AIR
AMERICAN INSTITUTES FOR RESEARCH®

Study of Physical Science and Engineering Invention Kit Curriculum for Middle School

External Evaluation of the Investing in Innovation Central Virginia Advanced Manufacturing Development Grant 78

JULY 2019

Christina LiCalsi | Kelly Reese | Dionisio Garcia-Piriz

MAKING RESEARCH **RELEVANT**

Study of Physical Science and Engineering Invention Kit Curriculum for Middle School

External Evaluation of the Investing in Innovation Central Virginia Advanced Manufacturing Development Grant 78

JULY 2019

Christina LiCalsi | Kelly Reese | Dionisio Garcia-Piriz

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

10430_12/19

Contents

Page

Summary ... 1

Introduction ... 2

Invention Kit Overview ... 3
- Invention Kit Development Process ... 3
- Invention Kit Structure and Contents ... 4
- Implementation ... 6

Study Design and Methods ... 7
- Measuring Implementation Fidelity ... 8
- Assessment of Physical Science and Engineering Knowledge ... 9
- Assessment of STEM Interest and Confidence ... 10
- Data Collection for Impact Analyses ... 10
- Analytic Sample for Impact Analyses ... 12
- Methodology ... 15

Study Findings ... 20
- Implementation Fidelity ... 20
- Physical Science and Engineering Knowledge ... 22
- STEM Interest and Confidence ... 24

References ... 26

Appendix A. National Evaluation of i3 Implementation Reporting ... 27

Appendix B. National Evaluation of i3 Impact Reporting ... 31

Appendix C. Psychometric Analysis and Scaling of Outcome Measures ... 39

Exhibits

Page

Exhibit 1. Logic Model of Intervention 5

Exhibit 2. Engineering Elective Courses Using Invention Kits in 2017–18 7

Exhibit 3. Districts and Schools Participating in the Impact Study, by Condition 8

Exhibit 4. Assessment Response Rates by School and Condition 11

Exhibit 5. Survey Response Rates by School and Condition 12

Exhibit 6. Characteristics of Students in Study Schools in Fall 2017–18, by School and Condition 14

Exhibit 7. Characteristics of Students in Impact Analyses, by School 15

Exhibit 8. Balance on Treatment and Comparison Groups in Student Characteristics, Assessment 18

Exhibit 9. Balance on Treatment and Comparison Groups in Student Characteristics, Survey 19

Exhibit 10. Key Components to Measure Implementation Fidelity 20

Exhibit 11. Implementation of Solenoid Kit Components Reported by Engineering Elective Teachers 21

Exhibit 12. Implementation of Linear Motor Kit Components Reported by Engineering Elective Teachers 22

Exhibit 13. Implementation of the Linear Generator Kit Components Reported by Engineering Elective Teachers 22

Exhibit 14. Estimated Treatment Impacts on Students' Engineering and Physical Science Knowledge 23

Exhibit 15. Pre- and Posttest Average Scaled Scores for Students in the Treatment and Comparison Groups: Science Knowledge Assessment 24

Exhibit 16. Estimated Treatment Impacts on Students' Interest and Confidence in STEM Learning (STEM Attitudes) 24

Exhibit 17. Pre- and Posttest Average Scaled Scores for Students in the Treatment and Comparison Groups: STEM Survey Attitudes 25

Exhibit A1. Fidelity of Implementation Indicator Table, Development Grant 78, Updated in 2018 27

Exhibit B1. Impact Results of the Invention Kits on Students' Science Knowledge Achievement (Confirmatory Analysis) 31

Exhibit B2. Impact Results of the Invention Kits on Students' STEM Attitudes (Confirmatory Analysis) 32

Exhibit B3. Impact Results of the Invention Kits on Students' Math Attitudes (Exploratory Analysis) 33

Exhibit B4. Impact Results of the Invention Kits on Students' Science Attitudes (Exploratory Analysis) ... 34

Exhibit B5. Impact Results of the Invention Kits on Students' Engineering and Technology Attitudes (Exploratory Analysis) ... 35

Exhibit B6. Impact Results of the Invention Kits on Students' 21st-Century Learning Attitudes (Exploratory Analysis) ... 36

Exhibit B7. Contrasts for Science Knowledge and STEM Attitudes Outcomes ... 37

Exhibit B8. Impact Estimates ... 38

Exhibit B9. Baseline Equivalence of Students ... 38

Exhibit C1. Performance Criteria for Rasch Test Statistics ... 39

Exhibit C2. Reliability Statistics for the Overall S-STEM Survey and by Construct ... 41

Exhibit C3. Reliability Statistics of Assessment ... 41

Summary

Three central Virginia school districts and engineering education researchers at the University of Virginia were awarded an Investing in Innovation development grant to design, implement, test, and nationally disseminate a project-based engineering curriculum for middle school students. Referred to as invention kits, the curriculum is developed to teach key science and engineering principles and related skills to Grades 7 or 8 students by constructing modern interpretations of 19th-century inventions that sparked industrial activity within society: the solenoid, the linear motor, and the linear generator.

The American Institutes for Research (AIR) is the external evaluator of the grant. As part of the evaluation, AIR conducted an impact study to assess the invention kits' effect on students' engineering and physical science knowledge, as well as students' interest and confidence in STEM learning. The study used a quasi-experimental comparison group design investigating differences in student pre- and-posttests during the 2017–18 school year. Students in four schools across the three districts used a set of three invention kits in their engineering electives, as compared with students taking engineering electives in three schools within one district that had business-as-usual engineering curriculum. AIR studied implementation of the kits by collecting data reported by teachers on student use of kit components, interviews with teachers on how kits were incorporated into their engineering elective curriculum and adapted for use with their students, and observations of kits in use during site visits.

The results of the study as defined by the research questions are as follows:

Research Question 1: *Do students in Grades 7 or 8 who receive the intervention (i.e., construct three or more kits in an engineering class) have greater science knowledge compared with Grades 7 or 8 students in the business-as-usual condition (i.e., take engineering but do not construct or use kits)?*

The research team did not find a statistically significant difference between the physical science and engineering assessment scores of students who used the kits and comparison students.

Research Question 2: *Do students in Grades 7 or 8 who receive the intervention (i.e., construct three or more kits in an engineering class) have greater STEM interest and confidence compared with Grades 7 or 8 students in the business-as-usual condition (i.e., take engineering but do not construct or use kits)?*

The research team did not find a statistically significant difference between the measures of STEM interest and confidence of students who used the kits and comparison students.

Research Question 3: *Were the three invention kits identified by developers implemented by all engineering elective teachers with at least 75% of their students, using at least 60% of kit components?*

The research team found the answer was no. Teachers and students in two of the four schools in the treatment group implemented the three invention kits with fidelity. Only one of the three kits (Solenoid) was implemented with fidelity by all four participating schools.

Introduction

In late 2014, three central Virginia school districts—Albemarle County Public Schools (ACPS), Charlottesville City Schools, and Fluvanna County Schools—and engineering education researchers at the University of Virginia (UVA) were awarded an Investing in Innovation (i3) development grant to design, implement, test, and nationally disseminate a project-based engineering curriculum for middle school students. Referred to as invention kits, they are developed to teach key science and engineering principles and related skills to Grades 7 or 8 students by constructing modern interpretations of 19th-century inventions that sparked industrial activity within society: the solenoid, the linear motor, and the linear generator.

The American Institutes for Research (AIR) is the external evaluator of the grant. As part of the evaluation, AIR conducted an impact study to assess the invention kits' effect on students' engineering and physical science knowledge, as well as students' interest and confidence in STEM learning. The study examines these two main domains through a quasi-experimental comparison group design investigating differences in student pre- and posttests during the 2017–18 school year. Students in four schools across the three districts used a set of three invention kits in their engineering electives as compared with students taking engineering electives in three schools within ACPS that had business-as-usual engineering curriculum. AIR studied implementation of the kits by collecting data reported by teachers on student use of kit components, interviews with teachers on how kits were incorporated into their engineering elective curriculum and adapted for use with their students, and observations of kits in use during site visits.

This report describes the impact and implementation study conducted by AIR and its findings. We first provide an overview of the structure, development process, and implementation of the invention kits, followed by detail of the study design and methods used. Finally, the report includes an assessment of whether the kits were implemented with fidelity and the effect of invention kit use on student engineering and physical science knowledge and STEM interest and confidence.

Invention Kit Overview

The concept for the invention kits was originally developed through a partnership with the University of Virginia (Professors Glen Bull and Joe Garofalo), Princeton University (Professors Michael Littman and David Billington), and the Smithsonian Institute. Within its archives, the Smithsonian houses many original American inventions that transformed the American economy and society, a few of which are the focus of the kits. In the process of looking for ways to prioritize digital preservation of its archives, Smithsonian staff worked with content experts to determine possible uses of the digital replicas. Several of the kits that deal with pivotal 19th-century inventions are intended to be open source material and accessible from the Smithsonian in conjunction with its 3D replicas. An intention of the invention kit curriculum is that students will analyze the historical context and cultural significance of the inventions and be inspired by the power of new ideas within science and engineering to transform human life. The logic model in Exhibit 1 was created by AIR evaluators after interviews with invention kit developers and examination of kit materials. It displays the developers' intention for student interaction with the kits and the anticipated short- and long-term effects for students in their science and engineering learning.

Invention Kit Development Process

Developers at the University of Virginia (UVA) have been working on similar curriculum development for a number of years, and invention kits that introduce a variety of concepts in science and engineering, outside of those linked to 3D replicas in the Smithsonian archives, have been created.

A form of the intervention was first piloted in 2013 through collaboration between UVA, the Smithsonian, and Princeton University. In the initial pilot, students designed and manufactured a working reinterpretation of the Morse-Vail telegraph system using 3D printers, objects from the Smithsonian, and Vail's journals. The success of the pilot project inspired collaborators to develop the Summer Engineering Design Academy at the Laboratory School for Advanced Manufacturing, which is located in Sutherland Middle School in ACPS. Six teachers were trained on the historical reconstruction teaching methodology and worked with 12 students in the academy for two weeks.

In the 2014–15 school year, after the i3 grant award, Grades 7 and 8 students in the two partner laboratory schools that have an ongoing relationship with UVA—Sutherland Middle School (in ACPS) and Buford Middle School (in Charlottesville City Schools)—were exposed to invention kits in engineering and physical science classrooms while developers and high school educators worked to test and refine kit components and instructional approaches. In July 2015, physical science and engineering educators from the lab schools again worked with UVA developers and students entering eighth grade in the Summer Engineering Design Academy to further develop the content and pedagogical strategy included within the invention kits.

For the purposes of this grant, six new invention kits were slated to be developed and refined. The project leadership intended teachers and students to implement a minimum of three invention kits in their engineering science electives in the 2017–18 school year: (a) **the Solenoid,** (b) **the Linear Motor,** and (c) **the Linear Generator Invention kits.** Three additional kits were developed as part of the grant: (d) the Ammeter; (e) the Telegraph; and (f) Telephone/Speaker. Teachers in Sutherland Middle School and

Buford Middle School participated in the development of all kits during the grant period. The Summer Engineering Design Academy continued for the duration of the grant, and kits were tested and refined during this time as well. However, kit materials beyond the three core kits—solenoid, linear motor, and linear generator—were not available to the group of students participating in the impact testing in the 2017–18 school year, and these students did not have access to the kits prior to that school year.

After the first year of development under the grant, school districts and project leadership agreed that sufficient implementation of the kit curriculum would consist of teacher and student use of the first three kits in their engineering elective courses. In one school (Buford Middle School) students also used the kits with science teachers in their physical science classrooms through collaboration with the engineering instructor, and this use was tracked by AIR researchers. These three kits are **intended to function sequentially in concept and skill introduction** and act as a cohesive unit to build a foundation of student understanding of the concepts of **magnetism, electricity, and electromagnetism,** as well as skills related to computer-aided design, computer-aided manufacturing, basic maker-skills such as soldering, and scientific observation and process enactment.

Invention Kit Structure and Contents

The key components of the kits are identified in the logic model (Exhibit 1), listed as A–E:

A. *Make activity* instructions (via picture, video, written text) for students to build items (e.g., continuity tester, solenoid) to then use within labs.

B. *Lab activity* instructions for students, including material lists, step-by-step written guides, associated actions, and guiding questions to demonstrate principles within a lab.

C. *Design challenge* instructions for students to demonstrate their understanding of the concepts in practical, open-ended exercises in which they create a functioning object using the principles and maker-skills they have learned.

D. *Electronic computer-aided design (CAD) files* and/or electronic files and instructions for incorporating student-created invention components in 2D and 3D fabrication technologies (i.e., digital die cutters, 3D printers).

E. *Teacher guides* with information on related concepts, standards, material sourcing, associated student skills, safety, and assessment.

Exhibit 1. Logic Model of Intervention

INPUTS

1. Program Provision of Kits
Engineering teachers are provided with invention kits: a set of interdisciplinary curriculum materials incorporating engineering, physical science, mathematics, maker activities, and a description of the historical context of the pivotal invention. Kits have some or all of the following components:

A. Make activity instructions (via picture, video, written text) for students to build items (e.g., continuity tester, solenoid) to then use within labs.
B. Lab activity instructions for students, including material lists, step-by-step written guides, associated actions, and guiding questions to demonstrate principles within a lab.
C. Design challenge instructions for students to demonstrate their understanding of the concepts in practical, open-ended exercises.
D. Electronic CAD files and/or electronic files and instructions for incorporating student-created invention components in 2D and 3D fabrication technologies (i.e., digital die cutters and 3D printers).
E. Teacher guides with information on related concepts, standards, material sourcing, associated student skills, safety, and assessment.

ACTIVITIES

2. Teacher Use of Kits

A. Teachers accessed components A–E as a package for each kit on the maketolearn.org website.
B. Teachers incorporated the kits within their engineering elective courses for 7th or 8th graders.
C. Teachers used three fully developed invention kits with their students in sequential order: **(1)** the Solenoid kit, **(2)** Linear Motor kit, and **(3)** the Linear Generator kit.

3. Student Use of Kits

A. Students are exposed to the solenoid, linear motor, and linear generator kits in their engineering elective course.
B. Within the same academic year (2017–18), students take physical science. Physical science teachers are aware of the kit curriculum and reinforce concepts when appropriate.

MEDIATORS

Teachers' instructions of core engineering and physical science principles are strengthened.

Teachers' instructions to students are enhanced through interactive components of the kits.

Students' learning of engineering and science concepts and skills are enhanced by awareness of social contexts and cultural impacts of scientific invention.

Students engage in hands-on learning experiences.

Students are more interested in and motivated by engineering and science learning and more concretely understand what it may mean to pursue further coursework or careers in the space.

OUTCOMES

Proximal

Students who learned key concepts and skills through the invention kit construction and use in engineering courses perform better on assessments of their engineering and science knowledge and skills. ***(i3 Study Outcome Measure 1)***

In addition, these students report higher engagement with and confidence in learning STEM subjects. ***(i3 Study Outcome Measure 2)***

Distal

Students better retain:

- **Fundamental knowledge** in engineering, science, and math.
- **Skills** related to engineering design; construction; and problem solving, including ability to transfer knowledge, and to assess sequential events and historical impacts.
- **Confidence** in their learning abilities in these areas.

With these skills and knowledge, students are more likely to **pursue**—and be successful in—**advanced** engineering, science, and math **coursework**.

Students are more likely to pursue advanced manufacturing and engineering **careers** and are better prepared and, therefore, successful, when they do.

The kits come with teacher guides that provide an overview of activities for students and essential questions and key concepts and standards aligned to the activities. Suggestions for activities or resources to introduce key scientific concepts are sometimes included. A materials list is included with information on directly sourcing the necessary components to implement the kits in the lab. No physical materials are intended to accompany the kits outright and must be purchased or made by educators and students. Step-by-step instructions, often with pictures or short videos, are included for each sequential lab activity. Lab activities also include guiding questions for students. “Make” activities are identified separately, where students are intended to construct components (e.g., continuity tester, solenoid) that are then used to enact future lab activities. The kits incorporate features such as 2D and 3D fabrication and printing technology and electronic CAD files. As a part of the grant, the equipment necessary to use this technology (i.e., laser cutters, 3D printers) was purchased for the middle school engineering labs, and teachers and students received training on the equipment.

Finally, each kit includes at least one “design challenge” or “invent” activity. Students are intended to apply their understanding of the learned concepts and skills to create a functioning object, such as an articulated figure that moves and completes a task. Although presented sequentially with concepts and skills that build throughout, the labs, make activities, and invent activities or design challenges are intended to be adaptable for teachers and students who may be using the kit with different points of entry and prior knowledge in these areas. CAD files are presented in such a way that teachers and students can have entry-level knowledge of the program and still be able to use the files in the activities as intended.

Implementation

During the 2017–18 implementation of kits 1, 2, and 3 for the impact study, teachers and students accessed kit materials as a set from a UVA-maintained website.[1] Teachers used the kits with their students in a variety of engineering electives, which were either partial-year and or full-year experiences. All students included in the sample also took physical science during the 2017–18 school year and were expected to be learning the applicable aligned Virginia state science standards. Students included in the 2017–18 sample were in Grades 7 or 8 and had not previously used the invention kits in science or engineering courses. To monitor use of the kits at the teacher and student levels, AIR researchers gathered student rosters for all engineering courses directly from the school districts and verified student rosters with principals, teachers, and testing coordinators as a part of pre- and posttests for the study overseen by AIR. AIR confirmed that students were simultaneously enrolled in engineering electives and physical science courses. Exhibit 2 lists the names of the engineering elective courses within which the invention kits were embedded. In total, the kits were used with more than 300 middle school students across the four schools in their engineering electives.

[1] Make to Learn, retrieved from www.maketolearn.org.

Exhibit 2. Engineering Elective Courses Using Invention Kits in 2017–18

Course name	School	Course length	Number of students
Foundations of Engineering, Grade 7	Buford Middle School	Semester	90 *40, S1* *50, S2*
Engineering: Design and Build, Grade 8	Burley Middle School	Full year	37
Inventions and Innovations, Grade 8	Fluvanna High School	Quarter	89 *32, Q1* *45, Q3* *12, Q4*
Engineering, Grade 8	Sutherland Middle School	Full year	74
Mechatronics, Grade 8	Sutherland Middle School	Full year	28
		Total students	**318**

Study Design and Methods

AIR's impact study was designed to assess the effect of the invention kits on student engineering and physical science knowledge, as well as on students' interest and confidence in STEM learning. To measure engineering and physical science knowledge, AIR developed a pretest and posttest of 20 multiple choice and seven constructed response items aligned to specific Middle School Physical Science and Engineering Next Generation Science Standards. The standards—chosen by development team staff, including school district and UVA invention kit developers—were those that are applicable to the learning that should take place through kit use, but that, regardless, all middle school students should learn during the course of physical science and engineering curriculum. Assessment items were piloted and analyzed for reliability in May 2017. To measure STEM interest and confidence, AIR administered a pretest and posttest of a portion of the Student Attitudes toward STEM (S-STEM), a previously validated student survey, from which students are asked to rate agreement on a 5-point scale for 37 short statements related to STEM learning.

The study used a difference-in-difference quasi-experimental design (outcomes measured pretest to posttest in the 2017–18 school year) with treatment or comparison status assigned by the program staff at the school level. The seven schools included in the study are listed in Exhibit 3.

Exhibit 3. Districts and Schools Participating in the Impact Study, by Condition

District	School	Condition
Charlottesville City Schools	Buford Middle School[a]	Treatment
Fluvanna County Schools	Fluvanna County High School (Grades 8–12)	Treatment
Albemarle County Public Schools	Mortimer Y. Sutherland Middle School[a]	Treatment
Albemarle County Public Schools	Jackson P. Burley Middle School	Treatment
Albemarle County Public Schools	Joseph T. Henley Middle School	Comparison
Albemarle County Public Schools	Jack Jouett Middle School	Comparison
Albemarle County Public Schools	Leslie H. Walton Middle School	Comparison

Note. [a] Laboratory schools.

The impact study had two research questions:

1. Do students in Grades 7 or 8 who receive the intervention (i.e., construct three or more kits in an engineering class) have greater science knowledge compared with Grades 7 or 8 students in the business-as-usual condition (i.e., take engineering but do not construct or use kits)?
2. Do students in Grades 7 or 8 who receive the intervention (i.e., construct three or more kits in an engineering class) have greater STEM interest and confidence compared with Grades 7 or 8 students in the business-as-usual condition (i.e., take engineering but do not construct or use kits)?

The implementation study had one research question, defined by the threshold of implementation that was determined to be adequate by program staff:

3. Were the three invention kits identified by developers implemented by all engineering elective teachers with at least 75% of their students, using at least 60% of kit components?

Within this section, we describe how implementation fidelity was measured, how student physical science and engineering knowledge and STEM interest and confidence were measured, how data were collected for the impact study, how the analytic sample was constructed, and the methodology that was used to assess impacts on the student outcomes.

Measuring Implementation Fidelity

As previously mentioned, AIR researchers worked with each school district to gather course roster data for all students enrolled in engineering elective courses within the year that they also were enrolled in physical science. After receiving the data from district staff, AIR then confirmed student enrollment with principals and teachers in treatment and comparison schools to account for adds and drops of the courses. In Buford Middle School, this meant seventh graders were the target sample because physical science is offered in Grade 7 in Charlottesville City Schools. In the remaining treatment and comparison schools, physical science is offered in Grade 8, so the sample consisted of eighth-grade students.

Monitoring Kit Implementation. AIR implemented a Google form and walked teachers through how to complete the form in person or via webinar prior to the start of the 2017–18 school year. The form asked teachers to track their use of the invention kits during the 2017–18 school year and complete one form each time a separate invention kit use was completed with students.

The form asked teachers to report details including: which kit was used; what components of the kits were used (i.e., lab, make, and invent activities); in which courses and sections the teacher used the kit and over what specific time period; how much in-class time or out-of-class time the students dedicated to completing the kit; which, if any, students were chronically absent during the window of kit use; whether students experienced any specific difficulties with kit components; whether teachers altered the components of the kit they used, and if so how; and whether the teachers supplemented kit use with any additional instructional materials or activities for students to complete the kit and gain the key concepts intended to be conveyed as a part of the lesson or unit.

Teacher Interviews on Kit Use. AIR researchers first conducted telephone interviews in spring 2017 with four engineering and two physical science teachers who piloted the invention kits during the 2016–17 school year to gather their formative feedback on kit use so far. AIR researchers had interacted with nearly all the teachers previously in person during observations of kit use that took place in school classrooms, during sessions of the Summer Engineering Design Academy for a small group of students not in subsequent samples, and in public demonstrations of the kits at events associated with the grant. AIR researchers communicated with teachers at least quarterly during the 2017–18 school year, both to facilitate pre- and posttests for the impact study and to check in on implementation of the kits. Finally, telephone interviews were conducted, recorded, and professionally transcribed in late May and early June 2018 after the four engineering and two physical science teachers had concluded their use of the invention kits for the grant period. An AIR researcher first reviewed the forms submitted by the teachers on invention kit use throughout the school year and used the interviews as an opportunity to ask probing or follow-up questions for clarification regarding their responses and to give them an opportunity to elaborate further on their responses through the conversation.

This data collection enabled us to determine how many kits were implemented in each course between students' pretests and posttests, and with which students. It allowed us to determine the kit components used by teachers and students and in which contexts, and where and how adaptations were made.

Assessment of Physical Science and Engineering Knowledge

To measure engineering and physical science knowledge, AIR oversaw development of a pretest and posttest of 20 multiple choice and seven constructed response items aligned to specific Middle School Physical Science and Engineering Next Generation Science Standards. The science knowledge assessment items were developed by a test development vendor, with content collaboration from grantee partners managed by AIR. As noted in Study Designs and Methods, the standards were chosen by development team staff and applicable to the learning that should take place for students using the kit use but also for all middle school students in physical science courses and engineering curriculum.

Standard categories included forces and interactions, energy, waves and electromagnetic radiation, and engineering design. Assessment items were piloted and analyzed for reliability in May 2017. A continuous raw score was scaled using Rasch analysis (based on both the treatment and comparison students) and then standardized using the mean and standard deviations from the comparison group to provide estimates that are more easily interpreted (i.e., estimates are measured in standard deviations).[2]

Assessment of STEM Interest and Confidence

To measure STEM (science, technology, engineering, and mathematics) interest and confidence, AIR administered a pretest and posttest of a portion of the S-STEM, a previously validated student survey, from which students are asked to rate agreement on a 5-point scale for 37 short statements related to STEM learning. The S-STEM survey is a publicly available survey that measures students' confidence and efficacy in STEM subjects, as well as their interest in STEM subjects and careers.[3,4] This survey contains items similar to those on STEM interest/efficacy included on the Education Longitudinal Study survey of 2002—items that have been shown to predict postsecondary STEM success (Engberg & Wolniak, 2013; Riegle-Crumb & King, 2010; Wang, 2013; You, 2013). The S-STEM survey includes 37 Likert-type scale items and covers four constructs: math attitudes (eight items), science attitudes (nine items), engineering and technology attitudes (9 items), and 21st-century learning attitudes (11 items).[5] Rasch analysis was conducted to create scaled scores using both the treatment and comparison students as described in Appendix C and then standardized using the comparison group mean and standard deviation.

Data Collection for Impact Analyses

Pretests were administered electronically to students in seven Virginia middle schools during the students' first full week of school in either August or September 2017. Posttests were administered electronically to students in all schools from April 23 to 27, 2018, as amenable with all schools based on their state testing and end-of-year schedules.

Students took the online pretests and posttests under supervised testing conditions during school hours, overseen by AIR and test proctor staff. Students had up to 60 minutes to take the physical science and engineering assessment and up to 30 minutes to take the survey. Paper versions of both pretests secured by AIR were available to students if necessary because of special needs accommodations, but no students completed paper versions. Some students' accommodations required test items to be read aloud to them, either using software or by test proctor. No other resources (e.g., scrap paper or calculators) were available to students while taking the tests.

[2] See Appendix C for more details on the scaling process and standardization.
[3] For more details on the S-STEM survey, see North Carolina State University, "Maximizing the Impact of STEM Outreach," retrieved from https://miso.ncsu.edu/articles/s-stem-survey.
[4] For details on the development and validation of the S-STEM survey, see Unfried et al. (2015)
[5] The Likert-type item response scale had five categories: strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree.

A high percentage of students participated in the assessment and survey before (fall 2017) and after the intervention (spring 2018).

Overall, approximately 85% of target students in participating schools and classes took the assessment test before and after the intervention; a slightly smaller percentage (82%) took the survey in both administration periods (see Exhibits 4 and 5). By condition, there was higher participation among students in the treatment group than in the comparison group. For the assessment, approximately 91% of students in the treatment group took both pre- and posttests compared with 75% in the comparison group. For the survey, approximately 87% and 74% of students in the treatment and comparison group took the pre- and post-survey, respectively.

Exhibit 4. Assessment Response Rates by School and Condition

		Pretest		Posttest		Both pre- and posttests	
	Total	Yes	No	Yes	No	Yes	No
Treatment							
Buford	90	83	7	88	2	82	8
	%	92.22	7.78	97.78	2.22	91.11	8.89
Burley	37	30	7	36	1	29	8
	%	81.08	18.92	97.30	2.70	78.38	21.62
Fluvanna	89	85	4	88	1	84	5
	%	95.51	4.49	98.88	1.12	94.38	5.62
Sutherland	102	97	5	98	4	93	9
	%	95.10	4.90	96.08	3.92	91.18	8.82
Comparison							
Henley	55	30	25	54	1	29	26
	%	54.55	45.45	98.18	1.82	52.73	47.27
Jouett	69	63	6	66	3	60	9
	%	91.3	8.70	95.65	4.35	86.96	13.04
Walton	49	46	3	45	4	42	7
	%	93.88	6.12	91.84	8.16	85.71	14.29
Total	**491**	**434**	**57**	**475**	**16**	**419**	**72**
	%	**88.39**	**11.61**	**96.74**	**3.26**	**85.34**	**14.66**

Exhibit 5. Survey Response Rates by School and Condition

		Pretest		Posttest		Both pre- and posttests	
	Total	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
Treatment							
Buford	90	80	10	87	3	77	13
	%	88.89	11.11	96.67	3.33	85.56	14.44
Burley	37	34	3	36	1	33	4
	%	91.89	8.11	97.30	2.70	89.19	10.81
Fluvanna	89	81	8	87	2	79	10
	%	91.01	8.99	97.75	2.25	88.76	11.24
Sutherland	102	97	5	91	11	86	16
	%	95.10	4.90	89.22	10.78	84.31	15.69
Comparison							
Henley	55	30	25	53	2	29	26
	%	54.55	45.45	96.36	3.64	52.73	47.27
Jouett	69	58	11	66	3	56	13
	%	84.06	15.94	95.65	4.35	81.16	18.84
Walton	49	46	3	46	3	43	6
	%	93.88	6.12	93.88	6.12	87.76	12.24
Total	**491**	**426**	**65**	**466**	**25**	**403**	**88**
	%	**86.76**	**13.24**	**94.91**	**5.09**	**82.08**	**17.92**

For each outcome, an analytical sample was created based on participation rates before and after the intervention.

Two analytical samples were created based on the participation rates: (a) a sample composed of students taking engineering electives and physical science classes who took both pre- and post-assessment tests, and (b) a sample composed of those students who took both pre- and post-surveys. The assessment analytical sample was 419 students (288 and 131 students in the treatment and comparison group, respectively). The survey analytical sample was 403 of the same students (275 students in the treatment group and 128 students in the comparison group).

Analytic Sample for Impact Analyses

This study employed a sample of convenience, which was based on schools in neighboring districts to the laboratory schools in which the intervention could best be developed. Seven middle schools from the Virginia districts of Albemarle County, Charlottesville City, and Fluvanna County participated in the impact study (see Exhibit 3). Four schools from these three districts agreed to use the invention kits during the 2017–18 school year and composed the treatment group. All students from treatment schools enrolled in both engineering (elective course) and physical science (required course) participated in the intervention; this involved constructing the invention kits during engineering class and assisting their teacher in using them to facilitate understanding of scientific concepts in the physical science class. An additional three schools from Albemarle County were selected by the districts, in conjunction with AIR, to be part of the comparison group given their regional location and their overall similarities in demographic composition with treatment schools. All students taking engineering classes in comparison schools participated in the study; however, these classes did not have access to the invention kits and continued to use their standard science and engineering curriculum during the 2017–18 school year.

Among schools receiving the intervention, Sutherland Middle School and Buford Middle School were laboratory schools that participated in the development of the invention kits; therefore, their teachers had extensive background in and experience using them. The other two treatment schools, Fluvanna Middle School and Burley Middle School, were not involved in the development of the invention kits; as a result, teachers implementing the intervention only had access to the curriculum guides that accompany the kits.

Demographic Composition of Students in Participating Schools

Based on the fall membership counts for the 2017–18 school year, the composition of students enrolled in participating schools show similarities across treatment condition (Exhibit 6). All schools had an approximately even split in gender composition, with slightly fewer females than males in the comparison schools. Overall, students were predominantly White (38%–70% in treatment schools and 48%–89% in comparison schools). Schools with a lower percentage of White students had higher percentages of Black and Hispanic students. The percentage of students with disabilities was similar across condition (10%–16% and 11%–21% in treatment and comparison schools, respectively). There were more differences in the percentage of English language learners and economically disadvantaged students across schools, although there were schools with lower and higher numbers in both conditions.

Only students attending engineering and physical science classes participated in the study.

While the intervention occurred at the school level, not all students participated in the study and thus were not included in the impact analyses. Our analytic sample is based only on students who took engineering electives and physical science classes in either treatment and comparison schools during the 2017–18 school year and who participated in data collection activities as described in more detail in the following section.

Exhibit 6. Characteristics of Students in Study Schools in Fall 2017–18, by School and Condition

Student characteristics	Buford Middle School	Fluvanna Middle School	Sutherland Middle School	Burley Middle School	Henley Middle School	Jouett Middle School	Walton Middle School
Condition	Treatment	Treatment	Treatment	Treatment	Comparison	Comparison	Comparison
% Female	50%	51%	50%	54%	43%	47%	45%
% Black	33%	16%	8%	17%	2%	18%	11%
% Hispanic	15%	6%	6%	24%	3%	25%	11%
% White	38%	70%	70%	46%	89%	48%	71%
% Other races	14%	9%	16%	13%	5%	10%	7%
% English language learners	15%	3%	3%	18%	1%	19%	2%
% Students with disabilities	16%	11%	10%	15%	11%	16%	21%
% Economically disadvantaged	57%	36%	14%	42%	8%	50%	37%
N of students	260	322	212	190	295	187	112

Note. Fall membership reports elaborated by Virginia Department of Education. Percentages are for all students in the schools, not only our analytical sample.

There were some differences in the demographic composition of the analytical samples across schools and treatment condition.

Based on students included in both analytical samples, on average, participating students were mostly male, White, and did not have a special education or English language learner status (Exhibit 7). From a descriptive analysis, there were some differences in student demographic composition across treatment status. Comparing students in treatment and comparison schools, the percentage of female students ranged between 28% and 50% in treatment schools compared with 0% and 35% in comparison schools, with one school having no females in the analytical sample. The percentage of Black and Hispanic students ranged between 5% and 36% and 3% and 15%, respectively, in treatment schools compared with 0% and 16% and 3% and 22%, respectively, in comparison schools. Approximately 2% of students in treatment schools were English language learners and 8% had special education status compared with 6% and 14% of students, respectively, in comparison groups.

Exhibit 7. Characteristics of Students in Impact Analyses, by School

Student characteristics	Buford Middle School	Fluvanna Middle School	Sutherland Middle School	Burley Middle School	Henley Middle School	Jouett Middle School	Walton Middle School
Condition	Treatment	Treatment	Treatment	Treatment	Comparison	Comparison	Comparison
% Female	28%	37%	30%	50%	0%	29%	35%
% Black	36%	8%	5%	24%	0%	16%	4%
% Hispanic	11%	6%	3%	15%	3%	22%	9%
% White	48%	75%	72%	35%	97%	54%	78%
% Other races	5%	11%	19%	26%	0%	8%	9%
% English language learners	2%	2%	0%	9%	0%	13%	2%
% Students with disabilities	9%	4%	9%	21%	14%	16%	15%
N of students[a]	85	84	94	34	29	63	46

[a]Total number of students included in the impact analyses who participated in data collection activities in both fall 2017 and spring 2018. [b]Economically disadvantaged information from students in Albemarle County was not available to the study team.

Methodology

The hypotheses tested in this study are based on the expectation that students who participate in the unique process of interacting with the invention kits will display greater learning gains in the standards-aligned course concepts in science and engineering than students who did not learn the science and engineering concepts through the invention kits approach. It is also expected that the intervention will help to make the material generally more accessible to students to enable them to concretely see themselves with the confidence and interest to pursue additional coursework and ultimately a future career in the professional space.

The impact study used a difference-in-difference design. A difference-in-difference analysis approach was used to compare outcome scores between students in the comparison group with those in the treatment group pre- and postintervention. In a difference-in-difference approach, two (or more) groups are observed during two (or more) time periods. In our case, one of the groups had been exposed to the treatment by the second-time period (spring 2018), but not by the first (fall 2017). The second group, the comparison group, was not exposed to the treatment by either time period. To calculate the impact estimate of this intervention, the average outcome gain (from fall 2017 to spring 2018) in the comparison group was subtracted from the average gain in the treatment group. This approach thus relies on two sources of variation to inform the analyses: comparisons across groups and comparisons across time. By doing so, this design produces more robust impact estimates than a design that solely relies on change across time (e.g., pre- and postdesign) or on comparisons across groups (e.g., propensity score analysis) because it reduces biases caused by initial differences between treatment and comparison groups as well as biases from comparisons across time that could be the result of trends.

Impact Model

The following model specification was used for the two confirmatory analyses.

$$(1)\ Y_{tij} = \beta_{0ij} + \beta_1\ Treatment_j + \beta_2\ Post_{tij} + \beta_3\ Treatment_j{*}Post_{tij} + \beta_4 X_{ij} + u_j + r_{ij} + e_{tij}$$

In the equation, Y_{tij} is the student outcome measure for student *i* in school *j* at time *t*. The model includes an indicator for whether school *j* is a treatment school (*Treatment*) and for whether an outcome measure was taken during the postintervention time period (*Post*). The coefficient β_1 is therefore the difference in the pretest outcome measure for students in treatment schools compared with students in control schools, and β_2 is the pretest versus posttest difference in the outcome measure for control schools. β_3 represents the difference-in-difference estimate—the pretest and posttest difference between students in treatment schools and students in comparison schools. Therefore, β_3 is the estimated effect of the intervention. X_{ij} is a vector of student covariates, including: gender, race and ethnicity, disability, and English language learner status.[6] The student covariates are included in the model to increase the statistical precision of the impact estimates. The residuals μ_j, r_{ij}, and e_{tij} represent the random errors associated with schools, students, and time, respectively.

Baseline Equivalence

To assess whether treatment and comparison groups were similar at baseline, prior to the start of the intervention, the research team calculated effect-size differences (i.e., differences in standard deviations) for the analytic sample following the procedures of the What Works Clearinghouse (WWC, 2017a). The research team evaluated baseline equivalence of groups on both prior achievement and student demographic composition.

For continuous variables, such as pretest (fall 2017 outcomes), effect-size differences were computed using standardized mean differences (Hedges' *g,* with an adjustment for small-sample bias). These differences are defined as the difference in mean outcomes between the treatment and comparison groups, divided by the pooled within-group standard deviation of the outcome measure, as shown in the following equation.

[6] Economically disadvantaged status was not included in the model because it was available only for students in Charlottesville City and Fluvanna County; hence, all students in the comparison group were missing a value, resulting in collinearity issues.

$$g = \frac{\omega\left(y_i - y_c\right)}{\sqrt{\frac{\left(n_i - 1\right)s_i^2 + \left(n_c - 1\right)s_c^2}{n_i + n_c - 2}}}$$

Source: What Works Clearinghouse (WWC, 2017a)

Where y_i and y_c are the treatment and comparison means, respectively; n_i and n_c are the equivalent sample sizes; and S_i^2 and S_c^2 the variances. ω is a small sample size correction (see Appendix E in WWC, 2017a).

For dichotomous variables, such as gender, race, and English language learner status, Cox log odds ratios (d_{Cox}) are calculated as suggested by WWC. This is an alternative measure for binary outcomes that yields effect-size differences comparable to standardized mean differences for continuous variables. It is defined as the difference in probability of the occurrence of an event as indicated by the below equation.

$$d_{Cox} = \omega\left[ln\left(\frac{p_i}{1-p_i}\right) - ln\left(\frac{p_c}{1-p_c}\right)\right] / 1.65$$

Source: What Works Clearinghouse (WWC, 2017a)

Where p_i and p_c are the probability of being in the treatment or comparison group, respectively; and ω is a small sample size correction. To determine whether the effect-size difference was substantially important, the study also followed WWC's standards, where (a) effect sizes larger than 0.25 standard deviations (SDs) were considered to be substantively important and did not satisfy group equivalence, (b) effect-size differences larger than 0.05 and up to 0.25 SDs required statistical adjustment to satisfy equivalence, and (c) differences between 0.00 and 0.05 SDs satisfied baseline equivalence (WWC, 2017b).

For difference-in-difference analyses, however, the WWC establishes that when the baseline characteristic is the same as the outcome, a difference-in-difference adjustment may be acceptable as statistical adjustment under certain conditions (WWC, 2017b). First, the baseline and outcome measures must have the same units of measurement. In our case, this condition was met since forms of the same test[7] and identical surveys were administered in fall 2017 (baseline measure) and in spring 2018 (outcome measure). Furthermore, the same scaling procedure was used for both pre- and

[7] Twenty-three of the 27 items were identical in each form, with four items in each form exchanged and the item order changed to prevent students from taking an identical posttest.

postintervention measures, as described in Appendix C. Second, the correlation between the baseline and outcome measures need to be 0.6 or higher. This condition was met for the science knowledge assessment test, with a correlation of 0.7; however, it was not met for the survey STEM attitudes, with a correlation of 0.5.[8]

Baseline Equivalence Results

Students were similar in prior science knowledge but differed in student characteristics.

For students in the analytic sample for the assessment outcome, we did not find baseline differences in science knowledge based on the fall 2017 test (Exhibit 8). In other words, students' level of knowledge in science before participating in this intervention was similar with students in the treatment group compared with those in the comparison group. There were substantive differences, however, in many student demographics. These variables are included in all models as covariates to adjust for these differences.

Exhibit 8. Balance on Treatment and Comparison Groups in Student Characteristics, Assessment

Variable	Treatment (average)	Comparison (average)	Raw difference	Standardized difference[a]
Female (Proportion)	0.34	0.25	0.08	0.25[c]
Black (Proportion)	0.17	0.09	0.08	0.41[c]
Hispanic (Proportion)	0.07	0.13	-0.06	-0.42[c]
White (Proportion)	0.63	0.71	-0.08	-0.21[b]
Other races (Proportion)	0.13	0.07	0.06	0.44[c]
English language learners (Proportion)	0.02	0.07	-0.05	-0.86[c]
Students with disabilities (Proportion)	0.08	0.15	-0.07	-0.41[c]
Standardized pretest scaled score: Assessment test	0.00	0.00	0.00	0.00
N of students in analytic sample	288	131		

[a]Effect-size difference or standardized mean difference. [b]0.05 < effect-size difference ≤ 0.25. [c]Effect-size difference > 0.25.

There were some differences in prior STEM attitudes among students in each group.

Baseline equivalence results using the analytic sample for the survey outcome indicate that students in the comparison group had greater interest and confidence in STEM learning than students in the

[8] Correlations between pre- and post-measures of the science knowledge assessment by treatment condition were: 0.71 for students in the comparison group and 0.70 for those in the treatment group. Similarly, correlations of the STEM attitudes outcome by treatment condition were: 0.42 for the comparison group and 0.54 for the treatment group.

treatment group. The effect-size difference was 0.20 SD, which requires statistical adjustments to satisfy baseline equivalence (Exhibit 9). The student demographic differences observed in the assessment analytic sample were also found in the survey analytic sample, which is to be expected given the overlap in samples.

Exhibit 9. Balance on Treatment and Comparison Groups in Student Characteristics, Survey

Variable	Treatment (average)	Comparison (average)	Raw difference	Standardized difference[a]
Female (Proportion)	0.35	0.26	0.09	0.26[c]
Black (Proportion)	0.17	0.08	0.09	0.54[c]
Hispanic (Proportion)	0.07	0.13	-0.06	-0.40[c]
White (Proportion)	0.62	0.73	-0.11	-0.31[c]
Other races (Proportion)	0.13	0.05	0.08	0.60[c]
English language learners (Proportion)	0.02	0.06	-0.04	-0.66[c]
Students with disabilities (Proportion)	0.08	0.13	-0.06	-0.37[c]
Standardized pretest scaled score: STEM attitudes	-0.19	0.00	-0.19	-0.20[b]
N of students in analytic sample	275	128		

[a]Effect-size difference or standardized mean difference. [b]0.05 < effect-size difference ≤ 0.25. [c]Effect-size difference > 0.25.

Study Findings

Implementation Fidelity

The study met fidelity of implementation standards agreed to by all partners on only one of three key components identified.[9] The three measures of fidelity of implementation used in the study are outlined in Exhibit 10. The program provided three invention kits in an online format accessible to all treatment teachers (key component 1, met). In two of the four schools, teachers used at least 60% of the materials provided within each of the three kits (key component 2, not met). Finally, in two of the four schools, less than 75% of students in engineering elective courses used the majority of the kit components.

Exhibit 10. Key Components to Measure Implementation Fidelity

Key intervention components	Implementation measure[a]	Sample size (sample level)[b]	Component level threshold for fidelity[c]	Evaluator criteria[d]	Component level (fidelity score)[e]	Implemented with fidelity
(1) Program provision of invention kits	3	3 kits	Score of 1: kit was provided. Score of 3: all kits were provided.	Program provides 3 invention kits	3 (out of 3)	Yes
(2) Teacher use of invention kits	1	4 schools	Score of 1 for each school: engineering teacher incorporated at least 60% of the materials for all three kits in engineering courses.	4 out of 4 schools must use 3 kits	2 (out of 4)	No
(3) Student use of invention kits	1	4 schools	Score of 1 for each school: at least 75% of students in engineering courses used kits with fidelity (at least 60% of kit materials).	4 out of 4 schools must meet threshold	2 (out of 4)	No

[a]*N* of measurable indicators representing each component. [b]*N* of schools, districts, etc. [c]For the unit that is the basis for the sample level. [d]For "implemented with fidelity" at the sample level. [e]For the entire sample.

The following tables demonstrate which components of each invention kit were used by teachers in each of the engineering courses. All teachers (and 100% of students) implemented the Solenoid kit with fidelity (Exhibit 11). A majority of teachers and students did not implement the invention activity as provided but implemented another invention activity or design challenge at the conclusion of the kit.

[9] Appendix A includes a full description of the fidelity of implementation indicators and key components for NEi3 reporting.

Exhibit 11. Implementation of Solenoid Kit Components Reported by Engineering Elective Teachers

Solenoid Kit	Buford MS		Burley MS	Fluvanna HS			Sutherland MS	
	Foundations of Engineering, Grade 7		Engineering: Design and Build, Grade 8	Inventions and Innovations, Grade 8			Engin-eering, Grade 8	Mecha-tronics, Grade 8
	S1	S2	Full year	Q1	Q3	Q4	Full year	
Lab 1: Investigating Magnetism	X	X	X	X	X	X	X	X
Make Activity 1: Building a Continuity Tester	X		X	X	X	X	X	X
Lab 2: Investigating Conductivity	X	X	X	X	X	X	X	X
Lab 3: Detecting Magnetic Fields	X	X	X	X	X	X	X	X
Lab 4: Exploring Electromagnetism	X	X	X	X	X	X	X	X
Make Activity 2: Building a Solenoid	X	X	X	X	X	X	X	X
Lab 5: Investigating Solenoids	X	X	X	X	X	X	X	X
Invention Activity				X	X	X		X

Three of the four schools implemented at least 60% of the Linear Motor Kit components, and one school and engineering course did not (Exhibit 12). This translates to 72% of the 318 students in the treatment group using the Linear Motor Kit with fidelity.

Exhibit 12. Implementation of Linear Motor Kit Components Reported by Engineering Elective Teachers

Linear Motor Kit	Buford MS		Burley MS	Fluvanna HS			Sutherland MS	
	Foundations of Engineering, Grade 7		Engineering: Design and Build, Grade 8	Inventions and Innovations, Grade 8			Engin-eering, Grade 8	Mecha-tronics, Grade 8
	S1	S2	Full year	Q1	Q3	Q4	Full year	
Make Activity 1: Building the Linear Motor	X		X	X	X		X	X
Make Activity 2: Cardstock				X	X			X
Lab 1: Powering the Linear Motor			X	X	X		X	X
Lab 2: An AC Power Source			X	X	X			
Lab 3: Operating the Linear Motor with AC Power I	X		X	X	X		X	X
Lab 4: Operating the Linear Motor with AC Power II			X				X	X
Invention Activity: Articulated Figures							X	

Finally, the fewest teachers and students used the Linear Generator Kit during the 2017–18 school year (Exhibit 13). Only 139 students (44%) used the majority of the Linear Generator Kit components.

Exhibit 13. Implementation of the Linear Generator Kit Components Reported by Engineering Elective Teachers

Linear Generator Kit	Buford MS		Burley MS	Fluvanna HS			Sutherland MS	
	S1	S2	Design/ Build	Q1	Q3	Q4	Engin-eering	Mecha-tronics
Make Activity 1: Building the Linear Generator			X				X	X
Lab 1: Generating Electricity	X		X				X	X
Lab 2: Application of a Linear Generator			X				X	X
Lab 3: Visualizing a Voltage	X		X				X	X

Physical Science and Engineering Knowledge

The research team did not find an impact of the invention kits on students' science and engineering achievement. Results from the impact model suggest that the invention kits did not have a significant impact on students' engineering and physical science knowledge, with positive but statistically insignificant program impact estimates (Exhibit 14). The estimates indicate that students in the treatment group, on average, scored 0.06 SD higher on the spring 2018 assessment than students in the comparison group relative to their scores in fall 2017. This estimate, however, was statistically indistinguishable from zero.

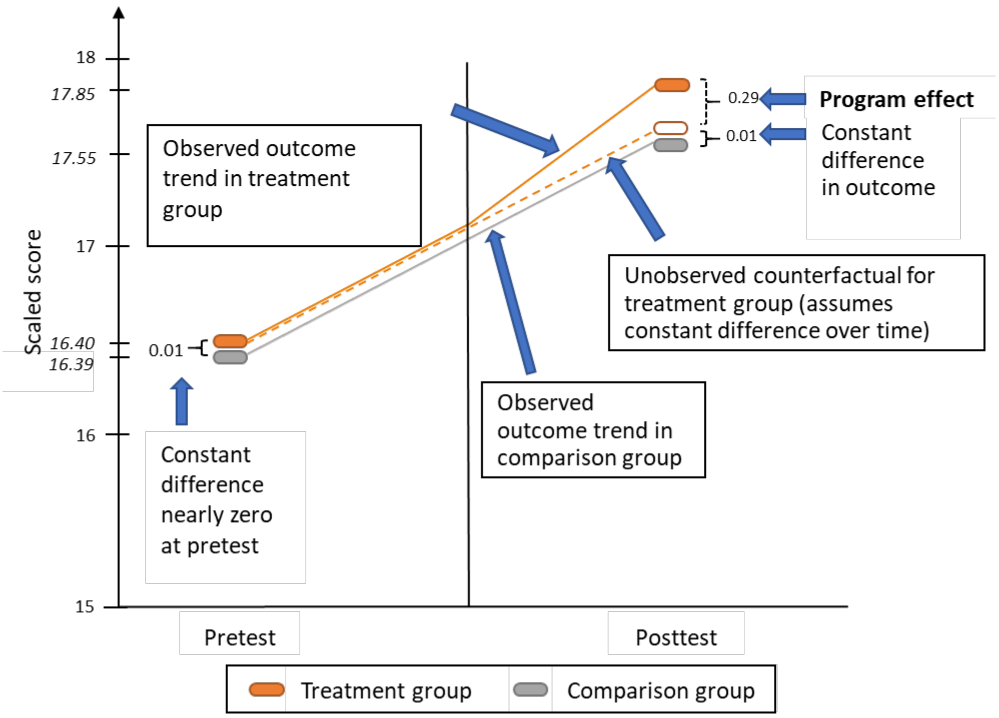
Exhibit 14. Estimated Treatment Impacts on Students' Engineering and Physical Science Knowledge

	Science knowledge
Treatment coefficient	0.06
Standard error	(0.13)
p-value	.66
N	419

The difference-in-difference observed between treatment and comparison student groups is depicted in Exhibit 15. At baseline, students in both groups had nearly identical knowledge in engineering and science based on the assessment test administered prior to the intervention (pretest). This minimal difference amounted to 0.01 SD on the scaled score. In absolute terms, students on average obtained 16.4 points out of a 34 scaled score.

After the intervention, we observed small gains in knowledge among students in both treatment and comparison groups (scoring 1.45 and 1.16 points higher on the scaled score in spring 2018 compared with fall 2017, respectively). This gain is equivalent to, on average, answering one to two additional questions correctly out of 27 total questions. The gain was larger in magnitude for students in the treatment group; however, this difference was statistically indistinguishable from zero.

Exhibit 15. Pre- and Posttest Average Scaled Scores for Students in the Treatment and Comparison Groups: Science Knowledge Assessment



Note. This figure is not drawn to scale: the difference in pretest scores (0.01 SD) is graphically disproportionally larger than the difference in program effect (0.29 SD).

STEM Interest and Confidence

The research team did not find an impact of the invention kits on student interest and confidence in STEM learning. Results from the impact model suggest that the invention kits did not have a significant impact on students' responses to the survey. Program impact estimates were positive but statistically insignificant (Exhibit 16).

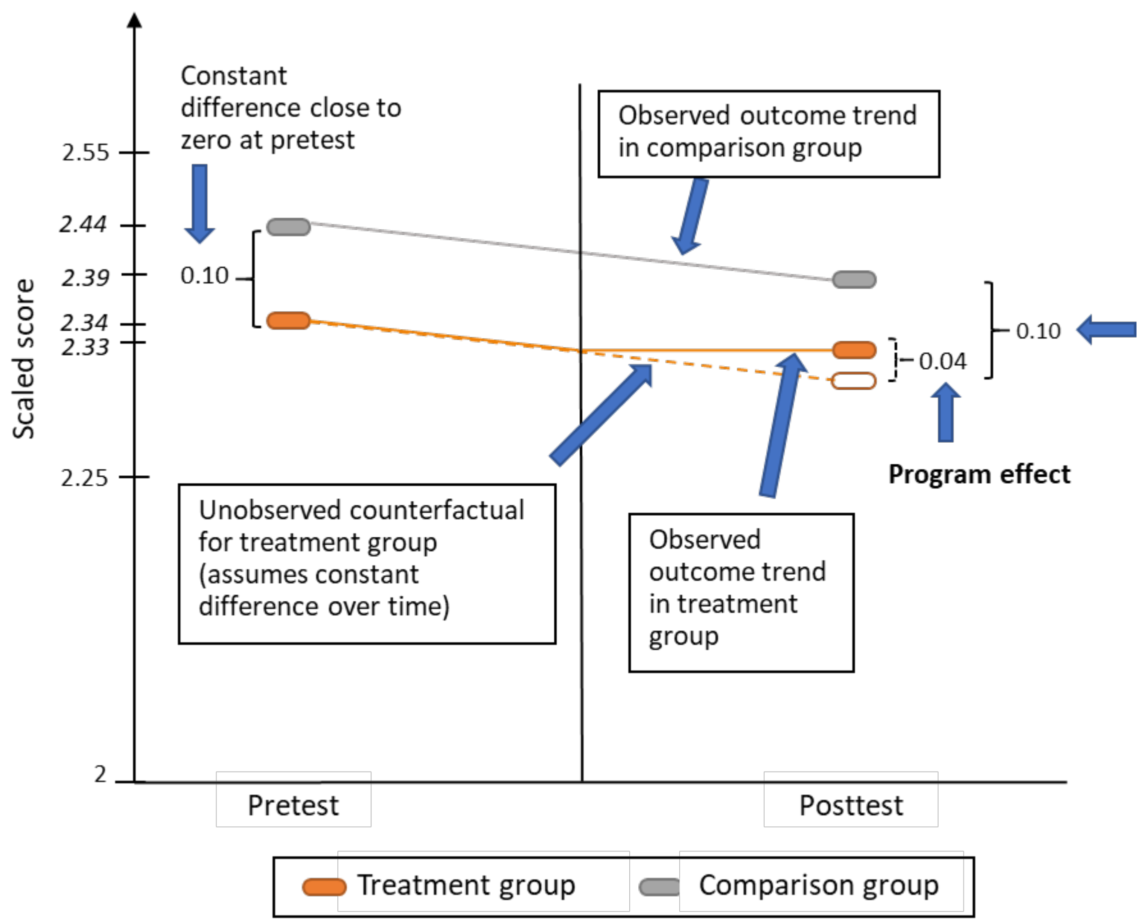
Exhibit 16. Estimated Treatment Impacts on Students' Interest and Confidence in STEM Learning (STEM Attitudes)

	STEM attitudes
Treatment coefficient	0.07
Standard error	(0.14)
p-value	.60
N	403

The average gains observed for students in the treatment and comparison groups for the STEM attitudes outcome are shown in Exhibit 17. At baseline, students in the intervention and comparison groups had similar levels of interest and confidence in STEM learning. Students in the comparison group indicated slightly higher levels than students in the treatment group, but the difference was not statistically significant (0.10 scaled score points at pretest).

After the intervention, we observed small and statistically insignificant losses in interest and confidence in STEM learning in both treatment and comparison groups (scoring 0.05 and 0.01 points lower in the scaled score during spring 2018 and fall 2017, respectively). The loss was larger in magnitude for students in the comparison group; however, this difference was statistically indistinguishable from zero. In conclusion, we did not observe any changes in students' interest and confidence in learning STEM that can be attributed to the intervention.

Exhibit 17. Pre- and Posttest Average Scaled Scores for Students in the Treatment and Comparison Groups: STEM Survey Attitudes



References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573.

Engberg & Wolniak. (2013).

Friday Institute for Educational Innovation. (2012). *Student Attitudes toward STEM Survey—Middle and high school students*. Raleigh, NC: Author.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Exp. ed.). Chicago, IL: University of Chicago Press.

Riegle-Crumb & King. (2010).

Unfried, A., Faber, M., Stanhope, D., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, mathematics, and engineering. *Journal of Psychoeducational Assessment*, 33(7), 622–639.

Wang (2013).

What Works Clearinghouse. (2017a). *What Works Clearinghouse: Standards handbook* (Version 4.0). Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

What Works Clearinghouse. (2017b). *What Works Clearinghouse: Procedures handbook* (Version 4.0). Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

Wright, B. D. (1996). Time 1 to time 2 (pre-test to post-test) comparison: Racking and stacking. *Rasch Measurement Transactions, 10*(1), 478.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement.* Chicago, IL: MESA Press.

You. (2013).

Appendix A. National Evaluation of i3 Implementation Reporting

Exhibit A1. Fidelity of Implementation Indicator Table, Development Grant 78, Updated in 2018

Indicator number	FOI indicator	Definition	Data source	Student level	Teacher level	School level	Sample level fidelity score	Sample in measurement	Years of fidelity measurement
Key Component 1 = Program Provision of Invention Kits									
1	Solenoid Invention Kit	All components of the kit, A–E as described on the logic model, are made available by program staff to be accessed online by the teachers.	AIR verification of kit availability via the website maketolearn.org; teachers in the 4 schools access the kits via the website.	N/A	N/A	N/A	Kit-level fidelity: Adequate implementation = 1. A score of 1 indicates kit is provided with components A–E. Inadequate implementation = 0. A score of 0 indicates kit is not provided with components A–E.	1 of 3 kits	1 year (2017–18): granted permission from OII due to 1 year of full program implementation.
2	Linear Motor Invention Kit	All components of the kit, A–E as described on the logic model, are made available by program staff to be accessed online by the teachers.	AIR verification of kit availability via the website maketolearn.org; teachers in the 4 schools access the kits via the website.	N/A	N/A	N/A	Kit-level fidelity: Adequate implementation = 1. A score of 1 indicates kit is provided with components A–E. Inadequate implementation = 0. A score of 0 indicates kit is not provided with components A–E.	1 of 3 kits	1 year (2017–18): granted permission from OII due to 1 year of full program implementation.

Indicator number	FOI indicator	Definition	Data source	Student level	Teacher level	School level	Sample level fidelity score	Sample in measurement	Years of fidelity measurement
3	Linear Generator Invention Kit	All components of the kit, A–E as described on the logic model, are made available by program staff to be accessed online by the teachers.	AIR verification of kit availability via the website maketolearn.org; teachers in the 4 schools access the kits via the website.	N/A	N/A	N/A	Kit-level fidelity: Adequate implementation = 1. A score of 1 indicates kit is provided with components A–E. Inadequate implementation = 0. A score of 0 indicates kit is not provided with components A–E.	1 of 3 kits	1 year (2017–18): granted permission from OII due to 1 year of full program implementation.
							Program-level fidelity: Adequate fidelity = 3. A score of 3 indicates 3 kits are provided (score of 1 each). Inadequate fidelity = < 3. A score of less than 3 indicates fewer than 3 kits are provided.		1 year (2017–18): granted permission from OII due to 1 year of full program implementation.

Indicator number	FOI indicator	Definition	Data source	Student level	Teacher level	School level	Sample level fidelity score	Sample in measurement	Years of fidelity measurement
Key Component 2 = Teacher Use of Invention Kits									
4	Teacher use of invention kits 1–3 in engineering elective, Grade 8.	Teachers report use of kits at the individual student level and are required to use all 3 kits in the engineering course.	Engineering teacher completes AIR's developed and maintained online form at kit conclusion, noting components used or alterations made.	N/A	1 = Teacher uses at least 60% of the materials in all 3 kits in each engineering course. 0 = Teacher did not use all 3 kits in each engineering course.	1 = 100% of teachers score a 1. 0 = < 100% of teachers score a 1.	Adequate Implementation = 4. A score of 4 indicates 4 schools receive a 1. Inadequate fidelity = < 4. A score of < 4 indicates fewer than 4 schools receive a 1.	All engineering teachers in 4 schools.	1 year (2017–18): granted permission from OII due to 1 year of full program implementation.
							Program-level fidelity: Adequate Implementation = 4. Inadequate fidelity = < 4.		1 year (2017–18): granted permission from OII due to 1 year of full program implementation.

Indicator number	FOI indicator	Definition	Data source	Student level	Teacher level	School level	Sample level fidelity score	Sample in measurement	Years of fidelity measurement
Key Component 3 = Student Use of Invention Kits									
5	Student exposure to each invention kit in engineering elective, Grade 8.	Teacher reports student use of kits at the individual student level.	Engineering teacher completes AIR's developed and maintained online form.	1 = Students use all 3 kits with fidelity. 0 = Students do not use each of the 3 kits with fidelity.	N/A	1 = ≥ 75% of students in each of 4 schools score a 1. 0 = < 75% of students in each of 4 schools score a 1.	Adequate Implementation = 1. A score of 1 indicates 4 schools receive a 1. Inadequate fidelity = 0. A score of 0 indicates < 4 schools receive a 1.	All students enroll in engineering electives (Grades 7 and 8) and enroll in physical science in the same year.	1 year (2017–18): granted permission from OII due to 1 year of full program implementation.
							Program-level fidelity: Adequate fidelity =1. A score of 1 indicates 100% of schools score 1. Inadequate fidelity = 0. A score of 0 indicates < 100% of schools score 1.		1 year (2017–18): granted permission from OII due to 1 year of full program implementation.

Appendix B. National Evaluation of i3 Impact Reporting

This appendix includes full results from impact models for the confirmatory analyses measuring the impact of the invention kits on students' science knowledge achievement and on students' STEM attitudes. Furthermore, it includes full results from exploratory analyses measuring the impact of this intervention on each of the survey constructs: math attitudes, science attitudes, engineering and technology attitudes, and 21st-century learning attitudes.

Exhibit B1. Impact Results of the Invention Kits on Students' Science Knowledge Achievement (Confirmatory Analysis)

Variable	Coefficient	Standard error	*p*-value
Treatment indicator (β_1)	-0.10	0.25	.678
Post indicator (β_2)	0.00	0.11	1.000
Treatment, postinteraction (β_3)	0.06	0.13	.664
Student covariate			
Female	0.15*	0.07	.034
Disability status	-0.42***	0.10	< .001
English language learner status	-0.56**	0.20	.005
Black	-0.83***	0.10	< .001
Hispanic	-0.58***	0.13	< .001
Other races[a]	-0.29**	0.10	.005
Intercept (β_0)	0.30	0.19	.119
School random-effects parameter	-1.22***	0.30	< .001
Residual random-effects parameter	-0.12***	0.02	< .001
Number (*N*) of observations	838		

Note. White students are the omitted racial group. The number of observations is the number of students ($N = 419$) participating in both the pretest and posttest.

[a] Other races include Asian, American Indian or Alaskan Native, multiracial, and other races.

*$p < .05$. **$p < .01$. ***$p < .001$.

Exhibit B2. Impact Results of the Invention Kits on Students' STEM Attitudes (Confirmatory Analysis)

Variable	Coefficient	Standard error	*p*-value
Treatment indicator (β_1)	-0.20	0.15	.166
Postindicator (β_2)	0.00	0.12	1.000
Treatment, postinteraction (β_3)	0.07	0.14	.604
Student covariates			
Female	-0.07	0.07	.373
Disability status	-0.21	0.12	.071
English language learner status	-0.47*	0.21	.024
Black	-0.23*	0.10	.026
Hispanic	-0.03	0.13	.822
Other races[a]	-0.05	0.11	.638
Intercept (β_0)	0.11	0.12	.371
School random-effects parameter	-2.00***	0.42	< .001
Residual random-effects parameter	-0.06*	0.03	.016
Number (*N*) of observations	806		

Note. White students are the omitted racial group.

[a] Other races include Asian, American Indian or Alaskan Native, multiracial, and other races.

*p < .05. **p < .01. ***p < .001.

Exhibit B3. Impact Results of the Invention Kits on Students' Math Attitudes (Exploratory Analysis)

Variable	Coefficient	Standard error	*p*-value
Treatment indicator (β_1)	-0.28*	0.13	.035
Postindicator (β_2)	0.00	0.13	1.000
Treatment, postinteraction (β_3)	0.12	0.15	.429
Student covariates			
Female	-0.01	0.08	.906
Disability status	-0.29*	0.13	.024
English language learner status	-0.32	0.22	.153
Black	-0.06	0.11	.559
Hispanic	-0.01	0.14	.945
Other races[a]	0.02	0.12	.884
Intercept (β_0)	0.07	0.11	.496
School random-effects parameter	-2.42**	0.78	.002
Residual random-effects parameter	0.01	0.03	.608
Number (*N*) of observations	806		

Note. White students are the omitted racial group.
[a] Other races include Asian, American Indian or Alaskan Native, multiracial, and other races.
*p < .05. **p < .01. ***p < .001.

Exhibit B4. Impact Results of the Invention Kits on Students' Science Attitudes (Exploratory Analysis)

Variable	Coefficient	Standard error	*p*-value
Treatment indicator (β_1)	-0.24	0.12	.054
Postindicator (β_2)	0.00	0.11	1.000
Treatment, postinteraction (β_3)	0.03	0.14	.847
Student covariates			
Female	0.06	0.07	.394
Disability status	-0.12	0.11	.298
English language learner status	-0.40*	0.20	.045
Black	-0.25**	0.10	.010
Hispanic	-0.08	0.13	.529
Other races[a]	0.01	0.11	.890
Intercept (β_0)	0.06	0.10	.543
School random-effects parameter	-2.31***	0.57	< .001
Residual random-effects parameter	-0.11***	0.03	< .001
Number (*N*) of observations	806		

Note. White students are the omitted racial group.

[a]Other races include Asian, American Indian or Alaskan Native, multiracial, and other races.

*p < .05. **p < .01. ***p < .001.

Exhibit B5. Impact Results of the Invention Kits on Students' Engineering and Technology Attitudes (Exploratory Analysis)

Variable	Coefficient	Standard error	*p*-value
Treatment indicator (β_1)	-0.04	0.15	.785
Postindicator (β_2)	0.00	0.12	1.000
Treatment, postinteraction (β_3)	0.07	0.14	.638
Student covariates			
Female	-0.36***	0.07	.000
Disability status	0.03	0.12	.791
English language learner status	-0.41*	0.21	.049
Black	-0.22*	0.10	.029
Hispanic	0.12	0.13	.353
Other races[a]	-0.14	0.11	.214
Intercept (β_0)	0.12	0.12	.296
School random-effects parameter	-1.99***	0.39	< .001
Residual random-effects parameter	-0.08**	0.03	.003
Number (*N*) of observations	806		

Note. White students are the omitted racial group.

[a] Other races include Asian, American Indian or Alaskan Native, multiracial, and other races.

*p < .05. **p < .01. ***p < .001.

Exhibit B6. Impact Results of the Invention Kits on Students' 21st-Century Learning Attitudes (Exploratory Analysis)

Variable	Coefficient	Standard error	*p*-value
Treatment indicator (β_1)	-0.08	0.11	.470
Postindicator (β_2)	0.00	0.11	1.000
Treatment, postinteraction (β_3)	0.06	0.14	.658
Student covariates			
Female	0.18*	0.07	.012
Disability status	-0.28*	0.11	.015
English language learner status	-0.51*	0.20	.012
Black	-0.07	0.10	.460
Hispanic	-0.01	0.13	.943
Other races[a]	-0.02	0.11	.874
Intercept (β_0)	0.03	0.09	.742
School random-effects parameter	-3.06	1.60	.056
Residual random-effects parameter	-0.09***	0.03	< .001
Number (*N*) of observations	806		

Note. White students are the omitted racial group.

[a] Other races include Asian, American Indian or Alaskan Native, multiracial, and other races.

*$p < .05$. **$p < .01$. ***$p < .001$.

Exhibit B7. Contrasts for Science Knowledge and STEM Attitudes Outcomes

Contrast ID #a	Contrast name [expected reporting date]	Design	Treatment group			Comparison group	Outcome				Baseline		
			[Condition] Description	Age/grade during intervention	Exposure	[Condition] Description	Domain	Measure [Scale]	Unit of observation	Timing of measurement	Measure [Scale]	Unit of observation	Timing of measurement
C-Sci	Science knowledge assessment [not specified] (Confirmatory)	QED with difference-in-difference approach. School-level intervention.	[Invention Kits] Invention kit schools: All Grades 7 and 8 students with both pre- and posttest scores in schools using the invention kit.	Grade 7 or Grade 8	1 year	[Business as Usual] All Grade 8 students with both pre- and posttest scores in schools using the traditional science and engineering curriculum.	Science Achievement	Science knowledge assessment developed ad hoc for this evaluation. Scaled score [Continuous]. (Posttest)	Student	Spring (April) 2018	Science knowledge assessment developed ad hoc for this evaluation. Scaled score [Continuous]. (Pretest)	Student	Fall 2017
C-STEM	S-STEM survey [not specified] (Confirmatory)	QED with difference-in-difference approach. School-level intervention.	[Invention Kits] Invention kit schools: All Grades 7 and 8 students with both pre- and posttest scores in schools using the invention kit.	Grade 7 or Grade 8	1 year	[Business as Usual] All Grade 8 students with both pre- and posttest scores in schools using the traditional science and engineering curriculum.	STEM interest and confidence	S-STEM survey: Scaled score [Continuous]. (Posttest)	Student	Spring 2018	S-STEM survey: Scaled score [Continuous]. (Pretest)	Student	Fall 2017

Exhibit B8. Impact Estimates

A	B	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S
Contrast ID #	**Contrast Name (Optional)**	**Posttest Measure Name**	**Treatment group *N* of clusters**	**Treatment group *N* of students**	**Comparison group *N* of clusters**	**Comparison group *N* of students**	**Unadjusted treatment group SD**	**Unadjusted comparison group SD**	**Standard deviation source (Code)**	**Comparison group mean (optional)**	**Impact estimate**	**Standardized effect size (optional)**	**Impact standard error**	***p*-value**	**Code for impact model description**	**Degrees of freedom**	**Source of data (optional)**	**Level of inference (optional)**
C-Sci		Science knowledge assessment	4	288	3	131	1.08	1	[a]	0.00	0.06		0.13	.66	[b]	826		
C-STEM		S-STEM	4	275	3	128	1.03	1	[a]	0.00	0.07		0.14	.60	[b]	794		

[a]Student-level SDs calculated from the sample shown on this row. [b]The model used to estimate this impact is shown in the design summary on file with the AIR team.

Exhibit B9. Baseline Equivalence of Students

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Contrast ID #	**Contrast name (optional)**	**Pretest measure name**	**Treatment group *N***	**Comparison group *N***	**Unadjusted treatment group SD**	**Unadjusted comparison group SD**	**Standard deviation source (Code)**	**Comparison group mean (optional)**	**Treatment – comparison difference**	**Standardized T-C difference (optional)**	**Pretest shown in this row was used as a control in the impact model for this contrast ? (Y/N)**	**Code for T-C difference calculation**	**Source of data (optional)**
C-Sci		Science knowledge assessment	288	131	0.97	1	[a]	0.00	0.00		Y	[b]	
C-STEM		S-STEM	275	128	0.85	1	[a]	0.00	0.19		Y	[b]	

[a]Student-level SDs calculated from the sample shown on this row. [b]The T-C difference shown in column J calculated as simple difference of unadjusted means (described in Method 1 of i3 findings; in Reporting Shells_09222014.docx.

Appendix C. Psychometric Analysis and Scaling of Outcome Measures

The S-STEM survey and the science knowledge assessment items were scaled using the Rasch model for ordered response categories (Andrich, 1978; Rasch, 1980; Wright & Masters, 1982) to determine whether the items reliably measure the constructs they intend to measure.[10] For the S-STEM survey, all items were scaled together to generate an overall STEM attitudes score that was used in the confirmatory analysis. Furthermore, items designed to measure single underlying constructs, such as student math attitudes, were also scaled separately and used in the exploratory analysis. For the science knowledge assessment, all items were scaled together.[11] Two sets of construct scale scores were generated for the two administrations (i.e., pretest and posttest). The scale scores provide a quantitative view of the frequency and intensity of respondents' answers across a set of items representing a given construct. Scale scores were equated across time (Wright, 1996) to ensure that they are comparable across administrations.

In addition to generating scale scores, the Rasch analysis yields several statistics that allow for assessment of reliability and validity. Reliability is an estimate of the precision of the measures (construct scale scores). Validity refers to the extent to which psychometric evidence supports the intended use of the scale scores. Here, we focus on two statistics: the Rasch person separation reliability index (also referred as Rasch reliability) and Cronbach's alpha statistic. The Rasch person separation reliability index is a measure of how well the scale can distinguish among individuals of varying levels on the scale. Cronbach's alpha is a measure of the internal consistency of a scale where internal consistency describes the extent to which all items in the scale measure the same concept. Reliability values for the two statistics range from 0 to 1, with values closest to 1 being considered best and values of 0.7 or higher considered as strong. Levels of performance criteria for these two reliabilities are summarized in Exhibit C1.

Exhibit C1. Performance Criteria for Rasch Test Statistics

Statistic	Performance
$x \geq 0.9$	Excellent
$0.8 \leq x < 0.9$	Good
$0.7 \leq x < 0.8$	Acceptable
$0.6 \leq x < 0.7$	Questionable
$0.5 \leq x < 0.6$	Poor
$x < 0.5$	Unacceptable

Although these are general guidelines, it is important to note that the criterion for an acceptable reliability is dependent on the intended use of the scores. If the scores are intended to differentiate

[10] For assessment, only the comparison group was used to investigate the psychometric properties, such as reliability and validity, to avoid the intervention possibly influencing how these items function psychometrically, because some items were poorly functioning when using both treatment and comparison groups.
[11] Two items were identified as misfit items during the psychometric analysis for both pre- and posttests; they were removed from the process that generated the scale score.

performance or rankings of individuals, as might be the case in a state assessment, then high levels of reliability are usually desirable. However, if scores are intended to differentiate between groups, low reliabilities are not overly problematic. If the scores are to be used to predict an outcome (as might be the case in a score that measures level of implementation), analytic methods can be employed that account for measurement error.

S-STEM Survey

The S-STEM survey is a publicly available survey that measures students' attitudes toward STEM subjects, as well as postsecondary pathways and career interests (Friday Institute for Educational Innovation, 2012).[12,13] There are two survey versions: one for elementary grades (Grades 4 and 5) and one for middle and high school grades (Grades 6–12). This study used the latter survey version given that the intervention targets Grade 8 students.

The survey contains six sections: four sections each designed to measure a single underlying construct (math, science, engineering and technology, and 21st-century learning attitudes); and two sections each containing items asking about students' future interest in STEM career fields, their expectations of future academic performance, and their plans for future coursework and postsecondary studies. Following is a description for each of the four constructs:

- **Math attitudes:** The survey contains eight items measuring self-efficacy toward math and expectations for future value gained from learning math.
- **Science attitudes:** The survey contains nine items measuring self-efficacy toward science and expectations for future value gained from learning science.
- **Engineering and technology attitudes:** The survey contains nine items measuring self-efficacy toward engineering and technology and expectations for future value gained from learning engineering and technology.
- **21st-century learning attitudes:** The survey contains 11 items measuring students' confidence in skills such as collaboration, communication, and self-directed learning.

For the confirmatory analysis, items from all four constructs were combined and psychometrically analyzed together to then create an overall survey of a "STEM" attitudes scale score for each survey administration (i.e., pre- and posttests). Scores were then equated as previously described.

Rasch analysis results indicate that overall the student survey scales functioned well. As reported in Exhibit C2, the Rasch reliabilities for the overall STEM attitudes survey scale scores were 0.93 at both pre- and posttests; and Cronbach's alpha values were 0.94, also for both survey administrations. By construct, Rasch reliabilities ranged from 0.82 to 0.93 at pretest and 0.83 to 0.93 at posttest; and Cronbach's alpha

[12] For general details on the S-STEM survey and to request access, see North Carolina State University, "Maximizing the Impact of STEM Outreach [MISO]," https://miso.ncsu.edu/articles/s-stem-survey.
[13] For more details on how it was developed and its psychometric properties, see Unfried et al. (2015) and MISO, "Student Attitudes toward STEM (S-STEM) Survey: *Development and Psychometric Properties*," retrieved from https://miso.fi.ncsu.edu/wp-content/uploads/2013/06/S-STEM_FridayInstitute_DevAndPsychometricProperties_FINAL.pdf

values ranged from 0.87 to 0.94 at pretest and 0.90 to 0.94 at posttest.[14] Whereas the scales functioned well for both the overall survey scores and by construct, review of the complete Rasch analysis results identified some areas to consider for improvement and revision in future administrations, including the wording of items related to item fit and multidimensionality issues (i.e., items designed to measure one construct measure more than one construct).[15]

Exhibit C2. Reliability Statistics for the Overall S-STEM Survey and by Construct

	Pretest		Posttest	
Construct	Cronbach's α	Rasch reliability	Cronbach's α	Rasch reliability
STEM attitudes	0.94	0.93	0.94	0.93
Math attitudes	0.87	0.82	0.90	0.83
Science attitudes	0.89	0.87	0.92	0.88
Engineering and technology attitudes	0.91	0.87	0.92	0.88
21st-century learning attitudes	0.93	0.87	0.94	0.88

Physical Science and Engineering Knowledge Assessment

A standards-aligned science knowledge assessment was developed by a test developer as part of this grant, with content collaboration from AIR and grantees. The assessment covered electromagnetic concepts and contained assessment items (20 multiple questions and seven open-ended questions) under the following standard categories: forces and interactions, energy, engineering design, and waves and electromagnetic radiation.

Rasch analysis results show that in general the assessment scale functioned acceptably (Rasch reliabilities and Cronbach's alpha assessments are in Exhibit C3). As reported in Exhibit C3, Rasch reliability for the assessment scale is 0.74 and 0.77 at pretest and posttest, respectively; and Cronbach's alpha values are 0.77 and 0.80 at the two time points, respectively.[16] Whereas the assessment scale functioned well, review of the complete Rasch analysis results identified two misfit items (i.e., the off-variable noise caught by the item is greater than useful information and degrades the measurement; in other words, it indicates that the item is measuring a different construct). As a result, these two items were removed before generating the scale scores.

Exhibit C3. Reliability Statistics of Assessment

	Pretest		Posttest	
Scale	Cronbach's α	Rasch reliability	Cronbach's α	Rasch reliability
Science Knowledge Assessment	0.77	0.74	0.80	0.77

[14] Cronbach alpha values are consistent with those found by survey developers: 0.90 for math attitudes, 0.89 for science attitudes, 0.89 for engineering and technology attitudes, and 0.91 for 21st-century learning attitudes (Unfried et al., 2015).
[15] Multidimensionality issues were expected from scaling the overall STEM attitudes survey scores since they are based on items from four single underlying constructs. Some potential multidimensionality issues were also observed at the construct level, particularly for science and 21st-century learning attitudes.
[16] These statistics are based on all assessment items.

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Established in 1946, the American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research on important social issues and delivers technical assistance, both domestically and internationally, in the areas of education, health, and workforce productivity.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL
Indianapolis, IN | Metairie, LA | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC
Austin, TX | Arlington and Reston, VA | Seattle, WA

International: Algeria | Ethiopia | Germany | Haiti | Zambia

10430_12/19