Developing Contingent Pedagogies: Integrating Technology-Enhanced Feedback into a Middle School Science Curriculum to Improve Conceptual Teaching and Learning

PROJECT SUMMARY

This proposal is for a full research and development project that addresses the contextual challenge, How can assessment of relevant STEM content improve K-12 teaching and learning? A team led by researchers at SRI International and comprised of experts in formative assessment, curriculum development, the geosciences, and evaluation will develop an assessment-driven intervention aimed at improving learning about the geosphere and that that incorporates classroom network technology, interactive assessments, and a set of contingent curricular activities. Emerging classroom network technologies will enable teachers to pose questions to elicit student thinking and understanding of core constructs and make visible to teachers and students alike what students do and do not understand via shared displays. Interactive assessment activities used in conjunction with these technologies will be designed as repeatable cycles of questioning, responding, discussing, and responding again to advance students’ understanding of concepts. Contingent curricular activities will address persistent student conceptions that diverge from target understandings about the geosphere that are the focus of 2 investigation-based modules of the Investigating Earth Systems middle-school science curriculum. Our hypothesis is that this synthesis of network technologies, interactive activities, and contingent curriculum offers a powerful exemplar for how assessment can improve K-12 teaching and learning.

The intellectual merit of this project follows from the importance of needs we address, an approach that builds systematically on the work both of our team and others, our unique focus on contingent curriculum activities, the quality of our empirical research plan; and the tight fit of our teams’ prior work and expertise to the requirements of this challenge. We address the need to harness the often recommended but rarely implemented practice of formative assessment to improve student learning of complex and foundationally important science content. Our approach builds from knowledge in assessment, cognitive science, and teacher learning in general. In particular, it builds on recent advances in three areas: classroom network technology that enables teachers to pose questions in a variety of formats and review responses of all (not just a few) students, knowledge of how to promote conceptual change in the geosciences, and knowledge of how teachers’ active participation in design activities can lead to changes in teaching and learning. Our unique advance is in developing a usable innovation that synthesizes what is known in contingent curricular activities in geoscience education, scaffolding teacher implementation of adaptive instruction.

This project has strong potential for broader impacts both through dissemination of its intervention in diverse public school settings and through application of its findings to other areas of science instruction. The potential for dissemination will be high because we are choosing to investigate affordable and robust technologies, identifying what teachers need to achieve a high-quality implementation and integrating with a widely-used middle school curriculum, Investigating Earth Systems (IES). The project staff includes an original writer from this curriculum, and there are plans to work with the publisher to incorporate activities developed through the project into a curriculum revision. By addressing the biggest barriers to collecting, aggregating, and using data, this project will present evidence of how and under what circumstances technology, curriculum materials, and teacher professional development can help teachers improve student learning by systematically improving their classroom assessment practice. If our concept of contingent curricular activities meets the challenge of providing teachers with the support they need to connect formative assessment practices to immediate selection of actions that enhance student learning, the concept could readily and powerfully generalize to other science topics and grade levels.
Project Goals

This research and development project addresses the contextual challenge How can assessment of relevant STEM content improve K-12 teaching and learning? A team led by researchers at SRI International and comprised of experts in assessment, curriculum development, the geosciences, and evaluation will develop a complete intervention to improve student understanding about the geosphere by systematically integrating classroom network technology, interactive formative assessments, and contingent curricular activities to guide the teacher from formative assessment to instructional choices and improved student learning. Emerging classroom network technologies will enable teachers to pose questions to elicit student thinking and understanding of core constructs and make visible to teachers and students alike what students do and do not understand via shared displays. Interactive assessment activities used in conjunction with these technologies will scaffold teachers in repeatable cycles of questioning, responding, discussing, and inviting students to respond again. Contingent curricular activities will address persistent student conceptions that diverge from target understandings; with them will be instructional guidance about when to use them, based on data gathered from interactive assessment activities. Our hypothesis is that this synthesis of network technologies, interactive activities, and contingent curriculum offers a powerful exemplar for how assessment can improve K-12 teaching and learning.

The project’s student learning goals emphasize the “big ideas” about the geosphere within an Earth system science context. Specifically, the intervention will help students develop understandings about the composition of the geosphere; how the movement of Earth’s lithospheric plates causes both slow changes in Earth’s surface (e.g., formation of mountains and ocean basins) and rapid ones (e.g., volcanic eruptions and earthquakes); how Earth’s surface is built up and worn down by natural processes, such as rock formation, erosion, and weathering; and how physical evidence, such as fossils and radioisotopic dating, provide evidence for the Earth system’s evolution and development. The project’s teacher learning goals address the critical needs to (1) increase teachers’ knowledge of the different facets of students’ thinking about the target concepts; (2) provide teachers with interactive assessment activities and contingent teaching activities that they integrate into their Earth science units with integrity to the design goals of the intervention; and (3) develop teachers’ proficiency with a comprehensive system that completes the links from formative assessment to adaptive instruction and student learning gains.

The intellectual merit of this project follows from the importance of needs we address, an approach that builds systematically on the work both of our team and others, our unique focus on contingent curriculum activities, the quality of our empirical research plan; and the tight fit of our teams’ prior work and expertise to the requirements of this challenge. We address the need to harness the often recommended but rarely implemented practice of formative assessment to improve student learning of complex and foundationally important science content. Our approach builds from knowledge in assessment, cognitive science, and teacher learning in general. In particular, it builds on recent advances in three areas: classroom network technology that enables teachers to pose questions in a variety of formats and review responses of all (not just a few) students, knowledge of how to promote conceptual change in the geosciences, and knowledge of how teachers’ active participation in design activities can lead to changes in teaching and learning. Our unique advance is in developing a usable innovation that synthesizes what is known in contingent curricular activities in geoscience education, scaffolding teacher implementation of adaptive instruction.

This project has strong potential for broader impacts both through dissemination of its intervention in diverse public schools and through application of its findings to other areas of science instruction. The potential for dissemination will be high because we are choosing to investigate affordable and robust technologies, identifying what teachers need to achieve a high-quality implementation and integrating with a widely-used middle school curriculum, Investigating Earth Systems (IES). The project staff includes an original writer from this curriculum, and there are plans to work with the publisher to incorporate activities
developed through the project into a curriculum revision. By addressing the biggest barriers to collecting, aggregating, and using data, this project will present evidence of how and under what circumstances technology, curriculum materials, and teacher professional development can help teachers improve student learning by systematically improving their classroom assessment practice. If our concept of contingent curricular activities meets the challenge of providing teachers with the support they need to connect formative assessment practices to immediate selection of actions that enhance student learning, the concept could readily and powerfully generalize to other science topics and grade levels.

**Needs of Students, Teachers, and Policymakers Addressed**

Improving classroom assessment is potentially one of the most powerful ways to improve student achievement (Black & Wiliam, 1998; Crooks, 1988; Fuchs & Fuchs, 1986). To realize such gains, however, teachers must use classroom assessments **formatively**, that is, to adjust and improve instruction (Fuchs & Fuchs). Put another way, it is not enough for teachers to collect assessment data; what and how teachers teach must be contingent on what assessments tell teachers about what their students know and can do (Raudenbush, 2002). In that respect, studies consistently indicate that teachers have trouble using assessment data to adjust their instruction (Franke, Carpenter, Levi, & Fennema, 2001; Penuel, Boscardin, Masyn, & Crawford, 2007). Not only are aggregating and interpreting data challenging for teachers, who typically have limited training in analysis of assessment data and face multiple demands on their time, but curricula rarely provide “what if” guidance to teachers about what to do in the case that students are having particular difficulty mastering a concept.

Policymakers and educational leaders have promoted formative assessment, but have not provided the comprehensive support teachers need to use formative assessment to improve instruction. Many localities are seeking to improve student achievement through the introduction of benchmark and interim assessments, coupled with initiatives to engage teachers and administrators in more “data-driven decisionmaking.” Professional development for these initiatives has often been limited, however, and the standardized test reports (whether end-of-year or mid-year) provide little in the way of diagnosis of student thinking that is useful to teacher, much less adequate guidance to teachers about **how to adjust their instruction** based on assessment results (Light, Wexler, & Heinze, 2004). Without adequate preparation for how to make valid decisions about instruction, teachers can draw inaccurate inferences from the data and choose counterproductive strategies to address the real needs of their students (Confrey & Makar, 2005).

An alternative to these tests and the wide range of so-called “benchmark” tests now being pursued by school districts is to **place formative, classroom-based assessment at the center of instructional change**. We need to provide teachers with better classroom assessment tools coupled with professional learning about using assessment to improve instruction (Shepard, 2000), and this goal requires investing in intervention research aimed at developing a complete system of curricular content, technological resources and teacher professional development. Implementation research will develop our understanding of how to build an intervention that systematically scaffolds teachers as they move from formative assessment to instructional decisionmaking to classroom practices that improve student learning. Further, research-based evidence of the effectiveness of such an intervention would allow policymakers advocating for improved opportunities for teachers to learn about classroom assessment to make a stronger case for such investments. This particular project will provide both an “existence-proof” study (Borko, 2004) that a systematic intervention can enable teachers to complete the links among formative assessment, instructional choices and student learning gains and that the necessary system components have low barriers to adoption (e.g., cost, usability, curricular scope) and consequently have high potential for wide diffusion and broad-scale positive impacts on teaching and learning in the geosciences and related subjects.
Description and Framing of the Project
Results of Prior NSF Support and Relevant Research

The Principal Investigator, Co-Principal Investigator, and partners on this project have been involved in or leading research in the areas that form the basis for the proposed project. We review this research as a way of presenting results of prior work relevant to the project and place it within the context of others’ research as it informs our project design.

Classroom Network Technologies to Support Interactive Classroom Assessment

As part of an earlier NSF grant (REC-0337793), the Principal Investigator for this proposed project conducted a systematic review of the potential of different forms of classroom network technologies for improving teaching and learning (Penuel, Roschelle, & Abrahamson, 2005). Broadly speaking, classroom network technologies are any form of technology that enables teachers and students to share questions, ideas, data, or responses via a local classroom network. To date, the most research has been conducted on a class of network technologies sometimes called audience or student response systems. Student response systems typically include handheld devices for every student that are networked together and to a teacher computer and a means to display the teacher’s screen that reveals the responses of students to a teacher question. With a student response system, teachers devise a question, such as “Why are most of the volcanoes of the world in the ‘Ring of Fire’?” and provide a list of possible answers to the questions from which students select. All students then use handheld input devices (sometimes called clickers) that are linked by a classroom communication or student response system and that allow them to answer the question. Once all students have answered, teachers can project a histogram showing the distribution of student responses. In one model of how to integrate this technology into instruction developed for undergraduate physics instruction called Peer Instruction (Mazur, 1997), the teacher then asks the students to discuss their answers with a peer or orchestrate a whole-class discussion. Evidence from a number of studies point to potential benefits of student response systems, including increased participation and engagement, improved conceptual understanding, and better feedback to students and teachers about what they know and can do (Crouch & Mazur, 2001; Penuel et al., 2005).

There are some limitations of student response systems and their use that will be addressed directly in the current study. First, most of today’s low-cost systems allow teachers to pose only one format of question (multiple choice). Researchers and teachers have been critical of multiple-choice items for diagnosing students’ preconceptions, (e.g., Berlak, 1992). Alternative network technologies enable teachers to pose questions that require students to construct a response; the responses, moreover, can be in the form of words, symbols, or student drawings (see Exhibit 1) and we expect that, over the grant period, the cost of these richer systems will steadily decrease to an affordable level. Building on the co-Principal Investigator’s prior work on the Calipers Project (DRL-0454772), we can leverage the more advanced network technologies to assess complex skills through visualization of phenomena and other advanced task types.

• In the current project, in order to assess the potential limitations of these new technologies, we will test out both response systems and GroupScribbles with teachers and students in the first phase of our project.
New classroom network technologies now offer the support real-time classroom assessment that makes thinking visible, breaking out of the limited multiple-choice paradigm. As part of an NSF Information Technology Research (ITR) grant (#0427783), SRI and its collaborators developed GroupScribbles (http://groupscribbles.sri.com/), collaborative software that allows individual students to compose “scribbles” on “post-it” notes in a private space in response to a public task. They can post these notes anonymously into a shared, public space that becomes the object of discussion (see below).

In one mathematics assessment, the teacher posed the question represented above, “What are different ways to represent the following fractions?” (i.e., 1/4, 2/3, and so on). Students then created their own representations and “moved” their answers into the public space for class discussion. A parallel question in Earth science is: “Draw a diagram to explain how mountains are formed.” A teacher may pose this question using the Peer Instruction model (i.e., each student constructs a diagram, puts it in the public space, and then tries to convince a peer of why her or his diagram is accurately represents the mountain formation process).

We intend to systematically explore the trade-offs between newer technologies like GroupScribbles and more traditional response systems. On the one hand, GroupScribbles enables a much broader array of question types that teachers or students can pose than do student response systems. On the other hand, GroupScribbles displays may be harder for teachers to interpret in the course of real-time instruction. Student response systems have the advantage that they condense student input to a histogram of the frequency of student choices; the multiple choice alternatives can be developed to represent different facets of student conceptions that are known to exist, so that teachers are not left to interpret student responses on their own. There is a tradeoff between “making thinking visible” with more richness and guiding teacher action through predictable contingencies and existing research is insufficient to tell us what to do.

- We will begin this project by comparing the trade-offs between different network technologies; in addition, at the beginning of the project, we will develop features to GroupScribbles to facilitate the interpretation of student responses through additional aggregation features.

Interactive Formative Assessment Activities and Contingent Curricular Activities

Black and William (1998) identify four features of effective formative assessments: they enhance feedback between teachers and students; they require active involvement of students; they need to be motivating to students; and they need to be used to adjust teaching and learning. Results from past studies of classroom networks suggest that networks greatly facilitate feedback, support classroom participation and engagement, and are motivating to students, but these studies also suggest that success is dependent upon teachers posing good questions and having resources that allow them to adjust instruction, contingent upon what they find out students know and can do (Beatty, Gerace, Leonard, & Dufresne, 2006).
Of particular importance in science education are questions designed to diagnose student thinking. A diagnostic line of questioning seeks to elicit those preconceptions and their subcomponents, which can be called “facets” of student thinking (Minstrell & Kraus, 2005). Diagnostic questioning is consistent with a growing body of cognitive science research that suggests it is necessary to engage rather than ignore problematic student ideas in order to promote conceptual change in science (di Sessa & Minstrell, 1998; National Research Council, 1999; Posner, Strike, Hewson, & Gerzog, 1982; Smith, di Sessa, & Roschelle, 1993-1994). The diagnostic approach, in its emphasis on eliciting student thinking, stands in contrast with the typical approach of science curricula and with the view of some advocates (e.g., Muthukrishna, Carnine, Grossen, & Miller, 1999) that focus solely on teaching goal facets or correct concepts to students.

In the approach developed by Minstrell and colleagues for diagnostic questioning, teachers elicit student facets of understanding through a sequence of multiple-choice questions that are focused on obtaining student explanations of phenomena, not just on recall of information (Minstrell & van Zee, 2003). Researchers develop these questions using think-aloud protocols and assessments of students aimed at eliciting the variety of student ways of thinking about a particular concept. Minstrell and colleagues have developed diagnostic questions and integrated them into interactive classroom discussions in physics (di Sessa & Minstrell, 1998). This approach is currently being applied in the domain of chemistry in project led by the co-Principal Investigator (DRL-0733169). In that project, SRI is working with Minstrell and his colleagues at FACET Innovations to develop both assessments of student thinking in chemistry and materials to support teacher learning about facets and use of the assessments. Other researchers have used the same approach to develop diagnostic questions in astronomy (e.g., Briggs, Alonzo, Schwab, & Wilson, 2006; Treagust, 1995). The resulting assessments have desirable psychometric properties: diagnostic questions in astronomy have been shown to successfully discriminate among students of different ability levels using IRT analysis (Sadler, 1998) and to aid in the identification of developmental sequences of learning of constructs (Briggs et al., 2006).

- In the current project, we will develop facet-based questions for teachers to use that are designed to elicit student understanding of core geoscience concepts addressed in the Investigating Earth Systems curriculum.

Developing recommendations for how to adjust instruction on the basis of particular patterns of student thinking requires additional knowledge about the kinds of strategies that are likely to be effective in addressing different facets. In geoscience education, the focus of the current project, experimental studies have demonstrated the efficacy of using models and analogies to teach the rock cycle (Blake, 2004) and the causes of volcanism and earthquakes (Gobert, 2005). Furthermore, experimental studies have demonstrated that having students generate diagrams of processes in the interior of the Earth or having them generate explanations is more effective than having them summarize text on the topic (Gobert, 2005; Gobert & Clement, 1999).

- In the current project, we will draw on cognitive science research on effective strategies for teaching about the geosphere as an initial basis for developing elicitation questions and to develop contingent curriculum activities for teachers to use, in case students have difficulty mastering concepts with existing activities in the Investigating Earth Systems curriculum, a curriculum developed with NSF funding and that is widely used in middle schools.

Feasibility of the Proposed Approach

The viability of using classroom network technology depends in part on its reliability and usability. Response system technologies have been in use for close to 30 years in classrooms, including K-12 settings. Teachers report that, relative to many other technologies, they are readily implemented into
classroom teaching routines, including the kinds of interactive assessment activities described above (Penuel et al., 2007). The newer GroupScribbles technology builds upon what we know about why teachers find response systems easy to adopt. GroupScribbles is simple, based upon a familiar metaphor, and builds upon typical classroom interaction patterns. Our experience is that teachers can learn the basics of using GroupScribbles in 3 to 5 minutes and that many teachers can readily invent new activities to fit their classroom needs. It is a robust, flexible, and simple technology that has been designed with classroom realities in mind. We plan to use co-design (Penuel, Roschelle, Shechtman, 2007) as our strategy for enhancing the usability of the technology-supported intervention as a whole.

We also believe the feasibility of our project is enhanced by the fact that in this area of the geosciences, we can leverage significant prior work in both assessment and instruction. The Geoscience Concept Inventory (Libarkin & Anderson, 2005) is a 73-item multiple choice assessment of student understanding of concepts about the geosphere; its distractors include common student misconceptions. Although developed for undergraduates, it forms a useful starting point for developing interview protocols to elicit middle-schoolers’ concepts. With respect to instruction, cognitive scientists have identified a number of potential strategies for addressing student difficulties, including diagramming (Gobert & Clement, 1999; Sibley et al., 2007), using technology-supported models of phenomena, and peer critique of student models (Gobert, Snyder, & Houghton, 2002). We can use knowledge of these strategies to bootstrap the design and development of contingent curriculum activities.

**Logic Model Guiding the Project’s Activities**

The logic guiding the design of the project and selection of core components reflects the research presented above about both the affordances of network technologies and interactive pedagogies. It also reflects the research on how professional development activities, teacher characteristics, and student characteristics could impact the efficacy of the proposed intervention. These factors are integral to our thinking about sampling and about the design of the intervention.

**Nature and Format of Resources to Be Developed and Tested**

Over the course of 4 years, SRI proposes to develop and test the following key resources for teachers and students:

- pedagogical routines for teachers to follow when using classroom network technology
- diagnostic questions teachers can use to elicit student preconceptions
• decision rules for teachers to follow for using alternative learning activities to supplement the IES curriculum and support contingent teaching
• training materials for preparing teachers to enact the intervention
• research- and classroom-based instruments for measuring changes in instructional practice and student achievement in targeted Earth science concepts

Curricular Scope

The intervention will cover between 15 to 20 weeks of the Investigating Earth Systems (IES) middle-school Earth science curriculum. IES is a National Science Education Standards (NSES)-based, Earth science curriculum for the middle grades. The American Geological Institute (AGI), in association with It's About Time Publishing, developed IES with support from the NSF, the AGI Foundation, and Chevron Corporation. Field tested and content reviewed, IES is one aspect of AGI’s efforts at implementing effective Earth science education reform. The intervention will focus, in particular, on part of the curriculum that addresses the geosphere, including interactions with the hydrosphere, atmosphere, and biosphere. The student learning goals that are the target of this intervention are of great practical significance to teachers and students, since both are constructs central to Strand D of the NSES (National Research Council, 1996).

Beyond content coverage, these constructs afford opportunities to develop and assess key student skills in science. For example, exploring the history of the theory of plate tectonics, which is included as part of these materials, provides an opportunity for students to develop an appreciation not only for what the theory explains but also for scientific discoveries and social processes that resulted in it eventually becoming accepted by the scientific community (Oreskes, 2003). In addition, developing an understanding of phenomena linked to plate tectonics requires the development of students’ model-based reasoning abilities (Gobert, Buckley, Levy, & Wilensky, 2007). In particular, developing an understanding of Earth’s dynamic structure requires that students construct and be able to use a model of the Earth’s interior to reason about and relate the phenomena of earthquakes, volcanic activity, and major physiographic features of the Earth.

Classroom Network Technology Development (Year 1; Refinements: Years 2-3)

In this project, we will primarily be selecting existing classroom network technologies that are best suited for supporting classroom assessment in Earth science. The two sets of technologies have different affordances for classroom use, which we will investigate through a pilot in the first phase of our project, will be GroupScribbles (used with carts of Tablet PCs shared by teachers in a school) and eInstruction’s CPS system (a student response system, requiring a teacher laptop and individual student IR handheld devices). The process to be used to select which technology will be used is described on pp. 10-11.

As part of the project, we will use existing CPS systems but will make some minor modifications to GroupScribbles to enable some forms of data aggregation. For example, it is a relatively simple matter for computational questions to develop the capacity of the technology to aggregate student answers and display a distribution of responses, much as the CPS system does. A more advanced feature we will explore is to develop “building blocks” for drawing and representing dynamic geospheric phenomena (e.g., representations of plates that can be put together to represent different types of faults). We will be able to develop algorithms to recognize target facets in student constructions from these blocks, which are not “free-form” but require students to construct their own representations using a fixed set of tools.

Diagnostic Questions for Eliciting Student Conceptions (Years 1-2; Revisions: Year 3)

Diagnostic questions for eliciting student conceptions will be based on an analysis we conduct in this project of the different facets of student thinking with respect to students’ ideas about the geosphere as encountered in the curriculum materials. Each facet cluster will include one or more goal understandings, as well as preconceptions (problematic facets) related to the goal understandings. The development team will collaborate to create 10 to 12 facet clusters related to topics addressed in the two modules, with approximately 1 cluster per investigation. Associated with these clusters, we will develop a set of questions
designed to evoke student preconceptions. We will develop between one to three questions to stimulate students’ thinking about the topic of each curricular investigation.

**Interactive Assessment Activities (Year 1; Revisions: Years 2-4)**

Peer Instruction (Mazur, 1997) offers a model of interactive assessment activities to be integrated with response system technology for use in classrooms. SRI has found that K-12 teachers can and do implement interactive assessment components (Penuel et al., 2006), but the prevalence of these components has not been studied among K-12 teachers. Relying on teacher-leader input to the design process, the project will adapt Peer Instruction strategies for middle school Earth science instructors, with particular attention to articulating for teachers how to support both paired and whole-class discussion with diagnostic questioning. Both are potentially advantageous for teaching with response systems: paired discussions have the potential to increase student involvement, but limit teacher monitoring. Class discussions allow the teachers to know more of what particular students are thinking, but they limit involvement of all students and can be difficult for teachers with limited content knowledge to orchestrate.

**Supplementary Contingent Curriculum Activities (Year 2; Revisions: Years 3-4)**

We will develop contingent or alternative learning activities embedded within the Investigating Earth Systems (IES) modules to provide students with additional instruction in areas in which they need additional instruction. The number and types of activities will be determined by reviewing student performance on existing classroom assessments and identifying content that is persistently challenging for students. Contingent or alternative learning activities will be developed that are based on what is known today about strategies for improving students’ conceptual understanding of the geosphere. The team will develop at least one alternative activity for each investigation; however, depending on what is learned from studying student thinking, alternative activities may be concentrated on investigations that focus on preconceptions that are particularly difficult to change. These materials will have the same format as the rest of the IES curriculum; we hope to distribute them as part of a planned cycle of revision with the curriculum’s publisher.

**Decision Rules for Using Contingent Learning Activities (Year 2; Revisions: Years 3-4)**

A key resource for teachers in the project will be a set of decision rules or heuristics for selecting and using alternative learning activities. As part of the project, we will explore the feasibility of providing specific guidance for teachers about what implications of particular distributions of facets in a class or group of students have for instruction. We see this aspect of the project as exploratory, since little has been done in our field to establishing an empirical basis for recommending, for example, what to do in class if 30 percent of the class has a target facet compared to 50 percent, and whether small group or individualized instruction might be warranted.

**Training Materials to Prepare Teachers for Enactment (Years 1-2; Revisions: Years 2-4)**

The training materials will include specific hands-on activities for teachers to gain practice in the pedagogies, opportunities to receive feedback from leaders of the professional development, and classroom-based follow-up after initial training. For example, in the training SRI will lead interactive assessment activities using the technology, initially with teachers playing the role of students. Then, teachers will engage as leaders in one cycle of questioning with their peers. In addition, at least part of the training will provide teachers with additional grounding in the subject matter focus of the modules. This will take the form of a presentation of the rationale for each concept’s inclusion in the IES modules.

Additional training materials are needed to help teachers learn what questions to pose, how to pose them, and the background behind the diagnostic questions we will provide as part of the curriculum. These materials will specify the different questions to ask and the facets of student thinking they are intended to reveal. The training materials will also specify and provide teachers with opportunities to discuss variations
in student response, as well as provide them with practice in using the data to make decisions about what
contingent activities they might use to advance students’ conceptual development.

Instruments for Measuring Implementation of Formative Assessment System (Years 1-2)

SRI has developed and piloted a questionnaire of K-12 teachers’ goals and approaches to integrating
response system technologies in their classrooms. SRI will adapt the goals and teacher practices items
from the questionnaire to align with the model of interactive pedagogies developed for the project and on
the basis of what technology is ultimately selected for use in the classroom. SRI will also develop new items
about the reliability of the technology and about teachers’ use of specific questions from the IES modules.

SRI will develop four new instruments to measure the implementation of the pedagogies. Multiple
measures are needed because each source of data has both advantages and limitations. The student
questionnaire will be the primary source of data about students’ perceptions of the intervention, especially
their attitudes toward participation in class. These data will be compared with teachers’ perceptions of
student participation, which will be gathered through interviews. Interviews will also be used as a source of
data on teachers’ cognition regarding instructional planning and decisionmaking. Logs or teacher diaries
provide more accurate portraits of teacher practice than do surveys because the problems associated with
memory are significantly reduced by the use of logs (Rowan, Camburn, & Correnti, 2004). Data obtained
from the logs will be compared with questionnaire data and use the logs on events that might be difficult for
teachers to remember, such as small technical problems with operating response systems. Finally, the
observation protocol will be our primary source of data on classroom discussion and participation, which are
difficult to capture by retrospective logs, surveys, or interviews. Observations will also afford opportunities
for uncovering unforeseen challenges to implementation that are critical to informing revisions to materials.

Expertise

One of the most important reasons why this project is feasible at the present time is that we have
assembled a team uniquely poised to bring the breadth of expertise required to create a comprehensive,
viable intervention through a co-design process. Creating a robust intervention requires expertise in
developing usable technology, questions that reflect the underlying science of how students learn, curricular
activities that can address student difficulties, and researchers to study implementation and outcomes in a
way that provides compelling, preliminary evidence of the success of the approach.

Project Leadership

Dr. William R. Penuel, PI, is Director of Evaluation Research at the Center for Technology in Learning
(CTL) at SRI, and has extensive experience both in evaluating science education programs and in
designing technology-supported innovations. He is currently PI on a study funded by the Institute of
Education Science that is comparing the efficacy of three approaches to Earth science professional
development; in the past, he served as PI on an NSF-funded grant to develop and test handheld-supported
formative assessment tools in science classrooms. Dr. Penuel will coordinate all development and research
activities, facilitate the co-design process, lead the implementation research studies, and prepare all reports.
Co-Principal Investigator, Dr. Angela DeBarger, is a Research Social Scientist in CTL. Dr. DeBarger is an
expert in motivation and in educational assessment and is PI on an NSF-funded initiative to develop
assessments of student thinking in chemistry based on facets of student thinking. She will oversee
development of the facet clusters and their integration into the intervention, and lead the development and
validation of research instruments for the study.

Development Support

Mr. Colin Mably, independent consultant to the project, is an expert in science curriculum development
and teacher professional development and an author of the IES modules that are the focus of this project.
He will identify existing assessment questions for use with the interactive technologies and develop teacher
training materials for the first implementation study, develop and revise replacement or supplementary activities to use as resources to support teachers’ contingent teaching, and lead training to teachers for both implementation studies. **Dr. James Minstrell** of FACET Innovations is an expert in science assessment who has devoted more than 15 years to identifying student preconceptions of important concepts in K-12 science. He will facilitate the development facet clusters for Earth science content in the IES modules and elicitation questions and related activities for those modules based on the facets and support the implementation study by developing appropriate training materials for teachers. **Dr. Dina Venezky** of the U.S. Geological Survey is an Earth scientist with experience in working collaboratively with educational researchers; she will provide Earth science expertise. **Ms. Cheryl Mosier**, a teacher of Earth science at Columbine High School in Jefferson County, Colorado, has used response systems in teaching science and will participate as a member of the design team and summer workshop leader. **Two middle school science teachers** experienced in teaching Earth science will be invited to serve on the co-design team.

**Evaluation**

**Dr. Christy Kim Boscardin** will serve as the project evaluator. Dr. Boscardin is an expert in multi-level evaluation design and analysis, and has worked on two related projects, including the earlier project in which the Principal Investigator was involved, as well as a large scale randomized field trial being undertaken in mathematics that is funded by the U.S. Department of Education. Dr. Boscardin, who has been with the Center for Research on Standards and Student Testing (CRESST) at UCLA for the past several years, will be an independent consultant to the project.

**Advisory Board**

The project will assemble an Advisory Board of recognized experts in science education, formative assessment practices, Earth science, and the use of response system technology for teaching. Throughout the project, the Advisory Board will monitor the interactive pedagogies and instrument development activities, review drafts of materials, and provide feedback on data collection methodologies and analyses. The Advisory Board will convene at SRI International annually.

**Dr. Ann Benbow**, Director of Outreach and Development at the American Geological Institute, is an expert in Earth science and is one of the contributing authors of the IES curriculum. **Dr. Jere Confrey**, Professor of Math, Science, and Technology Education at North Carolina State University, brings expertise in using diagnostic assessments in mathematics. **Dr. Douglas Duncan**, Director of Astronomical Laboratories at the University of Colorado, is the author of *Clickers in the Classroom* and conducts faculty workshops on using response systems. **Dr. Richard Duschl**, Professor of Science Education at Rutgers University, brings expertise in helping teachers make students’ scientific thinking visible and also has a M.A.T. in geology. **Dr. Scott Linneman**, Associate Professor of Geology and Science Education at Western Washington University, will provide Earth science expertise. **Dr. Edward Roebeck**, Associate Professor of Education at Salisbury University will provide expertise in curriculum development. Letters are on file for Drs. Benbow and Duschl.

**Research and Development Methodology**

**Approach to Design and Development**

Many innovations developed by experts in curriculum or researchers fail because they are not applicable for real classrooms. SRI’s approach to this challenge is to *co-design* interventions. Co-design is a facilitated, team-based process in which teachers, researchers, and curriculum developers work together in defined roles to design an educational innovation, realize the design in one or more prototypes, and evaluate each prototype’s significance for addressing a concrete educational need (Penuel, Roschelle, & Shechtman, 2007). Co-design begins with the formation of a team and a specification of roles on the project. For this project, the PI will serve as leader of the team and be fully accountable for the quality of the team’s work and resulting intervention. The team will also consist of two teacher-leaders, to be selected from the Denver
Public Schools, two assessment researchers (Dr. DeBarger and Dr. Minstrell), a curriculum developers (Mr. Mably), and a subject matter expert in Earth science (Dr. Venezky). The task descriptions below outline the roles of each of these team members and their specific project responsibilities.

Co-design also follows an iterative design process that involves cycles of planning, testing, and revision of materials. As part of the planning process, co-design teams develop detailed scenarios of use (Carroll, 1995) that describe teaching situations and strategies that are expected to be part of the intervention. In addition, in co-design, teams use design personas, images of different types of likely users (teachers) of an intervention. From scenarios and personas, specific support materials are designed to prepare teachers who may share characteristics of design personas for enacting the intervention in a range of classroom situations. Those materials are then tested on real classrooms, and using data from the classroom tests, revised to address the difficulties faced by teachers in classroom implementation. As appropriate, the scenarios and personas to guide further revision of materials.

Approach to Instrument Development and Validation

There are two sets of instruments that are to be developed in the study: (1) diagnostic questions that elicit different facets of student thinking about the geosphere and (2) instruments to study implementation of the intervention and effects on teaching and learning. Below, we describe the approach to developing each.

**Facet Cluster Development.** Earth science teachers and project staff with Earth science teaching, Earth science content, and cognitive research expertise will participate in facet cluster development. They will (1) review the middle school standards (NSES, AAAS) related to Earth science, (2) discuss how content related to these topics is typically taught in the two modules (big ideas, sequence), (3) specify learning objectives for each topic and distill these learning objectives into goal facets with smaller – grain size that can be assessed, (4) discuss problematic ideas that students bring to the classroom about these topics, and (5) draft open-ended questions that can be used to elicit different facets of student understanding. During the meeting, the development team will create at least five facet clusters. During follow-up teleconference calls, the development team will collaborate to create 10 to 12 facet clusters related to topics addressed in the two modules, with approximately 1 cluster per investigation.

SRI will validate the facet clusters by conducting cognitive interviews with Earth science experts and novices. Questions in the cognitive interviews will be based on those drafted by the facet cluster development team. SRI will analyze responses for evidence of goal facets and problematic facets specified in the facet clusters. The cognitive interviews will be conducted with two or three Earth science teachers and/or Earth scientists (including members of the advisory board), two or three middle school students who will take Earth science in the coming school year, and two or three middle school students who have just completed their Earth science course. SRI anticipates that these interviews will each take 2 hours. In addition, Earth science teachers who participate on the facet cluster development team will be invited to present these questions to their students as they introduce the content in class, and students’ written responses will be gathered and analyzed for additional evidence of goal and problematic facets. Facet clusters will be revised on the basis of interview and written data.

**Research Instrument Development.** The approach to new instrument development will follow a principled assessment design process (Mislevy, Steinberg, Almond, Haertel, & Penuel, 2003), driven by the conceptual model for the study. Measures and scales are needed for each of the student characteristics, teacher characteristics, teacher preparation activities, and enactment components. Available items and scales have already been reviewed and will be adapted to align with the conceptual model; new items will be developed for constructs for which no available items exist. Before piloting, SRI will submit each instrument for review to both the design team and the external Advisory Board. SRI will ask these team members to rate the strength of alignment of each item with the conceptual framework and to suggest revisions to improve alignment and comprehensibility. As part of a pre-pilot before the classroom testing of
the intervention, we will conduct cognitive interviews with three to four middle school teachers using the protocols.

**Approach to Implementation and Outcome Research (Formative and Summative Evaluation)**

The *formative evaluation* research will focus on testing of technology-supported interactive pedagogies and training materials to support existing curricular activities. As part of the formative evaluation, we will address the following questions:

- How often and how reliably do teachers integrate the designed pedagogical routines into their instruction?
- How do students of different backgrounds and attitudes toward class participation perceive the benefits and drawbacks of using response systems in science class?

The *summative evaluation* research will focus on implementation of the complete intervention. In addition to continuing to investigate the questions above, we will address the following questions:

- How often and how reliably do teachers use diagnostic questions to measure students’ conceptual understanding and then adjust their instruction, using the designed decision rules of the interventions?
- How does implementation vary for teachers with different levels of content knowledge and prior experience with using technology in their classrooms with students?

The second phase of evaluation research will also include a study to assess student conceptual understanding and test teachers’ ability to implement diagnostic questions and engage in contingent teaching. To help guide planning for a later efficacy study, we plan to recruit 20 teachers to the study from Denver Public Schools and neighboring districts and 20 comparison teachers with similar demographics; to leverage what teachers have already learned, and balance the need for rigorous data with the costs of investing in new teacher training activities, we will conduct the evaluation as a quasi-experimental study, comparing gains in teaching and learning from the treatment group (which will have been testing components of the intervention for 2 years) with a comparison group implementing the same curriculum, but without enhancements.

Data on implementation will inform planning for revisions to the intervention in three ways. First, data will be used to identify additional technology support and training needs if teachers report that they experience significant difficulties that affect more than one to two students when they use their response system technology. Second, the data will be used to focus efforts to identify phases of instruction where teachers find it easier to use the instructional routines. If a majority of teachers successfully implement activities in a particular phase of a module, greater attention will be placed on developing additional questions for that phase. Third, data from activities that were not successfully enacted will be analyzed to determine whether the activity needs to be revised or eliminated, or whether additional training should be provided to teachers.

**Co-Design Activities**

The co-design of materials will occur in two distinct phases. In the first phase, the focus of the effort will be on selecting and implementing network technologies with interactive assessment activities, and in the second phase on augmenting activities with diagnostic questions and contingent curriculum activities.

**Phase 1: Selecting and Implementing Network Technologies with Interactive Assessment Activities**

SRI will form a co-design team comprised of four to five teachers from Vallejo City Unified schools, a curriculum developer, and SRI researchers to engage in this first phase of the work. During this phase of work the primary goals will be to test how well teachers can implement interactive assessment activities in conjunction with GroupScribbles and Student Response Systems and then choose a classroom network technology to implement as part of a field trial in Phase 2 on the basis of that test.
As a starting point, the design team will articulate how to adapt the Peer Instruction approach to integrating assessment and network technology and use existing assessment questions embedded within the Investigating Earth Systems materials as the basis for the trial. During this phase, we will design opportunities for teachers to learn about the technology; these designs will also form the basis for a first draft of the professional development materials related to technology use.

Teachers will then try out the pedagogies and questions using both forms of technology in the first year of the project over a period of 3 to 4 weeks. A local researcher who will consult to the project will take ethnographic field notes, and the researcher and teacher will together report back to the design team via teleconference about what is learned from the test. The teleconference debrief will follow a protocol that reviews technology-related issues, as well as teacher and student interactions with the technology and with each other, in order to surface issues. On the basis of these trials and debriefing sessions, the team will select one technology as the basis for subsequent development and implementation research. In addition, we will develop a set of materials for preparing teachers to implement interactive assessment routines with students in their class using the experience of co-design as a guide.

**Phase 2: Augmenting Activities with Diagnostic Questions and Contingent Curriculum Activities**

In parallel to the first phase, but extending beyond it, FACET Innovations will lead the development of diagnostic questions and contingent curriculum activities for inclusion as part of our implementation research. Elicitation questions will be based in part on questions developed to validate the facet clusters. These questions will be refined and additional ones will be developed, with a focus on diagnosing student thinking as part of the IES curriculum. Phase 2 will also include development of a set of teacher professional development materials designed to help teachers understand and interpret student responses to the eliciting questions. These materials are especially critical if GroupScribbles is chosen as the platform for development, since student responses will be open-ended and require interpretation by teachers.

**Formative and Summative Evaluation: Implementation and Outcome Research Design**

Sample

We will be working with two urban districts in this project selected for their diversity and for alignment of district leaders’ goals with the project goals. Teachers from Vallejo City Unified School District (VCUSD), will participate as a collaborator in the co-design process and as one site where SRI will conduct the study. VCUSD is a diverse district; the population of low-income students ranges from 32%-54%, and 84%-98% of students are from underrepresented groups. Denver Public Schools will participate as another research site. In Denver Public Schools, 67% of students receive free and reduced lunch and 80% of students are from underrepresented groups. Denver Public Schools is ideal because the district has adopted the IES curriculum at the middle school level. Because SRI hypothesizes that subject matter knowledge and technology proficiency are likely to affect teachers’ comfort in leading discussions that elicit and engage student conceptions, it will seek teachers who have both extensive and less extensive backgrounds in Earth science and experience with technology. For the study, we will select a total of 20 classrooms to participate.

**Sources of Data**

*Teacher Questionnaire.* A teacher questionnaire will address the following elements of the logic model: teacher characteristics, classroom characteristics, and enactment of interactive pedagogies. SRI will adapt the questionnaire from one developed as part of an earlier NSF-funded study of K-12 teachers’ uses of response system technologies (Penuel et al., 2006). The earlier questionnaire asked teachers about their goals for using response systems, the instructional strategies they employ when using the system, and perceived effects. SRI will add items to this survey to measure teachers’ prior general use of technology with students, their comfort with the subject matter of the modules, class size(s), and available instructional resources other than the IES modules for teaching the content. In addition, to support the development
process, SRI will include a section in each questionnaire that solicits teachers’ feedback on the ease of use and perceived value of each of the activities and on how the activities might be improved.

**Student Questionnaire.** A student questionnaire will address students’ attitudes and perceptions of participating in classes where response systems are used. In addition, the questionnaire will collect background data on students’ ethnicity, gender, and past achievement.

**Teacher Interviews.** Interviews will serve as a primary source of data about teachers’ instructional planning process. The interviews will include questions about how much time teachers spend preparing for classes in which they use the instructional routines compared with other classes; questions about shifts, if any, in their instructional planning process; and questions about their decisions to adjust instruction as a result of their use of the instructional routines.

**Classroom Observations.** A classroom observation protocol will provide data on enactment of the pedagogical routines and on the technical difficulties encountered. The observers will also rate the nature and quality of student thinking elicited and incorporated into classroom discussions they observe. To capture the immediate consequences of technical breakdowns, SRI will adapt items from its observation protocol developed for the Evaluating Educational Technology Interventions study for the Institute of Education Sciences (Dynarski et al., 2007).

**Test of Earth Science Concepts.** A test of Earth science concepts will measure student learning of the rock cycle, destructive and constructive forces of Earth’s surfaces, and plate tectonics. This test will be adapted from the assessment SRI is using as part of the study comparing the efficacy of IES to other approaches in preparing teachers to impart deep conceptual understanding in Earth science to students. That assessment includes items specifically developed for the efficacy study, as well as released items from international, national, and state examinations; the assessment is also aligned with the NSES Earth science standards. This test went through a process of expert review as a means to test content validity; data analyses also showed that test scores were related to student reports of content coverage for relevant items, indicating sensitivity to instruction. The test has good reliability ( = 0.82) and includes a number of items that adequately discriminate among students of different abilities.

**Implementation Procedures**

Co-design team member Mr. Colin Mably will collaborate with SRI staff to lead the initial training with the pilot group of 20 teachers. The initial training will last 4.5 days, with 1.5 days focused on interactive pedagogies and enhancing teachers’ content knowledge; 0.5 day focused on technology, and 2 days focused on learning the IES modules. The leaders of the training session will use the materials prepared as part of the Phase 1 development activities. In addition to the initial training, each teacher will receive approximately 2 days of follow-up throughout the year. Follow-up will entail a visit by a member of the design team to the teacher’s classroom to discuss and troubleshoot implementation. If necessary, design team members may also model use of the pedagogical routines with teachers to help those who are struggling determine how the routines can be enacted with their students. After each visit, design team members will complete a debriefing form describing their activities in the classroom and any difficulties the teacher has experienced.

SRI will ask teachers to implement the IES modules in accordance with the district curriculum demands, using the interactive pedagogies with the response system technology and with selected questions from the curriculum. As part of the test, teachers will be asked to allow an observer into their classroom and to be interviewed by project staff about their implementation of the intervention. In addition, teachers will be asked to administer pre- and post-assessments of Earth science knowledge to students in their classes.
Before the third year of the project (and as part of the second year of implementation research), SRI and Facet Innovations will collaborate to deliver an additional 3.5 days of professional development to the same teachers focused on the new diagnostic questions and contingent teaching activities. Teachers will participate in the same research activities; we will also identify an additional 20 comparison classrooms in which to collect research data on teacher practices and student learning outcomes.

**Analysis Plan**

Analysis of teachers’ ability to implement the pedagogies reliably will focus on four key elements: (1) technology reliability, (2) which intended uses of pedagogical routines do teachers enact most and least often, (3) fidelity to the pedagogical routines, and (4) teacher characteristics associated with higher levels of implementation. First, the evaluator will analyze the reliability of the technology by examining log data to ascertain the percent of uses that involved technical difficulties and by documenting the number of student response pads that experienced problems over the course of the two modules. Second, the evaluator will analyze which questions teachers asked most often and least often as reported on the survey and use interview data to determine why particular questions were included or excluded. Third, for each aspect of the interactive assessment activities, the evaluator will use observational data to analyze the percent of teachers who enacted the element whenever they used the technology. Finally, the evaluator will examine associations between teacher characteristics (goals, confidence in the subject matter, and prior use of technology) and reliability of implementation.

Analysis of student outcomes will vary by year of implementation research. In the first year of implementation research, the evaluator will analyze student gains and test whether implementation levels predict levels of student gains using a hierarchical linear modeling (HLM) approach (Raudenbush & Bryk, 2002). Models will include teacher-level variables, of which the primary variable of interest will be frequency with which teachers used interactive assessment activities in conjunction with network technologies. In the second year of the implementation research, the same HLM approach will be used, but implementation variables will be replaced by a dichotomous variable indicating treatment condition (treatment vs. comparison), and the effect of being in a treatment classroom on student gains will be analyzed.

**Dissemination**

During the last 2 years of the project, we will work with staff from the American Geological Institute and its consultants as they work on a planned revision to the IES curriculum. The goal will be to incorporate either as integral parts of the curriculum or as supplementary materials those materials we develop as part of this project, for publication by It's About Time. Because our team includes representation from that revision effort, we are in a good position to incorporate such revisions into a new version and thus achieve broader impact of our development efforts than the teachers that with whom we work most intensively.

Dissemination will also occur through traditional channels of academic publication, developing presentations specifically for educators and facilitating participation of local research teams studying classroom implementation. We will seek to publish results of our initial case study research and early findings with respect to the measurement of interactive pedagogies beginning in Year 2 in such venues as the *Journal of Learning, Technology, and Assessment* and the *Journal of Research in Science Teaching*. Subsequent measurement case study work will be submitted for publication in Year 3, to journals in educational technology, such as the *Journal of Computers in Mathematics and Science Education*. We plan to develop a manuscript from the quasi-experimental study to submit to the *American Educational Research Journal*. In addition to these publications, we will develop a workshop that could be delivered at the Association for Supervision and Curriculum Development (ASCD), which reaches many curriculum developers, particularly those aligned to an assessment-driven approach.
References


