

Abstract Title: Evaluating Short-Term Impacts on Student Achievement: What Does Student Motivation and Reflection Tell Us?

MSP Project Name: Rhode Island Technology Enhanced Science (RITES)

Author(s): Elise Arruda Laorenza, Stephanie Feger, & Joye Whitney

Presenter(s): Elise Arruda Laorenza

120 word summary:

This presentation focuses on the use of student motivation and reflection as indicators of short-term MSP program impact. With evidence and data from the RITES evaluation study the presentation will share; (1) the challenges of informing stakeholders of the research on student motivation and reflection and the importance of these outcomes as a means to student achievement, (2) the statistical success of adapting student motivation surveys from the literature to a particular content (i.e., science) and the qualitative successes of analyzing student reflections, (3) the process at which teachers easily integrated the student survey and reflection with the school day, and (4) the interpretation and dissemination of evaluation findings on students' motivation and reflection to inform MSP program decisions.

- **Section 1: Questions for dialogue at the MSP LNC.**

1. How useful are standardized test scores in measuring student success in MSP programs?
 - a. How do other projects align their achievement tests with program impact measures?
 - b. Can benchmarks of student success be appropriately aligned with standardized test scores?
2. What are useful strategies to achieve stakeholder buy-in for student measures and benchmarks to student achievement?
 - a. How much does stakeholder buy-in matter for measuring student outcomes?
 - b. Does stakeholder buy-in affect the usefulness of the data collected?
3. What does student motivation and reflection tell us about academic achievement?
 - a. How can the relationship between motivation and academic achievement be leveraged for program evaluation?
 - b. Are these suitable measures of student success?
 - c. What does a relationship between a program's implementation and student reflection, or a relationship between student reflection and student motivation tell us about program implementation and impact?

- **Section 2: Conceptual framework.**

The Education Alliance, the affiliation of the authors of this presentation, is the external evaluator on the Rhode Island Technology Enhanced Science (RITES). The RITES program has

the goals of engaging teachers and students in inquiry-based approaches to learning science, bringing cutting-edge research into the classroom, and improving the understanding of the relevance of science. Program stakeholders are ultimately interested in the impact of the RITES program on student performance in science; however, identifying these outcomes measures has been a methodological challenge.

The RITES project primarily defines student success as a score of proficiency on the New England Common Assessments Program (NECAP) science state assessment. At the time the RITES program applied for NSF funds, the science portion of the assessment had not been administered to students in Rhode Island. However, based on the NECAP mathematics assessment, RITES believed that the program would result in 50% of all students and student sub-groups achieving a proficient level in science across participating schools in five years. The RITES project expects that these gains will also be reflected in increased in the number and diversity of students who pursue STEM coursework and careers.

The project's definition of student success posed several methodological challenges for evaluators. The NECAP is geared toward federal accountability standards, rather than designed for evaluation use. Therefore, the tests are *not* administered at every grade level *nor* at one time point. For example, the science test is administered in May for fourth, eighth and eleventh graders, while the Reading and Mathematics tests are administered in October for grades three through eight and eleven. For programs that are implemented in middle and high school, NECAP data does not provide the level of analysis needed to accurately identify student level program impacts due to RITES participation. Additionally, the test is developed to assess students' proficiency level; these tests are *not* scored in an appropriate way for student scores to be used as a measure of changes in content knowledge as a result of program implementation.

Furthermore, teachers who participate in RITES span grade levels (5th through 12th grades), content areas (from biology to botany), and course tracks (from remediation to advanced placement). Given the context of the state achievement test administration and the variation of students who receive programming, the evaluation design required flexible measures that could be compared across various science courses and grade levels. Evaluators developed a model to examine student success (described below); however, this model continues to be adapted as program implementation becomes more solidified.

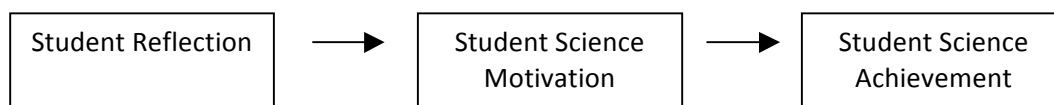
Student Motivation and Reflection: Instrument Development

Grounded in research that suggests that student motivation is a precursor to academic achievement, evaluators proposed measuring students' motivation toward science as an outcome associated with RITES program implementation. While several studies (Gibson & Chase, 2002; Haladyna, Olsen, & Shaughnessy, 1982; Napier & Riley, 1985; Thompson & Mitzes, 2002; Tuan, Chin, & Shieh, 2005) provided the foundation of a student motivation survey the items were adapted to reflect specific areas of program implementation. However, in efforts to maintain the reliability of the measures, much of the survey remained as implemented in previous studies. Tuan and colleagues (2005) developed six scales measuring students' motivation towards science learning, including: self-efficacy, active learning strategies, science learning value, performance goal, achievement goal, and learning environment stimulation. With

the exceptions of achievement goal and learning environment stimulation, which were not used (due to lack of relevance to the program and survey length), scales were minimally adapted for evaluation use.

Additionally, it became evident to evaluators that the RITES program was requiring a shift in teacher and student thinking around science, where classroom activities promoted students to “act and think like scientists.” Therefore, measures of reflection on how students interacted with science as a scientist were added by evaluators to offer valuable insight into short-term program impacts. Student reflection prompts were developed to be distributed by teachers directly following a program’s lesson or lab implementation, thereby capturing immediate feedback in response to RITES implementation. Development of the student reflection prompts were based on research examining strategies to promote productive student reflection and knowledge integration in science classrooms (Davis, 2003).

Based on published research both on student reflection prompts and metacognition (e.g. Davis, 2003; McMillan & Hearn, 2008), evaluators hypothesized a trajectory or benchmarking of student outcome measures as follows:



This model of benchmarking is intended to adapt to several methodological challenges that accompany a program like RITES, where state-wide implementation is expected on a rolling participation basis, and where teachers and students who participate make up several grade levels and science content areas. That is, traditional research models (cohort designs, or comparison groups) are not feasible within the context of RITES; however generating a benchmarking model allows evaluators to examine the ongoing progress the program makes as implementation builds.

Student Motivation and Reflection: Stakeholder Buy-in

Despite the alignment of student motivation to achievement, teacher buy-in was not immediate. Teacher buy-in was necessary due to the need for teachers to implement the surveys with their science students state-wide. Early on evaluators noticed that particular schools were not implementing the surveys with their students. Communications with teachers indicated that several teachers supported the widely held view that students, in general, are not interested in science. Many of these teachers thought there was no need to measure motivation because it was already known to be low. And others believed that the instrument would reflect on teacher instruction rather than student motivation. Evaluators shared literature with teachers as well as offered opportunity for questions to build understanding of measures and use of data.

In terms of the student reflection instrument, teachers were asked to implement this following a RITES investigation, or classroom lesson. This procedural strategy was helpful to gaining buy-in from teachers because teachers viewed this as feedback on how the RITES materials impacted student thinking. The drawback was that teachers often did not perceive student reflection as an important component of their lesson and therefore did not allow enough time for students to fully

complete or think through the reflection prompts. In observations, evaluators observed this instrument implemented similarly to a lab worksheet or satisfaction survey where students were asked to quickly complete a task, rather than critically think about their learning. This administration process for the student reflection instrument became a limitation to the data collection (noted below).

- **Section 3: Explanatory framework.**

Four variables were generated with high reliability (Cronbach's Alphas from .80 to .90; Table 1), including: science self-efficacy, active science learning strategies, science learning value, and science performance goal. Variables proved to be normally distributed and sample size was sufficient (N = 1412). Of note are the high means of each motivation variable representing a scale of one to four. That is, students reported high levels of student motivation toward science across all motivational variables. These data indicate that improvements in student motivation over the course of RITES implementation will be a challenging endeavor due to ceiling effects. These data also provide evidence counter to teachers' perspectives that the student population generally has low motivation toward science.

Table 1. Student Motivation Variable Characteristics

<i>Student Motivation Factors</i>	<i>Alpha</i>	<i>Mean Scale 1-4</i>
Science Self-Efficacy	.84	3.90
Active Science Learning	.81	3.76
Science Value	.80	3.74
Science Performance Goal	.63	4.00
Science Engagement	.84	3.91

N = 1412

Across the eight observations conducted by evaluators, 65 student reflections were analyzed. The student reflection prompts asked students to comment on how they “thought” and “acted” like a research scientist in completing the lab activities. Coding and analysis of the student reflections drew from the Depth of Knowledge (DOK) constructs used to develop the state science assessments (Hess, 2007). The DOK framework is based on four levels of classification: (1) recall and reproduction, (2) skills and concepts, (3) strategic thinking, and (4) extended thinking. Two coders from the evaluation team read the reflections and compared their results together to discuss and negotiate any scoring differences and produce a final aligned analysis. Each question was coded as a single unit to characterize the reflection and provide evidence of knowledge integration and productive reflection.

When asked what students liked most about their science lesson, the majority of high school students referenced conducting a hands-on experiment and watching the results. Some students said they had never worked with probes before and liked using new and different equipment. Most students at the middle school level said they liked that the lesson was about dragons. They also mentioned that they liked working with a partner. In contrast, when students were asked what they liked least about their lab activity, a few students in the Chemical Reactions lesson responded that creating a graph was difficult because they did not understand the directions.

Other students wrote that they were frustrated by slow computers or working on computers that “froze.” A few of the middle school students stated that some of the questions were “hard” or “confusing.” Across investigations, several high school students wrote that they thought there were too many questions before the activity started.

Student responses to the satisfaction questions received satisfaction-type responses. Although these data were not useful in determining student success, they provided insights into ways in which the program’s investigations could be improved upon. However, two student reflection questions were intended to generate student responses that could be coded for depth of student knowledge, including:

- In what ways did you act like a research scientist during this lesson?
- What would be the next step if you wanted to investigate this topic further?

Student responses to these questions did not result in common or code-able answers. There were several reasons for this. Often the RITES investigation was not complete at the end of the class and students were rushed, resulting in superficial reflective responses. Evaluators have implemented strategies for following data collection activities to not suffer from the same administration limitations.

In these two cases, the student measures were more useful in providing formative feedback on implementation than on program impacts (although long term impacts have yet to be measured). For example, as a baseline, the student motivation data provides a valuable tool to unpack teachers’ beliefs of student motivation and engagement with science. These data strongly suggest that students are motivated toward science even before high levels of program participation occur. Similarly, the student reflection data suggests that students were not frequently asked how they thought or acted like a scientist. These prompts were foreign to students, even after participating in a RITES investigation. Formatively, these data suggests that greater level of inquiry needed to be present within the investigation as well as within classroom instruction.