science measures

The final attitudinal measures used in these analyses are student views of the nature of science measures. For many years, many scientists and science educators have agreed that an objective of science education is that students have an informed conception of the nature of science (Abd-El-Khalick, Bell and Lederman, 1998, Duschl, 1990; Lederman, Abd-El-Khalick, Bell and Schwartz, 2002). Nature of Science (NOS) refers to the epistemic and sociology of science—science as a way of knowing as well as the values and beliefs inherent to scientific knowledge and it development (Lederman, 1992). While some aspects of the NOS are controversial (the issue of an objective reality) some are more accessible to K-12 students. Lederman et al. (2002) argue that views of NOS that are relevant to the daily lives of students are: "scientific knowledge is tentative, empirical, theory-laden, a produce of human inference, imagination and creativity and socially and culturally embedded" (p. 499).

In this study we are interested in the changes in students NOS views as they progress through the FAST curriculum. In the FAST curriculum students are considered the scientists and are asked to explore the world and come up with their own knowledge. We hypothesize that students in classrooms that have a higher degree FAST implementation fidelity will have a more realistic view of the NOS than students in the lower degree of implementation fidelity classrooms. Furthermore, in this pre-post test study design we would expect that students would have changes in their NOS views from more idealistic (unacceptable) to more realistic (acceptable) as they progress in the lessons and that in classrooms with higher degrees of FAST implementation fidelity, we would expect greater gains.

We use a Likert-scale questionnaire method to carry out this investigation. Although these standardized tests have been leveled with criticism because of their lack of validity, they reflect the researchers views of the nature of science and their usefulness. Since in this study we are interested in the student NOS views changes we believe the Likert-scale system works because we are not necessarily interested in the absolute NOS view value rather the differential view. Furthermore, in order to assure validity, we use items from several sources to develop our items, we verified the items with FAST curriculum developers, scientists and FAST trainers. This was to assure ourselves of a broader more valid view valid of the NOS.

We developed a 10 Likert-scale items that address the NOS issues related to validity to everyday life, a creative endeavor, absolute knowledge, represent diverse populations, amoral and developmental (Table 6).

Table 6.
Nature of Science Questionnaire: Problematic domains and sample items.

Nature of Science Problematic Domains	Sample Item
Scientific Theories and Laws as absolute	Scientists are always right.
Science as Socially Embedded	All people who study hard and are smart can learn to be a good scientist.
Science as amoral	Science knowledge is not good or bad.
Science does not involve creativity	Scientists always get the same results.
Science is tentative and developmental	Scientific knowledge can change over time.
Science is useful.	Scientific knowledge can be useful away from school.

Results

The purpose of this paper is to present the development of the measures as well as to argue to their validity. This section describes the piloting of the assessments. A total of 428 took both the pre-test assessment suite and the post-test assessment suite. Of those, 365 students have complete data sets. The validity analyses were completed on the 365 students. The students who took the assessment suite were 7th or 8th grade students in schools located in the State of Hawaii. All students were taught the FAST 1 materials.

Content Knowledge Student Achievement Test Validity

Students who took both the pretest and posttest, performed better in the posttest that the pre test (Table 7). The difference between the pre and posttest was significant ($T = 6.28$, $p = .000$). The instrument appears to be sensitive to student learning in the FAST classroom.

Table 7

Achievement test scores for student with complete data sets.

	N	Mean	STD
Pre Test	365	10.9	4.7
Post Test	365	12.3	4.9

Nature of Science

For the Nature of Science Survey, we used data from 7 of the items in the survey ($\alpha = .81$). We interpret that a high score on the nature of science survey as indicating that student has a great control over science than a lower score. Although we administered a series of 10 items to the students, 3 of the items did not behave as predicted and were dropped from the analysis. We found, as expected, that the pre and post nature of science surveys items were correlated with each other ($r = .67, p < .01$). We also found that the nature of science survey score was positively correlated with the posttest score ($r = .52, p < .01$). We expected that the high performing students would have higher nature of science scores than low performing students. We found that high performing students did have larger nature of science scores than low performing students and did not find differences between pre and posttests with in groups.

Student Science Investigation Self-Efficacy

For the 328 students with complete data sets including the science investigation self-efficacy, we found that all 12 items worked well together as expected ($\alpha = .90$). We found that the correlation between the pre and post surveys was significant ($r = .59, p < .01$) and that the correlation between posttest achievement scores and post survey scores was positive and significant ($r = .35, p < .01$). These relationship between the post survey and posttest achievement scores is as expected. We also found differences between high and low performing students on their scores, as expected.

Motivation

We present the results of the motivation survey in three parts, epistemic beliefs, confidence, and mastery learning.

Epistemic Beliefs. In our survey, there were a total of 4 items that were used to get students beliefs about the nature of knowing. A high score on this survey implies that the student believes that knowing in science is fixed and that the ability of learning this knowledge is predetermined at birth (i.e., You are born smart in science). The relationship between student epistemic beliefs between pre and post surveys was significant ($r = .48, p < .01$). As expected, the relationship between achievement and epistemic beliefs was negative ($r = -.35, p < .01$). Students with lower achievement believe that no matter how hard they try, they cannot learn, while high achieving students believe that if they try harder, they can learn more. Additionally, we found that student epistemic scores were negatively correlated with student self-efficacy scores ($r = -.28, p <$

.01). The more students believe that intelligence is fixed; the less they believe that they have control over their learning from science investigations. Using factor analysis only one Eigenvalue greater than 1 was found, which explained 51% of the variance in the values for Motivation Epistemic Beliefs.

Confidence. The six items that were used to measure students' confidence in their learning (e.g., I can learn science) worked well together ($\alpha = .79$). The correlation between pre and post surveys was positive ($r = .48$, $p < .01$). The correlation between post survey confidence scores with post achievement test was also positive ($r = .26$, $p < .01$). This relationship is weaker than expected. The post survey confidence scores vs. student self efficacy scores were related and strong ($r = .58$, $p < .01$). The more student believe that they are in charge of their own learning the more confidence they had that they would be successful. The post survey confidence scores vs. student epistemic beliefs scores were negative related, as expected ($r = -.32$, $p < .01$). The more students believe that their intelligence is fixed the less likely they were to have confidence in their learning. Using factor analysis only one Eigenvalue greater than 1 was found, which explained 50% of the variance in the values for Motivation Confidence.

Mastery. Students' beliefs about mastery learning were assessed using 5 times ($\alpha = .80$). In general, mastery oriented students are students who believe that intelligence is malleable and can grow over time. These students tend to be higher performing students. In our survey we found a positive correlation between pre and post surveys ($r = .56$, $p < .01$). We also found a positive correlation between post survey mastery scores and posttest scores ($r = .13$, $p < .05$) as well as mastery scores and gain scores ($r = .15$, $p < .01$). However, other indicators suggest the validity of the mastery survey score. Post survey mastery score was found to be positively related to student science self-efficacy ($r = .33$, $p < .01$) and post survey mastery scores were negatively related to epistemic belief scores ($r = -.32$, $p < .01$). Using factor analysis only one Eigenvalue greater than 1 was found, which explained 55% of the variance in the values for Motivation Mastery.

Science Value

Students rated the value of science (i.e., I find science interesting) using a five point Likert Scale. The value of science was measured using 7 items ($\alpha = .88$). In general, student science achievement is positively related to value of science. The more that the students find science as interesting and likeable, the more likely they are to perform well. Indeed in this study, we found that students post achievement scores were positively correlated with value of science scores ($r = .125$, $p < .05$). We also found that student self-efficacy scores were positively related to students' value of science scores ($r = .368$, $p < 0$). Using factor analysis only one factor was found with an Eigenvalue greater than 1, which explained 59% of the variance in the values for science value.

Science Anxiety

Students rated their anxiety to science (i.e., I get really uptight during science tests) using a five point Likert scale with 7 items ($\alpha = .83$). In general the more anxious a student is about performance in science the less that student achieves. And indeed we found this relationship as expected. Student anxiety towards science was negatively correlated with post test achievement scores ($r = -.38$, $p < .01$), negatively correlated with value of science scores ($r = -.41$, $p < .01$) and negatively correlated with student self efficacy scores ($r = -.50$, $p < .00$). Using factor analysis only one factor was found with an Eigenvalue greater than 1, which explained 50% of the variance in the values for science anxiety.

Conclusions

The purpose of this paper was to argue for the validity of the student achievement suite including the student attitudinal survey. We conducted think alouds on all the achievement measures. We built our attitudinal measures from known items and surveys. The reliability of these measures is strong and in line with those found in the studies from which we build out measures. Furthermore, the relationship between the different constructs were as anticipated. Based on these results we are confident in the validity of these measures.

References

Abd-El-Khalick, Bell & Lederman, (1998). The nature of science and instructional practice: Making the unnatural natural. Science Education 82, 417-437.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Britner & Pajares, (2001) Self efficacy beliefs, motivation, race and gender in middle school science. Journal of Women and Minorities in Science and Engineering vol 7, p 271-285.

Britner, S. L. (2002). Science self-efficacy of African American middle school students: Relationship to motivation self-beliefs, achievement, gender, and gender orientation. Unpublished doctoral dissertation, Emory University, Atlanta, GA.

de Jong, T., & Ferguson-Hesser, M. G. M. (1996). Types and qualities of knowledge. Educational Psychologist, 31, 105-113.

Duschl, R. (1990) Restructuring Science Education. New York: Teachers College Press

Duschl, R.A. (2003). Assessment of Inquiry. In M. Atkin & J.E. Coffey (Eds.), Everyday Assessment in the Science Classroom (pp. 41-59). Arlington, VA: NSTA press.

Dweck, C. & Henderson V. (1989) Predicting individual differences in school anxiety in early adolescence (ERIC Document Reproduction Service No. ED310884).

Dweck, C., Davidson, W., Nelson, S. & Enna, B., (1978) Sex difference in learned helplessness: The contingencies of evaluative feedback in the classroom and III. An experimental analysis. Developmental Psychology. 14, 268-276.

Haydel, A. & Roser, R. (2002). On motivation, ability and the perceived situation ins science test performance: a person centered approach with high school students. Educational Assessment, 8 2. 163-190.

Kagan, D. (1990). Ways of evaluating teacher cognition: Inferences concerning the Goldilocks Principle. Review of Educational Research, 60(3): 419-469.

Lederman, Abd-El-Khalick, Bell & Schwartz, (2002). Views of nature of science questionnaire: toward valid and meaningful assessment of learners' conceptions of nature of science, Journal of Research on Science Teaching, 39, 6, 497-512.

Li, M., & Shavelson, R.J. (2001). Examining the Links between Science Achievement and Assessment. A paper presented at the AERA Annual Meeting Seattle, WA.

Meece, J., Wigfield, A., & Eccles, J. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment and performance in mathematics. Journal of Educational Psychology, 82, 60-70.

National Research Council. (2000). Inquiry and the National Science Education Standards. A guide for teaching and learning. Washington, D.C.: National Academy Press.

National Research Council. (2001). Classroom Assessment and the National Science Education Standards. Washington, D.C.: National Academy Press.

Pajares, F. and Urdan, T. (1996) Exploratory factor analysis of the Mathematics Anxiety Scale. Measurement and Evaluation in Counseling and Development, 2, 9, 35-47.

Pintrich, P. R. (1999). Motivational beliefs as resources for and constrains on conceptual change. --: Pergamon Press.

Pintrich, P. R., Marx, R., & Boyle, R. (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.

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

Richardson, Virginia. (1996). The Role of Attitudes and Beliefs in Learning to Teach. In J. Sikula (Ed.), Handbook of Research on Teacher Education, 2nd Edition (pp. 102-118). New York: Macmillan.

Shavelson, R. & Stanford Educational Assessment Lab. (2004, June). CAESL and the Future of Science Assessment. Keynote address at the Center for Assessment and Evaluation of Student Learning Conference. Carmel, CA.

Shavelson, R. & Towne, L. (2001). Scientific Inquiry in Education. National Resource Council: National Academy Press.

Shavelson, R. & Young, D. (2001). Embedding Assessments in the FAST Curriculum: On Beginning the Romance among Curriculum, Teaching and Assessment. NSF Grant.

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52, 613-5629.

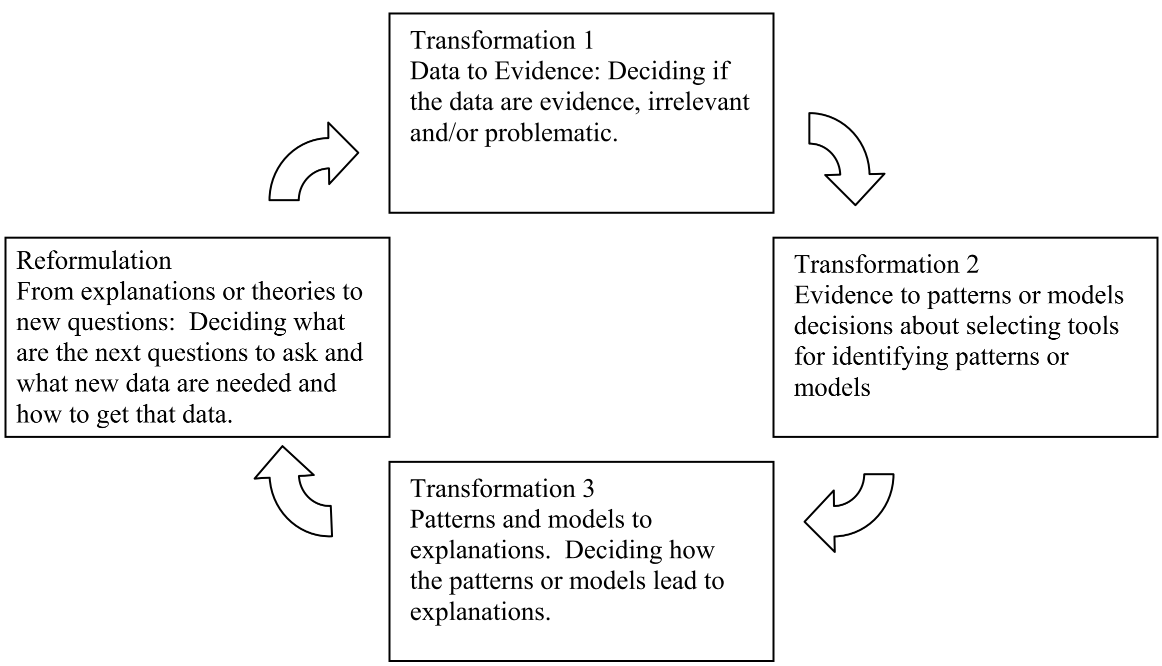

Figure 1: Three transformations and reformulation Science Inquiry Targets.