CRESST REPORT 811

MEASURING FIDELITY OF IMPLEMENTATION-METHODOLOGICAL AND CONCEPTUAL ISSUES AND CHALLENGES

JANUARY, 2012

Ellen Osmundson

Joan Herman

Cathy Ringstaff

Yunyun Dai

Mike Timms

National Center for Research on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

Measuring Fidelity of Implementation- Methodological and Conceptual Issues and Challenges

CRESST Report 811

Ellen Osmundson, Joan Herman, and Yunyun Dai
CRESST/University of California, Los Angeles

Cathy Ringstaff and Mike Timms
WestEd

January, 2012

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported by prime sponsor number R305B070354 from the US Department of Education to WestEd, grant number 5387S07-091, from the American Institutes for Research with funding to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

To cite from this report, please use the following as your APA reference: Osmundson, E., Herman, J., Ringstaff, C., Dai, Y., & Timms, M. (2012). *Measuring fidelity of implementation- Methodological and conceptual issues and challenges.* (CRESST Report 811). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

MEASURING FIDELITY OF IMPLEMENTATION -

METHODOLOGICAL AND CONCEPTUAL ISSUES AND CHALLENGES

Ellen Osmundson and Joan Herman
CRESST/University of California, Los Angeles

Cathy Ringstaff
WestEd

Yunyun Dai
CRESST/ University of California, Los Angeles

Mike Timms
WestEd

Abstract

A central challenge in efficacy studies centers on the issue of "fidelity of implementation," that is, the extent to which participants use the curriculum specified by curriculum developers. In this study, we describe and discuss a "fidelity of implementation" model using multiple methods and instruments to compare two versions of a science curriculum and embedded assessment system. We present results from our validation study and discuss the challenges in determining the extent to which teachers use a curriculum as designed. We focus as well on the practical curriculum implementation issues amidst the needs and perspectives from different stakeholders.

Introduction

The No Child Left Behind Act of 2001 produced an explosion of interest in the use of assessment to measure and improve student learning. Evidence from classroom-level assessment is clear: Teachers' ongoing use of assessment to guide and inform instruction—formative assessment—can lead to statistically significant gains in student learning (Black & Wiliam, 1998, 2001, 2004). Taking the evidence on formative assessment to heart, a well-known, well-established hands-on science curriculum was recently revised to include an embedded assessment system. This system is designed to provide teachers with evidence of student learning to effectively guide instruction and support student achievement.

In 2007, researchers received funding from the Institute of Education Sciences (IES) to conduct a Goal 3 Efficacy Study, utilizing a randomized controlled trial to determine which hands-on science curriculum (i.e., traditional versus one with an embedded assessment system) leads to higher student achievement. Both groups of teachers used the standard curriculum, and were expected to teach all investigations in the modules, as well as follow the intended instructional sequence. Teachers in the control group used the original curriculum assessments – general tasks that follow the instructional sequence. Treatment teachers used an enhanced assessment system, one comprised of a series of curriculum–aligned embedded tasks for each lesson, and for each series of investigations. The embedded assessment system was designed to support teachers' knowledge and understanding of students' conceptual development as they engage in hands-on science experiences.

A central challenge in efficacy studies centers on the issue of "fidelity of implementation," that is, the extent to which participants use the curriculum in the manner in which curriculum developers specify. Implementation of the curriculum is assumed, but often a realistic picture of "full implementation" is incomplete due to measurement and logistical challenges. For example, how do we know if teachers follow the instructional sequence outlined by the curriculum or if they deviate from the intended sequence? Are lessons or additional materials added to the curriculum, or portions of lessons omitted? Critical as well to understanding the efficacy of a curriculum is information on the *quality* of the implementation. How well, for example, do teachers understand the concepts they are teaching? Does their understanding support accurate interpretation of student responses? Can teachers respond to students' needs with appropriate instructional next steps? Answers to these questions are critical to determining which curriculum is "better," that is, more effective, as well as the *specific factors* associated with the implementation that contribute to student learning.

Perspective

To make valid comparisons between the outcomes associated with the use of the two versions of the science curriculum under investigation, we needed to determine the extent to which teachers implemented the curriculum with fidelity. Documentation and description of implementation levels are essential to determine if student performance can be associated with the use of the curriculum being studied. These data can also provide information about which aspects of the curriculum appear to contribute to improved student achievement. The information is particularly important if effects are inconclusive or negative (e.g., no difference between treatment and control conditions), and to determine if the outcomes can be attributed to the curriculum itself, or to curriculum implementation factors (e.g., partial or incomplete curriculum implementation).

This paper focuses specifically on how researchers conceptualized fidelity of implementation, designed instruments to measure it, and validated these instruments, based on a conceptual analysis of the curriculum. We also discuss the challenges in determining the extent to which teachers use a curriculum as designed. Further, we explore the importance of measuring fidelity of implementation in efficacy studies, and discuss methodological considerations for understanding curricular impact on student learning.

Overview of Current Study

One hundred and ninety two teachers (in seventy schools) with prior experience teaching the targeted science curriculum in its “traditional” form were randomly assigned by school to control (traditional curriculum) or treatment (revised curriculum with embedded assessments) condition. Both groups were expected to follow the curriculum, and make full use of all lessons, tasks, and strategies during instruction.

During Phase 1 of the study, a practice year, control and treatment teachers engaged in professional development to deepen their content knowledge and familiarity with the curriculum to support high fidelity of implementation. Teachers used the standard curriculum, were expected to teach all investigations in the modules, and use all of the assessment tasks provided. Control teachers used the original curriculum assessments – general tasks that follow the instructional sequence. Treatment teachers were provided with professional development to familiarize them with the embedded assessments in the curriculum, to support their analysis and interpretation of student work, and to learn to use embedded assessments to decide on appropriate next steps for instruction. Additionally, both groups of teachers were provided guidelines and models of “full implementation” of the curriculum, with specific information on how to use the curriculum to instruct and assess their students.

For treatment teachers, use of the embedded assessment system provided a detailed and conceptually based approach to assessment. In each lesson, treatment teachers used a document entitled “At a Glance,” a brief description of the important concepts in the lesson. During instruction, students in treatment classrooms used science notebooks to record their information and observations, and response sheets, worksheets that guided the investigations. After each science lesson, treatment teachers were asked to review student notebooks for 10 minutes using an informal coding guide – “What To Look For” – that described targeted student responses. At the end of a series of lessons, students in the treatment classrooms took benchmark assessments, called “I-checks” that included multiple-choice and open-ended items tightly aligned with the curricular concepts. Using a coding guide, treatment teachers examined student work and recorded student responses on an assessment sheet. Treatment teachers met with building colleagues in “Study Groups” to discuss results from the I-checks, examined patterns and trends in the data, and decided upon next instructional steps. These strategies formed the central components for measuring fidelity of implementation.

Phase 2 of the project was the study year, during which time participating teachers implemented two units of the curriculum under investigation. Data from Cohort 1 teachers and students have now been collected and analyzed. This paper presents data from Cohort 1 teachers who have completed both phases of the study (N=39 teachers).

Fidelity of Implementation Defined

In developing fidelity of implementation instruments for the efficacy study, researchers worked with curriculum developers to understand the important instructional and assessment concepts underlying the curriculum. Curriculum developers examined the curriculum, and developed a protocol for “full implementation,” outlining a model that described more than “business as usual” by emphasizing best use and best practices of the curriculum. We further analyzed the components of full implementation, and developed a conceptually based theory of action for full implementation. Table 1 displays the components for full implementation of the curriculum. This matrix formed the basis for the development of implementation instruments in the study.

Table 1

Full Implementation Matrix

Control	Treatment
Instruction	
• Teach all investigations in the outlined sequence • Ask all relevant questions • Follow lesson wrap-up • Use Content & Inquiry chart • Use Word Bank	• Teach all investigations in the outlined sequence • Ask all relevant questions • Follow lesson wrap-up • Use Content & Inquiry chart • Use Word Bank
Assessment	
1. Analyze students' work	1. Analyze students' science notebooks 10 minutes/day after instruction
2. Use a scoring/coding guide to analyze student work	2. Use a scoring/coding guide to analyze student work
3. Record observations of students during class (e.g., in small groups, 1:1 conversations)	3. Record observations of students during class (e.g., in small groups, 1:1 conversations)
4. Analyze student work for patterns and trends	4. Analyze student work for patterns and trends
5. Analyze observations for patterns and trends	5. Analyze observations for patterns and trends
6. Plan further instruction based on patterns and trends in student work	6. Plan and provide further instruction based on patterns and trends in student work using "next step" strategies
7. Check on students' understandings at the end of a lesson or an investigation	7. Check on students' understandings at the end of a lesson or an investigation
8. Engage students in self-assessment of science learning	8. Engage students in self-assessment of science learning
Treatment Only	
9. Not applicable	9. Use I-Checks after each investigation.
10. Not applicable	10. Code I-Checks, examine results for patterns and trends
11. Not applicable	11. Meet with Study Group to discuss I-Check results.
12. Not applicable	12. Engage students in next-step strategies for increasing understanding based on I-Check results.

Research Questions, Methods, and Findings

Measuring how well teachers implemented the specified curriculum is critical to understanding which approach, traditional or embedded assessment system, is more effective in supporting student learning. Several questions guided our study:

1. How do we effectively measure fidelity of implementation?

2. How can fidelity of implementation measures be validated?
3. What practical challenges are associated with developing effective measures of fidelity?

Research Question 1: How Do We Effectively Measure Fidelity of Implementation?

Because this was a complex curriculum, one that involved teaching challenging scientific concepts over an extended period of time, we needed more than a single measure to capture the interconnected, challenging components teachers were implementing in the study. Thus, we developed a suite of tools to measure fidelity of implementation, rather than relying on a single indicator. These tools were aligned with implementation concepts and curricular concepts and focused on implementation of instructional and assessment components. The measures are described in the following section.

Direct measure of teacher content knowledge and pedagogical content knowledge. This measure assessed content knowledge and pedagogical content knowledge about magnetism and electricity, the topic of one of two curriculum units that all study participants implemented. The measure was administered as a pre/post-test, before and after teachers implemented the curriculum twice, in two subsequent years. Three item types corresponded to different aspects of teacher knowledge:

a) content items, as a proxy for teachers' understanding of science concepts;

b) analysis and interpretation of student work items, as a proxy for teachers' pedagogical content knowledge; and

c) next-instructional steps items, as a proxy for teachers' instructional (pedagogical) knowledge.

This direct measure provided information about teacher knowledge, skills, and practices. We theorized that implementation of the curriculum might vary depending upon teacher knowledge, and developed a measure to assess that knowledge.

Teacher logs. Weekly online teacher logs were developed to document implementation of the curriculum and assessments, to provide an overall gauge of fidelity of implementation for program constructs. General reporting categories in the teacher log included: amount of time students engaged with the curriculum; amount of time teachers assessed student work; use of specific instructional strategies; use of assessment resources and strategies; and levels of student understanding. Teacher logs provided self-report information from teachers *as they were implementing* the curriculum – instruction and assessments. Because teachers were familiar with the curriculum (teachers had taught the curriculum a minimum of two times prior to participating in the study), the logs served as a reminder to teachers what they were expected to implement, and provided researchers with an ongoing stream of information about how and in what ways teachers were using the curriculum.

Observations and interviews. The *classroom observation protocol* paralleled the core instructional and assessment components of the curriculum. Observations served as reference for follow-up interviews regarding curriculum implementation. *Interviews* paralleled the classroom observation components, and provided information on teacher strategies for providing student feedback, making instructional decisions, observing students for patterns and trends, and approaches to analyzing and interpreting student work. Interviews highlighted processes and thinking not captured through teacher logs, observations, or surveys. Both interviews and observations were coded for the *frequency of assessment use,* as well as the *quality of assessment use.* These data represented information collected *during* instruction, and offered the opportunity for a third party observer to document the process, and assess both the frequency (quantity) and quality of assessment use.

Data Sources and Analyses

To address Research Question 1, we present information on the contributions of each measure of fidelity of implementation, and explore how these measures capture (or do not) the teachers' instructional and assessment practices. Table 2 below illustrates our data sources.

Table 2

Data Sources Used for Measuring Fidelity of Implementation

Data sources	Teacher assessment knowledge	Teacher science content knowledge	Teacher use of assessments	Teacher analysis and interpretation of student work
Observations	X	X	X	X
Interviews	X	X	X	X
Teacher logs			X	X

Teacher content assessment. The teacher content assessment addressed three concepts from one of two modules taught in the study: magnetism, electricity, and electromagnetism. Limited resources prevented development of a more than one content assessment. Research indicates that elementary teachers are relatively uncomfortable and unfamiliar with teaching science, particularly physical science (Olson, Martin & Mullis, 2008). Recent research has also highlighted teachers' limited formative assessment practice, particularly their capacity to analyze and interpret student work, and provide appropriate instructional next steps (e.g., Heritage & Vendlinski, 2006; Herman, Osmundson & Silver, 2010; Osmundson, Dai, & Herman, 2011).

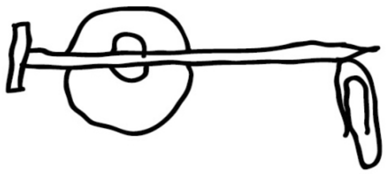
Content knowledge. The reliability of the 29-item multiple-choice items was moderate: subscales (magnetism, electricity, and electromagnetism) achieved alphas of .44 -.46. Even after deleting poorly performing items, the highest performing scale reached an alpha level of .65, as did the combined set of items. However, the items on the measure were designed for students (and paralleled the questions and question types on the student assessment), so the relatively low alphas are most likely due to range restriction and limited variation in teacher performance.

The decision to use items designed for students was made for several reasons. First, elementary teachers frequently have limited knowledge of physical science (Olson et al. 2008). Consequently, the assessment was designed to include items that were accessible to teachers. Second, during project professional development, teachers used the instructional materials they were responsible for implementing in classrooms, to learn or deepen understandings of concepts that they would be teaching. Thus, the content on the assessment reinforced the essential concepts teachers were expected to teach in the Magnetism and Electricity Module.

Analysis of teachers' pre/post content knowledge on multiple-choice items showed that both treatment and control teachers improved their scores for all magnetism and electricity concepts. These results suggest several things about the instrument: first, the measure is sensitive enough to detect changes in teachers' knowledge and understanding. Second, teachers were indeed learning content as they implemented the modules. Increased content knowledge is associated with greater efficacy in teaching (see for example, Black and Wiliam, 1998).

Pedagogical content and pedagogical knowledge. In addition to the multiple-choice items, the content assessment also asked teachers to answer open-ended questions such as the one in Figure 1, to elicit their understandings of content, their pedagogical content knowledge, and their pedagogical knowledge. The format of the assessment questions builds upon work by other researchers (see, for example, Heller, J., Daehler, K., Shinohara, M., & Kaskowitz, S., 2004) interested in measuring teacher pedagogical content knowledge.

1.22 Anne is investigating objects and magnets. She made this observation in her science journal.



"I was surprised! A nail was stuck to the magnet. When I accidentally touched the nail to a paper clip, the paper clip stuck to the nail. I wonder why that happened?"

a. Explain to Anne why the paper clip stuck to the nail. Use diagrams or pictures if necessary.

Anne and her friend were asked by her teacher why they thought the paper clip stuck to the nail. Here are their responses to the question:

Anne's response: The paper clip turned into a magnet too.

Anne's friend's response: The nail gets stuck on the magnet, and the nail turns into a magnet, so the paper clip can stick on the nail.

b. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

c. If these students were in your class, what would you do next in your instruction to help the students learning progress?

Figure 1. Teacher content assessment: Magnetism and electricity module.

There were three scales used to code the content assessment items:

1) content items were scored based on 1-correct/0-incorrect scale.

2) pedagogical content knowledge items were scored based on a 4-point-scale. A score of 0 was used for a non-response or irrelevant response, while 3 reflected a complete and accurate description of student understandings.

3) pedagogical knowledge items were scored based on a 4-point-scale. A score of 0 indicated a non-response or irrelevant response, and 3 indicated a complete, accurate understanding of appropriate next-steps for instruction.

Three raters participated in the scoring, all experienced science educators who were specially trained on the scoring rubric and familiar with the curriculum module. Pre- and post-test responses were scored together, with scorers blind to testing occasion. Based on a 25% sample of the responses that were double scored, reliability of scoring ranged between 76% agreement to 96% agreement (see Tables 3 and 4).

Table 3

Cohort 1: Pre-assessment, Inter-rater Reliability, Open-ended Responses

Comparison rater	Rater 1	Rater 2
2	0.96	
	<.0001	
3	0.90	0.86
	<.0001	<.0001

Note. Pearson correlation coefficients, N = 63.
Prob > |r| under H0: Rho=0.

Table 4

Cohort 1: Post-assessment, Inter-rater Reliability, Open-ended Responses

Comparison rater	Rater 1	Rater 2
2	0.86	
	<.0001	
3	0.91	0.76
	<.0001	<.0001

Note. Pearson correlation coefficients, N = 126.
Prob > |r| under H0: Rho=0.

Table 5 displays score reliabilities for the pre- and post-performance teacher assessment. Results show reasonable reliability for the analysis and interpretation and next step subscales, particularly given the small number of items constituting each. Scores for the content knowledge questions were less reliable than the other two areas, which may be in part due to the small number of items and potential ceiling effects (a total of seven content knowledge items).

Table 5

Cohort 1 Score Reliabilities for Performance Items on Content Assessment

Items	Pre	Post
Content knowledge	0.51	0.48
Analysis and interpretation	0.73	0.81
Next instructional steps	0.79	0.84

As a reminder, we developed this measure for the efficacy study because we believe that if teachers are to implement a curriculum well, they need to understand the content. Further, teachers must also have the capacity to analyze and interpret student work, and provide appropriate next instructional steps for students based on these sound analyses to fully implement the curriculum. The content assessment is designed to capture that knowledge and practice.

How important are the data from the content assessment to our model of fidelity of implementation? Results suggest that teachers, both control and treatment, benefited from implementing the curriculum with fidelity. In essence, teachers learned important concepts by teaching the curriculum. The teacher content assessment, closely aligned with curriculum concepts, captured these changes in teacher knowledge. Future analyses of student achievement data will provide evidence of the impact teacher knowledge has on assessment use, and student impact.

Research Question 2: How Can Fidelity of Implementation Measures Be Validated?

To help answer research question 2, we describe and present information on three conceptually aligned measures: a) teacher logs designed to reflect curriculum concepts and strategies; b) classroom observations tied to the science curriculum; and c) interview questions aligned with the curriculum. Each of the measures contained concepts that were aligned with the curriculum and were designed to provide evidence of fidelity of implementation of the curriculum under investigation. We also discuss our validation process for the measures and how the measures, when combined, support our fidelity of implementation model.

Teacher logs. Teacher logs were developed for use in this study as a time-efficient, cost-effective tool for gathering implementation data for *all* teachers. Log questions paralleled instructional and assessment practices critical to effective curriculum implementation. These items in the logs were culled directly from the curriculum, and were tied to the model of full implementation. Log completion rates varied by teacher and were equivalent for treatment and control teachers; most teachers completed 8 logs during a 12-week-unit, with a few teachers completing as many as 14 logs, and one teacher submitting only 2 logs.

During initial and subsequent professional development sessions, teachers were provided with examples and opportunities from the logs to develop their understandings of the elements or components of the curriculum. For example, what was the project definition of "providing feedback" to students based on analysis and interpretation of the work? Significant time and attention was dedicated to developing clarity and understanding of the components so that all project teachers shared the same conception of full implementation of the curriculum. This work was particularly important because of teachers' familiarity with the curriculum, given that only teachers who had taught the curriculum previously were recruited for this study. We wanted to ensure a common definition of "feedback," and "looking at student work" for all project teachers. Absent that common understanding, accurate measurement of fidelity of implementation is flawed.

Control and treatment log questions were parallel to provide reasonable assurance that both groups were engaged in the same types of instructional and assessment activities. Treatment logs contained additional items that were specific to the embedded assessments. All teachers reported how much time they engaged in instruction and assessment each week, as well as specific details about the nature of those activities.

Table 6 shows the sample of teacher log questions. This measure was aligned with important curriculum constructs, and parallels the model for "full implementation."

Table 6

Teacher Log Questions: 2009 – 2010

All Units Magnetism and Electricity, Structures of Life, Water
Science Curriculum Time
1. How many days did you teach science using the curriculum this week
2. On the days that you did teach science, approximately how many minutes did you spend each day?
3. This week, approximately how many minutes each day did you spend looking at student work after teaching science?
Use of Assessment[a]
During science instruction, how many days did you engage in the following activities?
4. Plan and use an assessment (e.g., student response sheet, student sheet, notebook entry).
5. Use a scoring/coding guide to analyze student work
6. Record observations of students during class (e.g., in small groups, 1:1 conversations)
7. Analyze student work for patterns and trends
8. Analyze observations for patterns and trends
9. Plan and provide further instruction based on patterns and trends in student work using “next step” strategies
10. Check on students’ understandings at the end of a lesson or an investigation
11. Engage students in self-assessment of science learning
Treatment-Specific Questions[a]
12. Administered an I-Check Benchmark Assessment
13. Used coding guides in the Benchmark Folio to code I-Check items
14. Selected and used a next-step strategy
15. Conducted student self-assessment sessions based on I-Check analysis

[a]Scale = Number of times/week teacher reported engaging in activities.

Log data suggest that teachers regularly engaged in assessment, including providing individual, written feedback; using scoring guides; recoding observations; checking student understandings at the end of investigations; and using data to guide subsequent instruction. Teacher logs, in conjunction with teacher observations and interviews, provided information about how and in what ways control and treatment teachers were implementing the curriculum.

We conducted a preliminary factor analyses to better understand how logs functioned as an indicator of fidelity of implementation (Table 7). Factor 1, a proxy for general information about implementation (which included frequency of and amount of time teaching the science curriculum, and the evaluation of and feedback on student work), accounted for

56% of the total variance in the model. Factor 2 identified a useful single-item measure of the minutes/day spent teaching the science curriculum, which is only moderately correlated with days per week teaching and time spent looking at student work after teaching. In other words, Factor 2 addresses the degree of intensity with which class time is focused on the science curriculum. Factor 2 accounts for 12% of the total variance among the log items. Overall, the alpha for the general implementation factor was 0.81.

Table 7

Teacher Log: Factors Component Matrix

	Component	
Factor	1	2
Number of times science taught/week	.623	.444
Minutes/day > 40 on science instruction	.367	.738
Minutes/day > 5 on analysis of student work	.678	.307
Provided written feedback on individual student work (notebooks or other) to most students	.833	.100
Used a scoring guide to analyze student work	.783	-.281
Figured out a next instructional step based on student assessment data	.806	-.068
Recorded observations of students during class	.880	-.097
Checked student understandings at the end of an investigation	.780	-.192
Conducted student self-assessment sessions	.853	-.386

Note. Extraction method: Principal component analysis.

Classroom observations. Classroom observations were conducted in six randomly selected Cohort 1 schools (twelve total observations: eight treatment teachers; four control teachers) representing a range of school and student demographics, as well as a range of teacher knowledge and experience with the science curriculum and teaching. Observation data were collected to provide a context in which to understand fidelity of implementation.

We used the observation data to create two quantitative variables to provide another perspective on implementation and to validate log findings. The first variable focused on the *extent to which* teachers implemented curriculum assessment guidelines and the second characterized the *quality* of that implementation.

The first, which we term *frequency of assessment use*, is a summary of whether each assessment component (see Table 1) was in evidence during the observation or follow-up interviews, for example, evidence that a teacher analyzed work in student notebooks, analyzed work on student response sheets, recorded observations of students during class, provided feedback, or engaged students in self-assessment. Teachers received a score of "1" for evidence of implementation of the assessment component, and a score of "0" if there was no evidence of the teacher using the assessment component. Nine components were used in the analysis, with a possible score range of 0-9 points.

Additionally, a four-point coding scheme was used to rate the *quality of assessment use* with which each assessment component was used (Table 8). The maximum possible score for quality of assessment implementation was 27 points (9 assessment components x "3," the maximum score for each component). Codes were as follows:

Table 8

Classroom Observation Codes: Quality of Assessment Use

Code	Description of assessment use
3	Use and analysis of assessment component is detailed and specific.
2	Use and analysis of assessment component is general.
1	Use and analysis of assessment component is broad and unspecified.
0	No use or analysis of assessment component.

The decision to include *quality* of assessment use (in addition to frequency of use) is important to our model of fidelity of implementation because we believed that it is not only the use of an assessment, but also how well that assessment is used that impacts student learning and achievement.

Interviews. A 50% sample of Cohort 1 teachers was randomly selected to participate in phone interviews. Interview questions were designed to parallel the fidelity of implementation constructs as well as classroom observation components, and to be carried out in a 30-minute timeframe. Interviews were designed to provide data on fidelity of implementation, including teachers' thinking and reasoning behind specific instructional and assessment decisions, and in their analysis and interpretation of student work. These data provided a way for us to triangulate information from other sources of data.

Similar to the observation coding, eight curriculum assessment components implemented by both treatment and control teachers (see Table 1) were coded, as well as treatment-specific assessment components teachers described in interviews. Interviews were first coded according to a "yes/no" scale for use of specific assessments and strategies specified in the curriculum full implementation model. As in the classroom observations, we describe this as *frequency of assessment use*. Next, interviews were coded for the *quality of assessment use* of each component as described by the teacher during the interview. These codes were devised by reading the teacher interviews, collecting evidence of the teacher comments, and then coding the interviews. Assessment quality ratings ranged from "0," meaning the assessment or assessment strategy was not used by the teacher, to "3," signifying that the teacher used the assessment component, and provided detailed and specific information about how the tool was used. See Table 9 for details on the quality of assessment use ratings for interviews.

Table 9

Interview Codes: Quality of Assessment Use

Code	Description and example
3	Use and analysis of assessment component is detailed and specific. E.g., "I recorded observations of students during the investigations, and used these data to help me figure out which students understood the different structures of the crayfish, and the function of each part to provide additional learning experiences for specific students."
2	Use and analysis of assessment component is general. E.g., "I recorded observations of students, and used them to help regroup students."
1	Use and analysis of assessment component is broad and unspecified. E.g., "I made some observations of students but didn't record them in a formal way – kept track in my head."
0	No use or analysis of assessment component. E.g., "No, I didn't make formal observations of students in this module."

Similar to our thinking about the frequency of assessment use during the observations, we also coded the interviews for frequency of assessment use, as well as the quality of assessment use. By aligning the observation and interview components with the teacher logs, and with the components of the full implementation model, we had a variety of methods to capture how and in what ways teachers were using the curriculum and assessments.

Relationship between Indicators

We conducted correlation analyses to explore the relationships among the teacher log items, interview items, and observation items. The analyses drew on total scores for the frequency and quality of use variables from the observations and interviews, and scores for each of the factors identified in the logs. In addition, we included aggregated items from the logs characterizing theoretically important aspects of assessment use (i.e., use of feedback and time teachers spent analyzing student work). Because observations and interviews were conducted during different modules, we correlated scores with logs for the relevant units. That is, all observations were conducted during the Magnetism and Electricity Module and thus we correlated observation scores with log scores from that module.

Table 10 shows the significant correlation coefficients found between classroom observations and log scores for the Magnetism and Electricity Module. Results show moderately strong relationships in three areas: 1) Factor 1 log scores and the overall *quality* of assessment use during observations; 2) Factor 1 log scores and the *frequency* of teachers' use of feedback during observations, and 3) Factor 1 log scores and the *quality* of teachers' use of feedback during observations. Similar correlations were found for specific assessment components from the logs. Specifically, there were moderate correlations between the amount of time teachers spent daily outside of class assessing students' work, the frequency of teacher analysis of student notebooks, and teachers' use of feedback reported in the logs with observation scores for quality of assessment use, with the frequency of teachers' feedback, and with the quality of teacher's feedback.

Table 10

Correlations between Observations and Log Variables (Magnetism and Electricity Module)

Item	Factor 1	Time on analysis of student work (Q1D_AVE)	Frequency of analysis of student notebooks (q3b_ave)	Use of feedback (q3g_ave)
Overall quality of assessment use during observations	0.75*	0.41	0.62*	0.78*
Frequency of teachers' use of feedback during observations	0.82*	0.64*	0.57	0.65*
Quality of teachers' use of feedback during observations	0.71*	0.38	0.62*	0.73*

*Statistically significant at alpha <0.05 level.

In contrast, correlations between interview and log indices for the second module, which generally was Structures of Life, were not statistically significant. The only exception was the relationship between log data - time spent on analysis of student work (Q1D_AVE) and the interview total score on frequency of assessment use (r=.49).

We also examined correlations among scores summarized over all modules, which include observations of Magnetism and Electricity, interviews associated with Structures of Life (and one teacher who implemented the Water module), and all log responses. Results shown in Table 12 generally show an absence of relationship between log variables and quality of assessment use, as measured by observations and interviews, and moderate relationships between the primary factor emerging from the logs, Factor 1, and interview and observation ratings of frequency of assessment use. Selected items from the log (i.e., minutes teachers spend a day analyzing student work) show similar relationships to the observation and interview assessment use ratings. Note that Factor 2 and Factor 3 scores from the logs show no relationship with interview or observation scores.

Table 12

Statistically Significant (<.05) Correlations Based on Teacher Log Factor Scores, Interviews, and Observations

Teacher logs	Interviews and observations	Correlation coefficient
Q1D_AVE (minutes on analysis of student's work)	Frequency of assessment use, interview total score	0.50
Q1D_AVE (minutes on analysis of student's work)	Frequency of assessment use, observation total score	0.59
Q1D_AVE (minutes on analysis of student's work)	Frequency of assessment use, observation-specific items	0.68
Q3G_ave Provided feedback to individual students (days/week)	Quality of assessment use, interview total score	0.49
Factor 1 (assessment factor)	Frequency of assessment use, interview total scores	0.55

Correlations between teachers' interview scores and classroom observation scores are generally high, despite the different module contexts for each. Table 13 presents the details. Note in particular the high correlations between the quality of assessment use ratings from each instrument.

Table 13

Statistically Significant (<.05) Interview and Observation Correlations

Observations	Interview total, frequency of assessment use	Interview total, quality of assessment use	Interview total, treatment-only items, quality of assessment use
Observation total, frequency of assessment use	0.93	0.86	0.75
Observation total, quality of assessment use	0.86	0.93	0.82
Observation total for treatment-only items, quality of assessment use	0.70	0.87	0.87
Total score for observation-specific items, frequency of assessment use	0.82	0.84	0.94
Total score for observation- specific items, quality of assessment use	0.81	0.94	0.92

Finally, we explored the correlation between the teacher content assessment and the other measures; note that the correlations between measures of the same content area are high. The correlation matrix for the open-ended items on the teacher content assessment and other measures can be found in Appendix A, with significant correlations highlighted.

Taken together, the measures developed and used to help answer research question 2 were able to capture curriculum implementation and teacher practices. Because the correlations are moderately high, we are fairly confident that what we are seeing is a relatively accurate picture of what teachers did and didn’t do. By moving away from reliance on a single indicator of implementation, we are able to better understand frequency of assessment use as well as the quality of that assessment use.

Teacher logs, in concert with selected observations and interviews, provided evidence of the accuracy, reliability, and validity of the instruments for the study. The relationship between the teacher content assessment and use of the curriculum plays an important role in understanding the factors at work in fidelity of implementation. Each instrument captures unique elements of implementation, teacher knowledge, and assessment strategies. The use of multiple data sources allows for the triangulation of data to provide a more complete picture of curriculum implementation, thus laying the foundation for understanding curriculum efficacy.

Figure 2 presents our model for fidelity of implementation developed for the efficacy study.

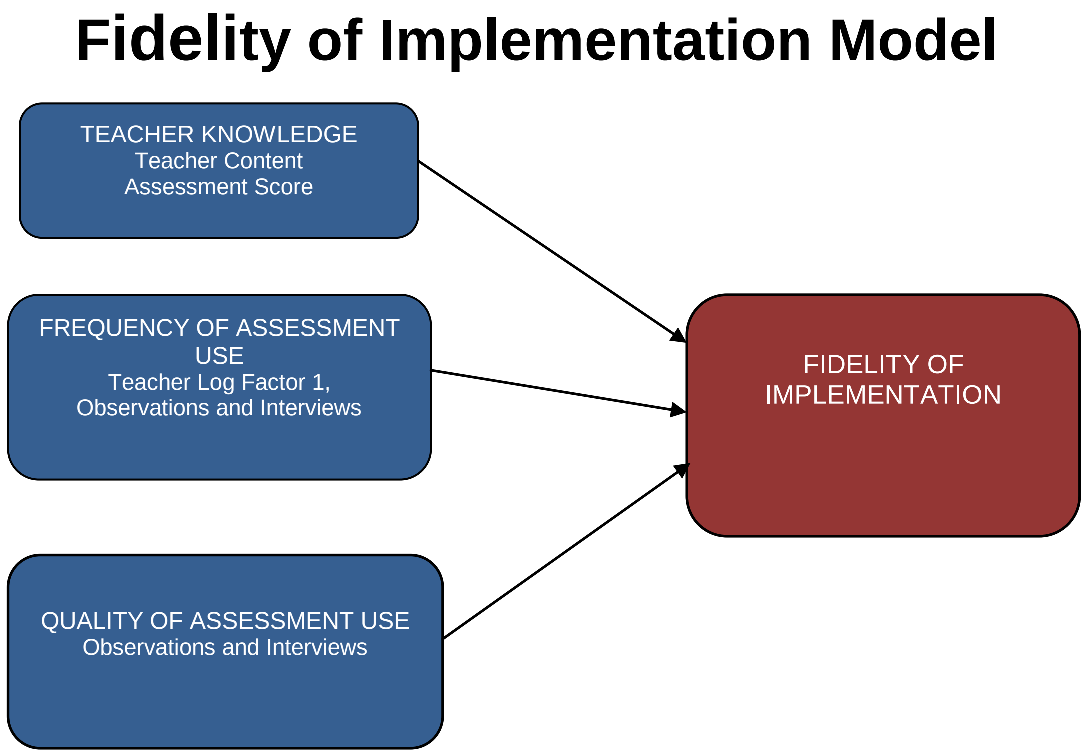


Figure 2. Fidelity of implementation model.

No single measure adequately captured all of the elements of fidelity of implementation, leading us to believe that the measures, while related, capture unique contributions to fidelity of implementation.

Research Question 3: What Practical Challenges are Associated with Developing Effective Measures of Fidelity?

As with all curriculum programs and research efforts**,** our study existed within a complex set of goals, tensions, and interactions between different stakeholders. We discuss these interactions and highlight the challenges in understanding fidelity of curriculum implementation based on various needs and perspectives from several key participant groups.

Teachers. Teachers volunteered for the study primarily because they were interested in the specific curriculum. Their focus was on using the curriculum, and supporting student learning. Teachers were not necessarily focused on recording the details of what they did, nor when or how they did these specific activities. Teachers were similarly less focused on the extent to which they understood the subtleties of specific assessment components, or in carefully reflecting on their own understandings of assessing student learning. Teachers face significant challenges in reporting data, in part because detailed documentation falls outside the realm of typical teacher activities, and because science time is often shortened or moved to another time slot during the week. The flexing science schedule challenges even the most organized teachers to keep their lesson plans current, and represent what they actually did, rather than what they had planned to do, when reporting their work in the logs.

Another issue for teachers with respect to the logs relates to the computer interface of the logs themselves. To increase accuracy in reporting implementation data, researchers designed the logs to close automatically 10 days after completing a week of instruction. This timeframe was wide enough to allow teachers to complete the logs, but short enough to minimize teachers' forgetting what they had accomplished. Additionally, teachers had to complete the log in a single sitting, meaning they couldn't start a log and complete it at a later time. This feature was designed intentionally by researchers to help avoid the "re-creating history" tendency. Several factors came into play that may have limited, unintentionally, the accuracy and completeness of the logs. While teachers intended to complete logs during the instructional day, or at the end of the day, there were often interruptions that meant teachers occasionally submitted incomplete logs, or logs that they did not have the opportunity to review for accuracy. Some teachers reported that they would have welcomed the opportunity to review the log completely before submitting it, to clarify a point, add more detail, or check their logs against their lesson plans.

Curriculum and professional developers. Curriculum developers were interested in understanding whether the new curriculum was more effective in supporting student learning than the traditional curriculum. But the specific details about what teachers should be doing, for how long, and when – conditions that determine the fidelity of the implementation – may have fallen outside the scope or focus of the intentions of curriculum developers. Moreover, details about the level of professional development and the support teachers would need to implement the curriculum were somewhat unspecified or were not necessarily compatible with classroom demands.

Professional developers attempted to develop a common understanding of the various components of the curriculum, so that when teachers reported on implementing specific components of the curriculum (e.g., embedded assessments or feedback techniques), teachers reported on the same component, construct, or activity. Despite the potential usefulness of measuring whether teachers reached this common understanding, collecting data on this variable was outside the scope of the research.

Researchers. Finally, in this efficacy study, researchers were interested in understanding who was doing what, how well they were doing it, and how well the curriculum (and embedded assessments) worked and why. Researchers were concerned with collecting adequate data to get reliable information, including enough items to create reasonable scales for measuring impact. Yet this need/quest for implementation data may have become burdensome and/or uninteresting to teachers, who were working hard to learn new components of a curriculum, and, for treatment teachers, to implement the curriculum in ways that may have been unfamiliar and challenging.

Importance of a "Practice" Year

This study involved the implementation of a complex curriculum with multiple components, and challenging content. Teachers, both control and treatment, benefitted by having a "practice" year that allowed both groups of teachers time for reflection and to learn to use the curriculum according to the study specifications. Treatment teachers learned to use a new embedded assessment system, one that required them to engage in practices and strategies outside their typical comfort zone. Teachers also completed logs during the practice year, which allowed them time to become familiarized with the computerized logging system and to develop a routine whereby they would enter the requested implementation data.

Log Development and Refinement

Development of logs directly tied to the essential components of the curriculum and its implementation was a critical feature of this efficacy study. In particular, it was important to develop teacher understanding of the importance of recording accurate information in the logs, and to help teachers clearly identify what teaching and assessing activities corresponded to log items. The issue of the amount of time logs should remain open to teachers to complete after instruction, and the extent to which teachers have the option to review and refine their log entries, is worth additional consideration when deciding how to best measure fidelity of implementation.

Of additional importance is understanding log completion variability. Log completion raises questions about incomplete data sets: how accurately do the log data reported represent teachers' typical practice? What are effective methods for estimating the average or typical profile for each teacher during a week of science instruction? These are important considerations when examining log data, and estimating fidelity of implementation, and that we will need to address with the complete teacher log sample

Conclusions

In this report, we presented an approach to addressing challenges inherent in efficacy studies, that is, how to determine levels of fidelity of implementation. We developed theory-based measures of fidelity of implementation, validated these measures by examining curriculum implementation, and examined the relationships between the measures. Data generated by these measures helped us to examine complex interactions in the study, and implementation factors that may influence student outcomes.

A key lesson, in our opinion, is that embedded assessment—the “value added” curriculum component in this study—is a complex, iterative process that involves more than simply implementing a curriculum and making use of accompanying assessment tools. The data generated by the tools in this study provided critical information about the extent to which teachers implemented this “value added” component. Going beyond measurement of whether teachers “did it/didn’t do it” is critical to understanding the impact of specific approaches, materials, and curriculum on students’ science learning.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles, Policy and Practice. 5*(1), 7 -74.

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability* (pp. 20-50). Chicago, IL: University of Chicago Press.

Black, P., & Wiliam, D. (2009) Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5-31.

Heller, J. I., Daehler, K. R., Shinohara, M., & Kaskowitz, S. R. (2004, April). *Fostering pedagogical content knowledge about electric circuits through case-based professional development.* Paper presented at the annual meeting of the National Association for Research in Science Teaching (NARST), Vancouver, Canada.

Heritage, M., & Vendlinski, T. (2006). *Measuring teachers' mathematical knowledge*. (CSE Tech. Rep. 696). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Herman, J., Osmundson, E., & Silver, D (2010). *Capturing quality in formative assessment practice: Measurement challenges.* (CRESST Report 770). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (2004). *TIMSS 2003 technical report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

No Child Left Behind Act of 2001, Pub. L No. 107-110, 115 Stat. 1425 (2002).

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (2008). *TIMSS 2007 technical report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Osmundson, E., Dai, Y., & Herman, J. (2011). *Year 3 ASK/FOSS efficacy study*. (CRESST Report 782). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Appendix A:
Magnetism and Electricity Module: Pre/post Assessment

Table A1

Inter-rater reliability, open-ended items

Kappa	Pretest	Posttest
MagElec1	0.92	0.89
MagElec2	0.79	0.74
MagElec3	0.71	0.82
MagElec4	0.91	0.89
MagElec5	0.90	0.77
MagElec6	0.94	0.59
MagElec7	0.81	0.96
MagElec8	0.65	0.39
MagElec9	0.81	0.56
MagElec10	0.90	0.78
MagElec11	0.79	0.62
MagElec12	0.88	0.82
MagElec13	0.83	0.78
MagElec14	0.52	0.61
MagElec15	0.96	0.97
MagElec16	0.94	0.88
MagElec17	0.99	0.91
Mean kappa	0.84	0.76

Overall, the inter-rater reliability indicates that the raters scored students' responses with high level of agreement (Landis & Koch, 1977).

Table A2

HLM Analyses 1: Post scores controlling for pretest scores

Effect	Estimate	*SE*	*df*	*t* value	Pr > |t|
Intercept	28.84	0.74	28.00	38.87	<.0001
gp_AZC1_me_pre	0.40	0.07	618.00	5.42	<.0001
gpgd_AZC1_me_pre	0.23	0.11	618.00	2.08	0.04
Treatment vs. control	3.23	1.02	28.00	3.18	0.00
gp_AZC1_me_pre*tc	-0.19	0.10	618.00	-1.87	0.06

Note. HLM1 outcome variable: AZ Cohort 1 Magnetism and electricity posttest total.

There is a statistically significant and noticeable treatment effect on students' posttest scores when pretest scores are controlled as displayed in Table A2 (treatment vs. control).

Table A3

HLM Analyses 2: Post scores controlling for background variables and pre-test scores

Effect	Estimate	*SE*	*df*	*t* value	Pr > |t|
Intercept	28.78	0.95	27.00	30.45	<.0001
gp_AZC1_me_pre	0.38	0.07	568.00	5.15	<.0001
gpgd_AZC1_me_pre	0.18	0.10	27.00	1.78	0.09
Treatment vs. control	3.20	0.92	27.00	3.47	0.00
gp_AZC1_me_pre*tc	-0.21	0.10	568.00	-2.05	0.04
Third_Ethnicity_H	0.01	0.75	568.00	0.02	0.99
Third_Ethnicity_W	1.36	0.70	568.00	1.95	0.05
Third_ELL_di	-1.28	0.70	568.00	-1.82	0.07
Third_FRL_di	-0.36	0.48	568.00	-0.75	0.45
Third_Gender_di	-0.22	0.38	568.00	-0.58	0.56

Note. HLM2 outcome variable: AZ Cohort 1 Magnetism and electricity posttest total.

After taking into account differences in student backgrounds (e.g., ethnicity, gender, language proficiency status and free/reduced lunch eligibility), results indicate a statistically significant treatment effect on students' posttest scores while controlling for their pretest scores as displayed in Table A2 (treatment vs. control)

Appendix B:
Teacher Content Survey

TEACHER CONTENT SURVEY

FOSS STUDY

PURPOSE OF THIS INSTRUMENT

This measure is designed to collect information about teacher understandings of magnetism and electricity and approaches teachers use to understand student thinking. Results from the survey will help us to better understand how FOSS works to help students learn science.

INSTRUCTIONS

1. You have been alloted 30–45 minutes to complete this measure. However, if you wish, you may use more time during your break in order to finish it. You may choose to not answer questions and/or stop your work at any point during the time period.

The content survey includes questions with a wide range of difficulty, and we expect you to encounter items for which you may not know the answers. If you are not sure of an answer, ***please make your best guess***—there is no penalty for guessing.

2. Please fill in your name and ID numbers below and your ID on the next page.

First name Last name Date

Your ID Number: T ☐ ☐–☐ ☐–☐ ☐ ☐

IMPORTANT:

To keep your data confidential, this cover sheet with your name will be removed upon receipt by the research staff, leaving only your ID number on the next page of the survey. This cover sheet will be stored in a locked cabinet, separate from the completed surveys.

Your ID Number: T ☐ ☐–☐ ☐–☐ ☐ ☐

1.11 Julie placed a paper clip, piece of cardboard, and magnet together like you see in the picture.

Why did the paper clip stay in place next to the cardboard instead of falling to the floor? Choose the **best** answer.

❍ A. The paper clip is made of iron and so is the magnet.

❍ B. The magnetic field goes around the cardboard and makes the paper clip stay there.

❍ C. The magnet has a magnetic field that is not blocked by the cardboard.

❍ D. The electric force field makes the paper clip attract to the magnet.

1.12 Arthur was playing with magnets. He had one magnet on the table, and one in his hand. As he moved the magnet in his hand closer to the one on the table, the magnets suddenly snapped together.

a. Explain why the magnets snapped together even though they were not touching.

Here are two students' responses to question 1.12:

Student 1 Response: Both magnets are made of iron, and the magnets are both facing south and south.

Student 2 Response: The magnets snapped together because the electric fields got close.

b. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

c. If these students were in your class, what would you do next in your instruction to help the students learning progress?

1.21 A nail that was stuck to a permanent magnet picked up a small metal washer. The nail could pick up the metal piece because:

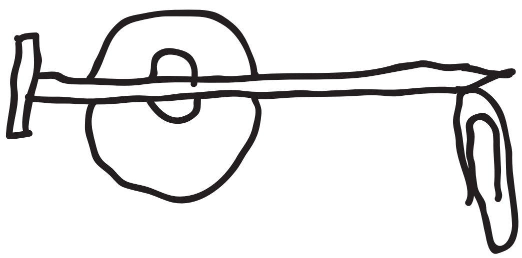
❍ A. Nails have magnetic fields.

❍ B. Magnetism was induced in the nail.

❍ C. The nail and the washer are both made of iron.

❍ D. The washer is still in the range of the magnetism.

1.22 Anne is investigating objects and magnets. She made this observation in her science journal.



"I was surprised! A nail was stuck to the magnet. When I accidentally touched the nail to a paper clip, the paper clip stuck to the nail. I wonder why that happened?"

a. Explain to Anne why the paper clip stuck to the nail. Use diagrams or pictures if necessary.

Anne and her friend were asked by her teacher why they thought the paper clip stuck to the nail. Here are their responses to the question:

Anne's response: The paper clip turned into a magnet too.

Anne's friend's response: The nail gets stuck on the magnet, and the nail turns into a magnet, so the paper clip can stick on the nail.

b. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

c. If these students were in your class, what would you do next in your instruction to help the students learning progress?

1.31 a. Complete the following table. Put an "X" in the second column of the table if the object sticks to a magnet. Put an "X" in the third column of the table if the object conducts electricity.

Object	Sticks to a magnet	Conducts electricity
Iron nail		
Plastic straw		
Steel wire screen		
Wooden craft stick		
Brass ring		
Rubber band		
Copper penny		
Piece of aluminum foil		

b. Why did you choose the objects that you did in the "Sticks to a magnet" column? Use diagrams or pictures to show your thinking.

c. Why did you choose the objects that you did in the "Conducts electricity" column? Use diagrams or pictures to show your thinking.

1.32 Here is how one student completed the table.

Object	Sticks to a magnet	Conducts electricity
Iron nail	X	X
Plastic straw		
Steel wire screen	X	X
Wooden craft stick		
Brass ring	X	X
Rubber band		
Copper penny	X	X
Piece of aluminum foil	X	X

Here are one student's responses to questions 1.31b and 1.31c (see page 5):

Student 1 Response:

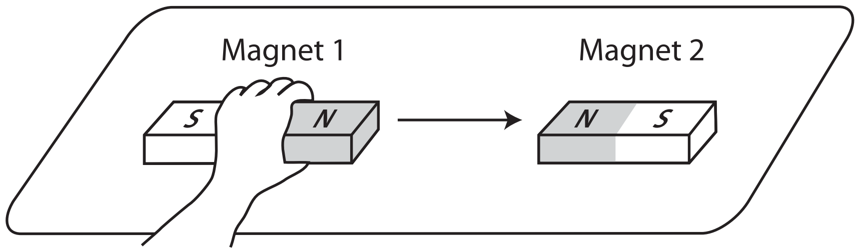
1.31 b. These things stick to the magnet because they are all metal.

1.31 c. These things are all made of metal and metal conducts electricity.

a. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

b. If these students were in your class, what would you do next in your instruction to help the students learning progress?

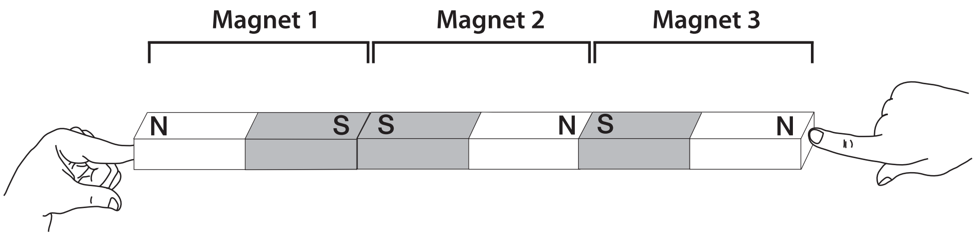
1.41 The picture below shows Maria pushing Magnet 1 toward Magnet 2 on a smooth table. Both magnets are lying on a smooth table.



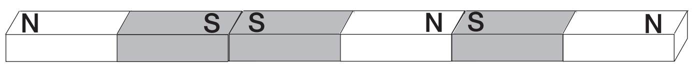
a. What will happen as Magnet 1 moves towards Magnet 2?

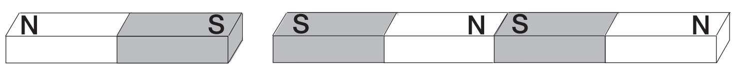
b. Why will this happen?

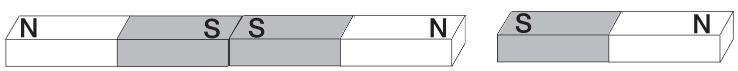
1.42 Three bar magnets are held together as shown in the picture below.

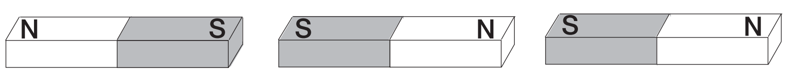


a. What will the magnets do when they are released? Circle the correct answer.

A. 

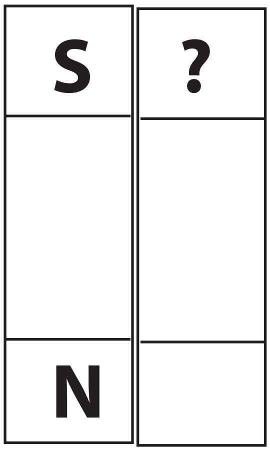
B. 

C. 

D. 

b. Why does that happen?

1.46 Lisa found a magnet with no labels on the poles. She found another magnet with correctly labeled poles and put the magnets together. They **attracted**.



a. The pole labeled with the "?" is most likely which pole?

❍ A. south pole
❍ B. north pole
❍ C. not enough information provided

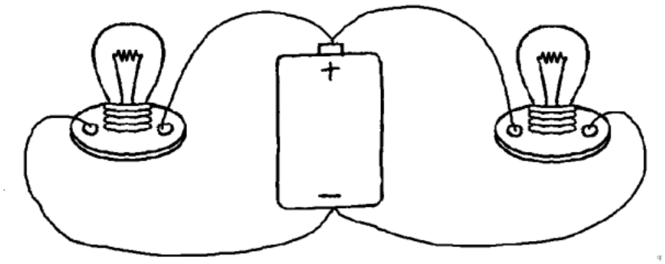
b. Why? Please explain your answer.

WRAPPING IT UP

1. What is/are the key concept/s addressed by the assessments in Section 1?

2. Why is it important for students to learn these magnetism and electricity concepts?

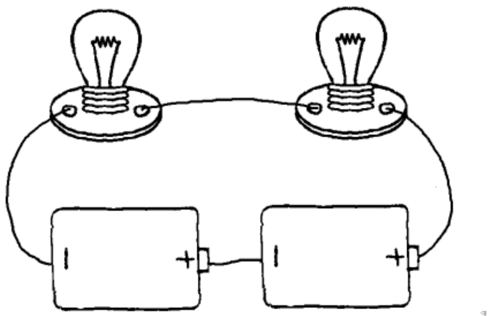
2.11 Look at the picture below. What kind of circuit is this?



How do you know?

❍ A. network circuit
❍ B. series circuit
❍ C. parallel circuit
❍ D. short circuit

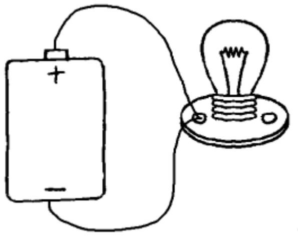
2.12 Look at the picture below. What kind of circuit is this?



How do you know?

❍ A. simple circuit
❍ B. series circuit
❍ C. parallel circuit
❍ D. short circuit

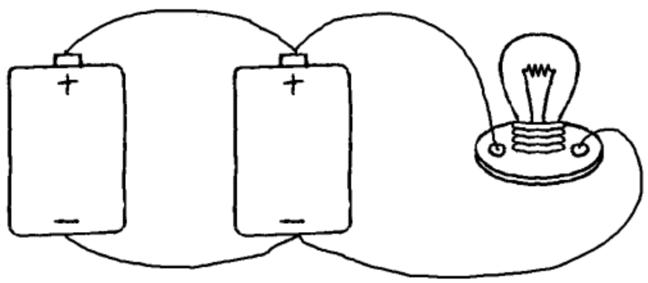
2.13 Look at the picture below. What kind of circuit is this?



How do you know?

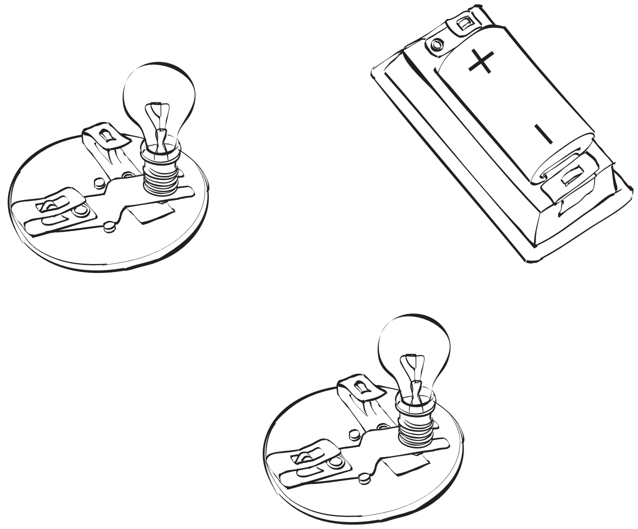
❍ A. simple circuit
❍ B. series circuit
❍ C. parallel circuit
❍ D. short circuit

2.14 Look at the picture below. What kind of circuit is this?



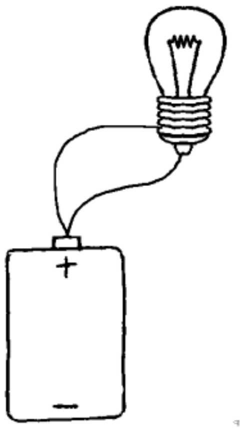
❍ A. simple circuit
❍ B. network circuit
❍ C. series circuit
❍ D. parallel circuit

2.15 a. Draw in lines representing wires to make a ***parallel*** circuit.



Explain your drawing: what features make this a parallel circuit?

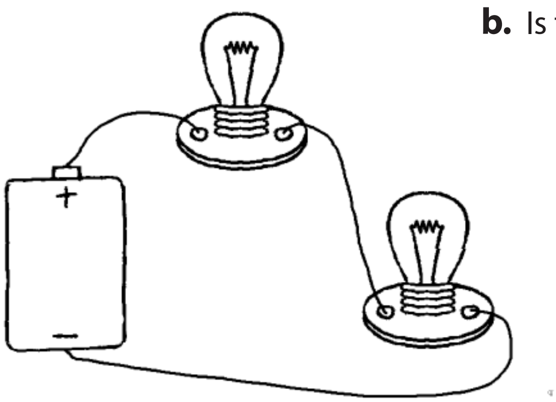
2.21 Look at the picture below.



a. Will the bulb light? ❍ Yes ❍ No

b. Is the circuit complete? ❍ Yes ❍ No

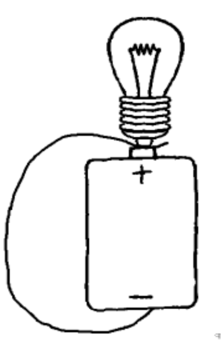
2.22 Look at the picture below.



a. Will the bulbs light? ❍ Yes ❍ No

b. Is the circuit complete? ❍ Yes ❍ No

2.23 Look at the picture below.



a. Will the bulb light? ❍ Yes ❍ No

b. Is the circuit complete? ❍ Yes ❍ No

c. Explain why you think the circuit is or is not complete.

2.24 Look at the picture below. The round object in the middle of the picture is an empty bulb holder.

a. Will the bulb light? ❍ Yes ❍ No

b. Explain why you think the bulb will or will not light.

This is how a Student 1 responded to question 2.24.

a. Will the bulb light? ❍ Yes ⊗ No

b. Bulb won't light because it's not connected to the battery.

This is how a Student 2 responded to question 2.24.

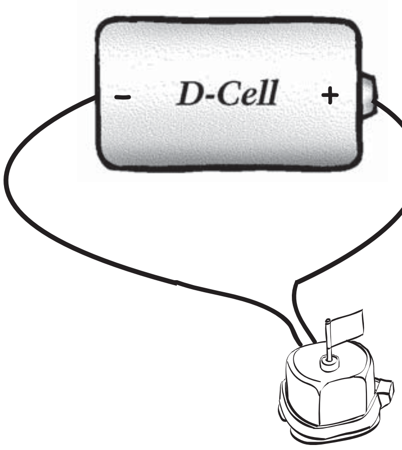
a. Will the bulb light? ❍ Yes ⊗ No

b. Bulb won't light because it's a short circuit.

c. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

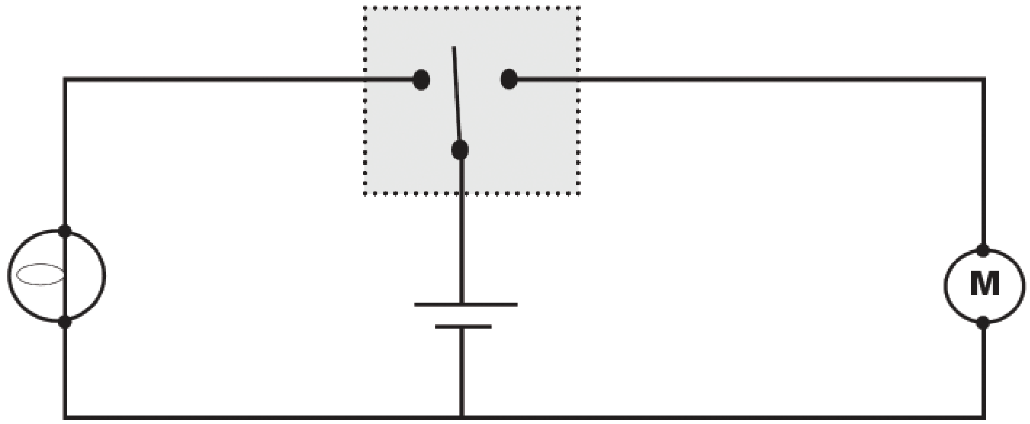
d. If these students were in your class, what would you do next in your instruction to help the students learning progress?

2.31 Draw arrows on the picture to show which direction electricity will flow through the circuit to run the motor.



Explain your answer.

2.32 Denise wants to build a circuit that will light up a bulb and run a motor at the same time. She drew the diagram of the circuit she planned to build. She used a special switch in the circuit. The switch is shown in the gray box.



a. Look at the diagram Denise drew. Explain to her why you think her circuit would or would not work the way she wants it to work.

2.33 Below are two student's responses to question 2.32.

Student 1 response: I think it would work because all the parts of connected. But it might not work because the battery might not have enough juice to carry all on one circuit.

Student 2 response: It probably won't because the energy can't go two different ways.

a. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

b. If these students were in your class, what would you do next in your instruction to help the students learning progress?

2.41 Electricity can be changed into other forms of energy. Complete the sentences below:

a. The bulb in a lamp changes electric energy into ______________________________

____

b. A motor changes electric energy into ______________________________

____

2.42 Which of the following items converts electric energy into motion?

❍ A. light switch
❍ B. electric stove
❍ C. light bulb
❍ D. electric fan

2.43 When an electric stove is turned on, most of the incoming electrical energy changes into:

❍ A. heat energy
❍ B. light energy
❍ C. mechanical energy
❍ D. sound energy

2.44 Which of the following items converts electric energy into light?

❍ A. light switch
❍ B. doorbell
❍ C. light bulb
❍ D. electric fan

2.45 When an electric fan is running, most of the incoming electric energy is converted into:

❍ A. heat energy
❍ B. light energy
❍ C. motion energy
❍ D. sound energy

2.46 Household appliances convert electricity into one or more different forms of energy. An electric fan can best be described as converting electricity into:

❍ A. heat energy only
❍ B. heat energy, and sound energy
❍ C. heat energy, sound energy and motion energy
❍ D. heat energy, sound energy, motion energy and chemical energy

WRAPPING IT UP

1. What is/are the key concept/s addressed by the questions in Section 2?

2. Why do students need to know these concepts about magnetism and electricity?

3.11 Annie had three rivets. One was copper, one was iron and one was steel. Which rivet or rivets could she use to make an electromagnet? Why?

3.12 Here are two students' response to question 3.11

Student 1: Annie should use the iron and steel rivets because they conduct electricity and they stick to magnets.

Student 2: Annie could use the iron, copper or steel rivets because they are all metal.

a. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

b. If these students were in your class, what would you do next in your instruction to help the students learning progress?

3.21 Imagine you have the following materials: a large iron nail, several permanent magnets, lots of insulated wire, a D-cell and a switch.

a. Describe one way to make the nail a temporary magnet.

b. Describe another way to make a temporary magnet.

3.31 Samuel Morse, the inventor of the telegraph, had a problem. His telegraph's signal was too weak. He needed a stronger electromagnet. What are two ways he might have used to increase the strength of the electromagnet for his telegraph?

3.41 Wendy is making an electromagnet. First, she wrapped a long, insulated wire around an iron nail. What should Wendy do to complete the electromagnet?

Here are two student responses to question 3.41:

Student 1: Attach the wire to the D-cell and switch, rub the magnet on the nail a few times and then try it.

Student 2: Wendy should connect the iron nail to the D-cell to make a complete circuit.

a. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

b. If these students were in your class, what would you do next in your instruction to help the students learning progress?

3.42 Which of the following materials is **NOT** necessary to build an electromagnet?

❍ A. a magnet

❍ B. a steel rivet

❍ C. a D-cell battery

❍ D. wire

WRAPPING IT UP

1. What is/are the key concept/s addressed by the assessments in Section 3?

2. Why do students need to know these concepts about magnetism and electricity?

Appendix C:
ASK/FOSS Classroom Assessment Observation Protocol

ASK/FOSS Classroom Assessment Observation Protocol

Observation Notes

	Descriptions
Observer	
Date	
School	
Teacher	
Module & Investigation	

Concepts Addressed	
Lesson	
Introduction: Focus question	
Activity 1 Describe the activity, what teacher is doing, what students are doing, interactions.	
Activity 1 Assessment To what extent is T involved with assessing Ss? (1=not at all, 3=moderate extent, 5=great extent)	
Activity 2 Describe the activity, what teacher is doing, what students are doing, interactions.	
Activity 2 Assessment To what extent is T involved with assessing Ss? (1=not at all, 3=moderate extent, 5=great extent)	
Activity 3 Describe the activity, what teacher is doing, what students are doing, interactions.	
Activity 3 Assessment To what extent is T involved with	

assessing Ss**?** (1=not at all, 3=moderate extent, 5=great extent)	
Other observational data (fill out as observing)	
Classroom description	
Assessment materials in evidence (per activity/task if appropriate**)**	
Other: please indicate	

Note: observer should take notes during the observation and complete the scaled items at the conclusion of the lesson and/or after reviewing notes.

	Check if observed	Description	Congruence/Alignment with FOSS/ASK assessment system 1=not at all 3=moderate 5=to a great extent see below
Prior to the lesson/investigation			
Used the "At a Glance" to review science content and assessment opportunities for teaching and assessment			
During the lesson/investigation			
Analysis and Interpretation			
Analyzed students' science notebooks			
Used a scoring guide to analyze response sheets			
Recorded observations of students' during class			
Analyzed **student work** for patterns and trends			
Analyzed **observations** for patterns and trends			
Feedback to Students			

	Check if observed	Description	Congruence/Alignment with FOSS/ASK assessment system 1=not at all 3=moderate 5=to a great extent see below
Individual Individual students provided ongoing, clear feedback regarding progress toward targeted goals.			
Small Group Targeted, specific, descriptive feedback is provided to students working in small groups regarding progress towards targeted goals.			
Whole Class Targeted, specific, descriptive feedback is provided to whole class regarding progress towards targeted goals.			
Notebooks Provided feedback to students on notebook entries			
Provided opportunities for students to work in small groups to discuss ideas			
Asked open-ended questions			
Guide for Instruction			
Planned and implemented additional instruction based on **observations of students** during class			
Planned and implemented additional instruction based on **assessment results**			

	Check if observed	Description	Congruence/Alignment with FOSS/ASK assessment system 1=not at all 3=moderate 5=to a great extent see below
End of lesson/investigation			
Checked on students' understandings of science concepts			
Engaged students in self-assessment of science learning			
Other: please specify			

Appendix D:
ASK Study Research-Phone Interview

ASK Study Research: Phone Interview

[Purpose: provide more detailed and specific information on teacher assessment practices, based on guidelines from LHS on "full implementation" model, and certain components of CRESST's Quality Assessment Model]

Note: it may be helpful to provide the teacher with a copy of the interview protocol to help him/her follow the questions and the conversation.

Introduction

(Interviewer introduces self)

Hi. As you know, we're conducting interviews with teachers in the ASK/FOSS study to help us better understand your use of FOSS and how are you are assessing students. This interview to bring me up to date on the _______ (Module and Investigation).

Do I have your permission to audiotape this conversation? I will use the tape only to ensure I have complete notes. As we outlined in our information letter and permission documents, your confidentiality is assured, and you have the right not to answer any questions and to terminate the conversation at any time.

Do you have any questions before we begin?

Interview Questions

1. **General Update**: I want to get a general sense of the ______ (Module and Investigation) you've been teaching. Note: Section 1 should take 2 – 3 minutes.
   a. Based on the information you've provided in the Teacher Logs, I see that you have just finished ______ Module and Investigation (interviewer needs to check Teacher Logs in advance of interview). How are things going – what has worked well so far with this Module and Investigation? What has been a challenge? (keep very brief)
   b. Which assessments have you used to date (check all that apply: pre-test, I-checks, student response sheets, notebooks). In general, how are things going with assessing students' learning – what has worked well so far? What has been a challenge? (keep very brief)

2. Use of Assessments

Now I'd like to ask you more specific questions about your ASK assessment practices.

Note: interviewer will take brief notes here to describe the process. Audiotape can be used to supplement the details, but does not have to be transcribed verbatim.

In this current Investigation, have you:	**Yes/No**	**If yes, then:** **a. *how* did you use the tool/do it?** **b. what did you find out about student learning from this process or work?** **c. what do the results mean for your teaching?**
a. Analyzed work in students' science notebooks		
b. Analyzed student work on the response sheets		
c. Recorded observations of students' during class		
d. Analyzed **student work** for patterns and trends		
e. Analyzed **observations** for patterns and trends		
f. Planned and used a next-step strategy based on student work		
g. Provided feedback to students about their work and learning		

3. **<u>End of Investigation</u>**

Note: these questions apply (need to be asked) only if the teacher indicates that s/he is at the conclusion of an Investigation.

Interviewer: Next, I'd like to ask you about the end of investigation assessments. (refer back to information in #1 to guide next set of questions).

In this current Investigation, have you:	Yes/No	If yes, a. *how* did you use the tool/do it? b. what did you find out about student learning from this work? c. what do the results mean for your teaching?
a. Administered the I-Check Benchmark Assessment		
b. Used coding guides in the Benchmark Folio to analyze I-Check		
c. Recorded I-Check data on the Benchmark Coding sheets		
d. Conducted student self-assessment session after I-Checks were returned to students		
e. Checked student reflections (revisions) after self-assessment session		
f. Made instructional decisions based on I-Check results		
g. Other: please specify		

4. Study Groups

In this current Investigation, have you:	Yes/No	If yes, a. describe what you did Note: see other specific questions below
a. met as a Study Group		
b. scored work in your Study Group		**a. describe what you did** **b. what did you find out about student learning from this work?** **c. what do the results mean for your teaching?**
c. figured out next steps strategies based on the combined student work		**a. describe what you did**
d. planned next instructional steps		**a. describe what you did**
e. other: please describe		**a. describe what you did**

5. Wrap Up

Do you have any other questions or comments to add? Thanks very much for your time.

Appendix E:

FOSS Study Research-Phone Interview

FOSS Study Research: Phone Interview

[Purpose: provide more detailed and specific information on teacher assessment practices, based on guidelines from LHS on "full implementation" model, and certain components of CRESST's Quality Assessment Model]

Introduction

(Interviewer introduces self)

Hi. As you know, we're conducting interviews with teachers in the ASK/FOSS study to help us better understand your use of FOSS and how are you are assessing students. This interview to bring me up to date on the _______ (Module and Investigation).

Do I have your permission to audiotape this conversation? I will use the tape only to ensure I have complete notes. As we outlined in our information letter and permission documents, your confidentiality is assured, and you have the right not to answer any questions and to terminate the conversation at any time.

Do you have any questions before we begin?

Interview Questions

1. **<u>General Update</u>**: I want to get a general sense of the ______ (Module and Investigation) you've been teaching. Note: Section 1 should take 2 – 3 minutes.
   a. Based on the information you've provided in the Teacher Logs, I see that you have just finished ______ Module and Investigation. How are things going – what has worked well so far with ? What has been a challenge? (keep very brief)
   b. Which assessments have you used to date (check all that apply: pre-test, student response sheets, notebooks, other). In general, how are things going with assessing students' learning – what has worked well so far? What has been a challenge? (keep very brief

2. <u>Use of Assessments</u>

Now I'd like to ask you more specific questions about your assessment practices when teaching FOSS.

Note: interviewer will take brief notes here to describe the process. Audiotape can be used to supplement the details, but does not have to be transcribed verbatim.

In this current Investigation, have you:	**Yes/No**	**If yes,** **a. *how* did you use the tool/do it?** **b. what did you find out about student learning from this work?** **c. what do the results mean for your teaching?**
a. Analyzed students' science notebooks (if applicable)		
b. Used a scoring guide (or coding guide) to analyze response sheets		
c. Recorded observations of student's during class (e.g., in small groups, 1:1 conversations)		
d. Analyzed **student work** for patterns and trends		
e. Analyzed **observations** for patterns and trends		
f. Planned further instruction based on patterns and trends in student work (specify which work)		
g. Provided feedback to students about their work and learning		

3. End of Investigation

Note: these questions apply (need to be asked) only if the teacher indicates that s/he is at the conclusion of an Investigation.

Interviewer: Next, I'd like to ask you about the end of investigation assessments. (refer back to information in #1 to guide next set of questions).

In this current Investigation, have you:	**Yes/No**	**If yes,** **a. *how* did you use the tool/do it?** **b. what did you find out about student learning from this work?** **c. what do the results mean for your teaching?**
a. Checked on students' understandings at the end of a lesson or an investigation (describe)		
b. Engaged students in self-assessment of science learning		
c. Other: please specify		

4. Wrap Up

Do you have any other questions or comments to add? Thanks very much for your time.

Appendix F:
Weekly Teacher Log-FOSS/ASK Water Module

Weekly Teacher Log: FOSS/ASK Water Module

Your responses to these questions will be confidential except for two items, which are clearly marked in red below.

1. Date:

2. Which Investigation(s) did you work on this week? (This information will be made available to support staff and contractors to allow them to better support you.)
(check the appropriate boxes)
Water:
Survey (pretest)
1.1
1.2
1.3
2.1
2.2
2.3
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
Posttest

3. How many days did you teach FOSS/ASK this week? (check the appropriate box)

No FOSS/ASK this week   1 day   2 days   3 days   4 days   5 days

4. On the days that you taught science, approximately how much time did you spend teaching FOSS/ASK? (check the appropriate box)

0 – 20 minutes/day   21 – 40 minutes/day   41-60 minutes/day   more than 60 minutes/day

5. This week, approximately how much time did you spend each day looking at student work after teaching FOSS/ASK? (check the appropriate box)

No time   5 minutes/day   10 minutes/day   20 minutes/day   more than 30 minutes/day

6. How many times when using FOSS/ASK this week did you provide written feedback on individual student work (notebooks or other) to most students?

0 times   1 time   2 times   3 times   4 times   5 times

7. How many times when using FOSS/ASK this week did you use a next-step strategy (feedback to the entire class at one time)?

0 times   1 time   2 times   3 times   4 times   5 times

8. How many times when using FOSS/ASK this week did you reteach content?

0 times 1 time 2 times 3 times 4 times 5 times

9. In the past week during FOSS/ASK instruction, how many times did you engage in the following items and activities?

	Used 1x	Used 2x	Used 3x	Used 4x	Used 5x	N/A
Resources						
a. Used the "At a Glance" to review focus question, science content, and assessment opportunities for your teaching	1	2	3	4	5	6
Embedded assessment for each lesson						
b. Analyzed student work in science notebooks	1	2	3	4	5	6
c. Analyzed student work on the response sheets	1	2	3	4	5	6
d. Recorded observations of students during class	1	2	3	4	5	6
e. Analyzed **student work** for patterns and trends	1	2	3	4	5	6
f. Analyzed **observations** for patterns and trends	1	2	3	4	5	6
g. Planned and used a next-step strategy based on patterns and trends	1	2	3	4	5	6
h. What did you learn about students' understanding of science concepts from your analysis of student work? Please provide examples and specific details.						
Benchmark assessments for each investigation						
i. Administered the I-Check Benchmark Assessment	1	2	3	4	5	6
j. Used coding guides in the Benchmark Folio to analyze I-Check.	1	2	3	4	5	6
k. Recorded I-Check data on the "Benchmark Coding Sheets"	1	2	3	4	5	6
l. Conducted student self-assessment session based on I-Check results	1	2	3	4	5	6
m. Checked students' reflections after self- assessment.	1	2	3	4	5	6
n. Made instructional decisions based on I-Check results	1	2	3	4	5	6
o. Other: please specify	1	2	3	4	5	6
p. What did you learn about students' understanding about science concepts based on information from the I-Checks? Please provide examples and specific details.						

Comments:

10. What percentage of your students do you think understand the core concepts of the Investigation you taught this week? (check the appropriate box)

0%-25% 25%-50% 50%-75% 75%-95% 100% Not sure

11. Do you have any questions or feedback about your experience with the project this week? This question is not confidential and responses will be made available to support staff and contractors to allow them to better support you.

Thank you!

Appendix G:
Weekly Teacher Log-FOSS Water Module

Weekly Teacher Log: FOSS Water Module

Your responses to these questions will be confidential except for two items, which are clearly marked in red below.

1. Date:

2. Which Investigation(s) did you work on this week? (This information will be made available to support staff and contractors to allow them to better support you.)
(check the appropriate boxes)
Water:
Survey (Pretest)
1.1
1.2
1.3
2.1
2.2
2.3
3.1
3.2
3.3
3.4
4.1
4.2
4.3
4.4
Posttest

3. How many days did you teach FOSS this week? (check the appropriate box)

No FOSS this week   1 day   2 days   3 days   4 days   5 days

4. On the days that you taught science, approximately how much time did you spend teaching FOSS? (check the appropriate box)

0 – 20 minutes/day   21 – 40 minutes/day   41-60 minutes/day   more than 60 minutes/day

5. This week, approximately how much time did you spend each day looking at student work after teaching FOSS? (check the appropriate box)

No time   5 minutes/day   10 minutes/day   20 minutes/day   more than 30 minutes/day

6. How many times when using FOSS this week did you provide written feedback on individual student work (notebooks or other) to most students?

0 times   1 time   2 times   3 times   4 times   5 times

7. How many times when using FOSS this week did you reteach content?

0 times	1 time	2 times	3 times	4 times	5 times

8. In the past week during FOSS instruction, how many times did you engage in the following items and activities?

	Used 1x	Used 2x	Used 3x	Used 4x	Used 5x	N/A
Resources						
a. Used the "At a Glance" to review science content and assessment opportunities for your teaching	1	2	3	4	5	6
Assessment work for each lesson						
b. Analyzed students' science notebooks	1	2	3	4	5	6
c. Used a scoring guide to analyze response sheets	1	2	3	4	5	6
d. Recorded observations of students' during class	1	2	3	4	5	6
e. Analyzed **student work** for patterns and trends	1	2	3	4	5	6
f. Analyzed **observations** for patterns and trends	1	2	3	4	5	6
g. Planned further instruction based on patterns and trends	1	2	3	4	5	6
h. What did you learn about students' understanding of science concepts from your analysis of student work? Please provide examples and specific details.						
End of each investigation						
i. Checked on students' understandings at the end of an investigation	1	2	3	4	5	6
j. Engaged students in self-assessment of science learning	1	2	3	4	5	6
k. Other: please specify	1	2	3	4	5	6
l. What did you learn about students' understanding about science concepts at the end of the Investigation? Please provide examples and specific details.						

Comments:

9. What percentage of your students do you think understand the core concepts of the Investigation you taught this week? (check the appropriate box)

0%-25%	25%-50%	50%-75%	75%-95%	100%	Not sure

10. Do you have any questions or feedback about your experience with the project this week? This question is not confidential and responses will be made available to support staff and contractors to allow them to better support you.

Thank you!