Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1642

# Cultivating STEM Interest in Middle School Students: Evidence from a Randomized Controlled Trial of an Out-of-School Time Experiential Learning Program

Christopher Doss
RAND

AK Keskin
RAND

A sustainable science, technology, engineering, and math (STEM) workforce is central to U.S. national interests and policy debates. Researchers and practitioners often overlook non-academic levers such as early interest and career aspirations. One way to foster these dispositions is through out-of-school time (OST) programs that provide learning experiences not typically offered in school settings. However, few causal studies on these programs exist. We field a randomized controlled trial to understand whether an experiential, space-themed STEM OST experience provided by the Challenger Center can affect sixth-grade students' STEM attitudes and beliefs. We find that the experience initially increased students' self-efficacy in performing STEM activities by 0.12 standard deviations (SD) and attitudes towards STEM careers by 0.16 SD. Exploratory analyses provide suggestive evidence that these effects were larger for students from underrepresented communities. Initial effects fade six months later and student participation in science course taking in seventh grade and in STEM OST programs did not change. Our results suggest short OST experiences can produce effects, but sustained exposure to novel experiences might be needed to maintain effects and affect behavior.

VERSION: November 2026

Suggested citation: Doss, Christopher, and AK Keskin. (2026). Cultivating STEM Interest in Middle School Students: Evidence from a Randomized Controlled Trial of an Out-of-School Time Experiential Learning Program. (EdWorkingPaper: 26-1642). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/f1md-k036

# Cultivating STEM Interest in Middle School Students: Evidence from a Randomized Controlled Trial of an Out-of-School Time Experiential Learning Program

Christopher Doss
AK Keskin
RAND

## Abstract

A sustainable science, technology, engineering, and math (STEM) workforce is central to U.S. national interests and policy debates. Researchers and practitioners often overlook non-academic levers such as early interest and career aspirations. One way to foster these dispositions is through out-of-school time (OST) programs that provide learning experiences not typically offered in school settings. However, few causal studies on these programs exist. We field a randomized controlled trial to understand whether an experiential, space-themed STEM OST experience provided by the Challenger Center can affect sixth-grade students' STEM attitudes and beliefs. We find that the experience initially increased students' self-efficacy in performing STEM activities by 0.12 standard deviations (SD) and attitudes towards STEM careers by 0.16 SD. Exploratory analyses provide suggestive evidence that these effects were larger for students from underrepresented communities. Initial effects fade six months later and student participation in science course taking in seventh grade and in STEM OST programs did not change. Our results suggest short OST experiences can produce effects, but sustained exposure to novel experiences might be needed to maintain effects and affect behavior.

Acknowledgements: The authors would like to acknowledge staff from Challenger Center and local Challenger Learning Centers without whom this study would not be possible including: Paula Dorfman, Melissa Duncan, Valerie Fitton-Kane, Kathleen Guitar, Kevin Harrell, Denise Kopecky, Jackie Shia, Heather Townley, Lisa Vernal, and Suzanne Wright. We would also like to acknowledge the work of the RAND Survey Research Group, led by Danielle Schlang, who worked to collect student surveys over the course of two years. We especially appreciate working with the educators and administrators in the participating schools and districts who took the time to participate in the study, including data collection, in addition to their professional responsibilities. Finally, we would like to thank the Overdeck Family Foundation for the generous funding for the study.

## 1. Introduction

Policymakers, researchers, and industry stakeholders have widely cited a sustainable science, technology, engineering, and mathematics (STEM) workforce as a national need to ensure the United States' economy and technological output remain globally competitive (Tai et al., 2006; Young et al., 2017; Allen et al., 2019; Saw, 2020; National Science Board [NSB], 2026). The demand for STEM labor is projected to outpace non-STEM labor threefold between 2024-2034 (Bureau of Labor Statistics, 2025), all while concerns about upholding the nation's STEM workforce pipeline remain (NSB, 2026). The latest 8$^{th}$-grade National Assessment of Educational Progress (NAEP) science assessment scores in 2024 fell significantly across all percentiles for the first time (National Center for Education Statistics [NCES], 2024) since it was updated with a new science framework in 2009 (National Assessment Governing Board, 2008). In both 2009 and 2024, students scored an average 150 points out of 300 points. Additionally, a recent landscape analysis of science and engineering in the United States by NSB (2026) found that only 38 percent of 9$^{th}$ graders in 2009 enrolled in a STEM postsecondary certificate or degree program, and only 55% then completed their degree by 2017. Overall, the NSB (2026) report argues that such levels of STEM degree attainment may result in a labor force that is unable to fulfill the STEM needs of the nation.

While policy debates about improving the STEM workforce pipeline largely center on raising test scores, non-academic factors such as early interest and career aspirations in STEM can play an equally crucial role (Tai et al., 2006; Maltese & Tai, 2011). For instance, students with eighth-grade STEM career aspirations are 1.9 times more likely to earn life sciences degrees and 3.4 times more likely to earn physical sciences/engineering degrees than those without these expectations (Tai et al., 2006). Other studies find evidence that early STEM-related beliefs and

attitudes not only predict STEM major uptake in college, but also future STEM career expectations. Further, key high school indicators such as the number of science courses attempted and a planned college STEM major in 12th grade predicted STEM degree attainment, even when controlling for academic achievement. (Maltese & Tai, 2011).

Ages 10-14 are a crucial window in the formation of student STEM aspirations, but students often find STEM less engaging than other school subjects (Archer et al., 2012). Exposing students to unique experiential learning opportunities related to real-world STEM settings and professionals is one way to foster students' STEM interest (Saw et al.,2019; Young et al. 2017), engagement (Allen et al. 2019), identity (Sonnert et al, 2025) and career interest (Dabney et al., 2012). Out-of-school time (OST) programs often offer these types of authentic experiences. Although there is some evidence that approaches used by OST STEM programs can be effective in improving students' STEM-related beliefs and attitudes, there is a dearth of rigorous evaluations of OST STEM programs (NRC, 2015; Saw et al., 2019; Allen & Noam, 2024; Sonnert et al., 2025). The few studies that exist have severe limitations: they are largely descriptive, have small sample sizes, and may not use validated measures of students' STEM attitudes or career aspirations (Saw et al., 2019).

In this study, we fielded a block-clustered randomized controlled trial (RCT) to understand whether an experiential, space-themed OST experience provided by the nonprofit Challenger Center can affect sixth-grade students' STEM attitudes and beliefs. Since its inception in 1986, Challenger Center engaged over 6.5 million students in simulation-based STEM missions (Challenger Center, 2026). Teachers led class field trips to a local Challenger Learning Center (CLC), where students stepped into a highly immersive environment to role-play as STEM professionals (e.g., scientists, engineers, astronauts, medical professionals) and worked

collaboratively on space exploration mission tasks, such as searching for evidence of life on Mars, exploring lava tubes on the Moon, or redirecting a comet from impacting Earth.

We estimate the effect of Challenger Center mission participation on measures of self-efficacy in science activities and coursework, attitudes and interest in STEM careers, and STEM identity immediately after the mission and six months later to assess the persistence of initial effects. We also estimate effects on two self-reported choices to further engage in STEM: enrollment in seventh grade STEM electives and subsequent participation in OST STEM programs or activities (i.e., afterschool and summer STEM programs). We find that Challenger Center missions increased students' self-efficacy in performing science activities by 0.12 standard deviations (SD; $p < 0.05$) and attitudes towards STEM careers by 0.16 SD ($p < 0.01$). Our exploratory analyses also provide suggestive evidence that these effects were larger for students from communities traditionally underrepresented in STEM (i.e., girls and racial/ethnic minorities). Initial effects, however, faded six months later and we do not find evidence that mission participation affected other measures, including seventh grade elective course-taking and participation in OST STEM activities.

## 2. Background

### 2.1 The Theory and Evidence Behind OST Programs

The factors that influence a student's attitudes towards STEM coursework and careers are dynamic and multifaceted. However, research indicates that access to high-quality STEM learning experiences that foster a student's STEM identity and self-efficacy promotes interest in STEM coursework and careers (Allen et al., 2019; Young et al., 2017). Practitioners and policymakers consider OST STEM programs a valuable approach to promoting STEM skills and dispositions in students because they can provide learning experiences that can be difficult to

incorporate into school settings. These programs come in a variety of forms, from experiences outside the classroom during the school day (e.g., museum visits) to afterschool programming at school (e.g., science clubs) or outside the school, to summer programs. These experiences often include hands-on learning opportunities that can involve meeting and working with STEM professionals in authentic STEM settings and applying STEM knowledge to real-world situations (NRC, 2015; Saw et al., 2019; Holian, 2024).

Research suggests that experiential learning, which includes approaches leveraged by Challenger Center such as simulation, role-play, and narratives, can influence how students view STEM studies and careers. Computer-driven simulations and game-based learning can help students test hypotheses, participate in teams, learn to use new technology, and get a sense of engaging in STEM beyond textbooks and lectures (Barab & Dede, 2007; Barton et al., 2008; Wills et al., 2007; Meluso et. al 2011; Rutten et al., 2012; Thisgaard & Makransky, 2017). Further, through simulations of actual environments, students can gain an understanding of real-world problems (Lunce, 2006). Learners become emotionally engaged when they are personally involved in a story (Raybourn, 2014), and this emotional involvement or psychological investment is underlies student engagement in academic learning (Newmann et al., 1992). Benefits of educational simulation include heightened engagement and motivation (Cantrell & Sudweeks, 2009), collaboration that teaches life skills (Nugent et al., 2011) and fosters deep learning (Meluso et al., 2011), and academic self-efficacy (Redmond et al., 2011) that impacts career choices and goals (Kim, 2011).

The empirical research literature on the effects of OST STEM programs on student outcomes provides some evidence to support the notion that OST programs and their approaches can improve student outcomes. A meta-analysis of 15 STEM OST studies found an overall

association between OST STEM programs and student interest in STEM of 0.37 SD, with larger associations from programs with a social and academic focus (0.45 SD) compared to those with only an academic focus (0.17SD). Female students, male students, students from underrepresented communities, and students from represented communities all experienced similar associations (Young et al., 2017). We see similar findings in other evaluations of STEM OST programs. In a cross-sectional national study of 158 afterschool STEM programs serving students in grades 4-12, Allen et al. (2019) found significant positive associations in a range of STEM attitudes (STEM identity, engagement, career interest, career knowledge) through retrospective pre-post student surveys on the order of 0.35-0.5 SD.

Given the breadth of OST programs, some of which are light touch and short term and some of which provide services for a longer period of time, the relationship between the length of services provided to students and effects on student outcomes is an open question. The literature indicates that programs which are longer in duration may produce larger effects on students, but with diminishing returns. Allen et al.'s (2019) study of OST STEM programs found the largest gains in STEM attitudes in programs lasting at least 4 weeks with at least an hour of STEM activities per week, though effects didn't differ between 4-8 weeks and more than 8 weeks of participation. Looking at the larger OST literature beyond STEM oriented programs, Lauer et al.'s (2006) meta-analysis of the effect 35 OST programs on mathematics and reading achievement revealed that effects were significantly larger in programs lasting more than 45 hours, though the longest-duration programs (greater than 210 hours) either were not significant or had smaller effect sizes relative to medium-duration programs (45-210 hours).

Although these results are promising, there are limitations to the cited studies. Maily, many are largely descriptive, often have small sample sizes, and may not use validated measures of

student attitudes towards STEM and STEM careers (McCombs, Whitaker, and Yoo; Saw at al., 2019). Further, while these studies suggest that longer duration programs can produce larger effects, we could not find any studies that sought to understand whether longer duration programs also produced effects that persist after the end of OST programs.

Rigorous studies that examine field trips are even scarcer. We identified two studies that employed causal methods. Whitesell (2016) used a school fixed effects design to understand whether teacher professional development combined with field trips to informal science learning opportunities, such as museums, affected student achievement in science. She found that they led to small improvements in science standardized test scores of 0.026 SD, with effects concentrated on Hispanic and lower-income students. Outside of the STEM domain, Kisida et al. (2016) conducted an RCT of an art museum field trip with 7,500 students enrolled in grades 3-12. The authors found treatment effects ranging from 0.08 to 0.13 SD on students' critical thinking skills when evaluating a work of art.

On the whole, there are theoretical reasons to believe that OST programs including field trips can affect student dispositions towards STEM. Although the evidence to date suggests positive correlations between OST experiences and student outcomes, the lack of rigorous studies, particularly RCTs, heightens the importance of continued evaluations.

### 2.2 Prior Research on Challenger Center

To date, three studies have been conducted on Challenger Center missions and all support the notion that the program's potential effects on student attitudes towards STEM content and careers can be sizable. Jarvis & Pell (2002) surveyed elementary school children in England prior to and after attending a mission and found that 24 percent expressed more interest in science-related careers and greater enthusiasm for science after the mission, with significantly larger

associations reported for female students' aspirations to become scientists on the order of 0.14 SD. Chen & Howard (2010) used a similar method to evaluate a mission delivered via distance-learning technology and found significant post-mission improvements in one subscale of student science interest with male students experiencing significantly larger gains than female students in both science attitudes and career interest. Finally, Crawford & Huscroft-D'Angelo (2015) fielded the largest Challenger Center mission study to date with approximately 3,000 students in grades 6-8 attending a CLC in Colorado. They found significant pre- post- increases in item-level responses on the order of 0.20 SD to 0.90 SD across all of their eight items on student attitudes toward math, science, and science careers, with larger effects for male students for most items.

The research base on the effect of Challenger Center missions mirrors that of the larger OST literature. Observational studies point to potentially large effects, but methodological limitations, including a lack of control students, necessitate a more rigorous study to confirm the magnitude of potential effects.

### 3. *Challenger Center Missions*

Challenger Center engages hundreds of thousands of fifth through eighth grade students annually with its Center Mission program–simulated missions to space, developed in collaboration with subject-matter experts from National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration, and other public and private partners.

Missions take place at local CLCs as part of a field trip during the school day. Challenger Center missions are live role-play experiences in which a group of around 20-40 students (usually children in the same classroom) work together to achieve a unified goal. The centers are located in communities across the U.S. spanning large cities, suburban communities, and rural

areas. Nearly all CLCs are in communities predominantly populated by underrepresented minorities, Title I schools, and/or other under-resourced areas. The CLCs recruit student groups from local public and private schools as well as homeschooled students. As a result of the CLC network's geographic range and foci in underserved communities, CLC missions are designed for the broadest set of students, which includes many individuals who have historically been excluded, underserved, or underrepresented in STEM, including girls, students from minority communities, students with physical disabilities, neurodiverse students, and English language learners (ELL).

Missions take place in a highly immersive environment that allows students to take on real-life roles such as astronauts working in space or technical experts in mission control. Challenger Center staff have set up the space to resemble real-life facilities, such as a spacecraft interior, and contain real science equipment, computer workstations, and audio-visual equipment. Students role-play as scientists, engineers, medical professionals, and other STEM occupations on missions to, for example, search for evidence of life on Mars, explore lava tubes on the moon, or redirect a comet from impacting Earth. Software provides instructions and information students use to conduct hands-on experiments and analyze data. The mission also includes background information on specific science topics, data analysis tools, simulations of specific concepts, and tools for communicating within and between teams (e.g., chat functionality, radio, and video communications). Challenger Center applies Universal Design for Learning principles and other strategies to accommodate students from all backgrounds.

Space exploration is naturally interdisciplinary, enabling Challenger Center to address a range of Next Generation Science Standards and Common Core State Standards. In each CLC Mission, teams align with a wide range of STEM occupations and topics, including biology,

geology, navigation, weather, robotics, engineering, communications, life support, rover operations, and medicine. Each team conducts experiments and analyzes data relevant to its topic (e.g., biology). They share data within the team and with other teams to achieve the simulation's overall goal.

Figure 1 presents the Theory of Change of the Challenger Center. The Challenger Center missions are meant to provide opportunities to engage in more authentic science activities that are tied to the real word. These activities may also change student attitudes towards science content where they are more likely to find them interesting and exciting. Stepping into the role of real-life STEM careers and observing the STEM career roles that their classmates engage in can also change students' attitudes towards and interest in careers. If attitudes towards activities, content, and careers change, students may also begin to regard coursework differently. Success with these authentic experiences is meant to help build of sense of STEM identity

Overall, missions aim at to 1) deepen student engagement and interest in STEM, 2) increase STEM self-efficacy, and 3) introduce students to a range of STEM careers. Improvements in these areas will then lead to longer-term outcomes of 1) improved STEM identity, 2) interest in pursuing STEM careers, 3) greater participation in OST STEM programs, and 4) greater participation in formal STEM coursework.

In this study, almost all students completed a Mars themed mission called *Expedition to Mars* while one district completed a moon-themed mission called *Lunar Quest*. Prior to the mission, teachers prepared students by completing a crew manifest to split students into teams, instructing them on the basic science needed for the missions, and asking students to complete a job application to rank their top career role for the mission. After the mission, CLCs provided teachers with activities that are designed to help process the experience. These activities

addressed misconceptions, reflected on STEM careers, provided follow-up exercises to extend analytic concepts used in the mission, asked students to write about the experience or create digital products (e.g., a video), and connected their mission to current NASA missions. While pre-mission activities were required, post-mission activities were optional and varied by CLC.

## 4. Setting and Methods

### *4.1* Setting and Research Questions

We conducted this study in seven districts containing with 25 unique teachers teaching 91 unique sixth-grade science classes across five states (New Mexico, Kentucky, Ohio, Massachusetts, and Pennsylvania). We worked with Challenger Center and the local CLCs to recruit the participants. We recruited in two cohorts with five districts and 20 teachers participating in cohort one and four districts and 12 teachers participating in cohort two. Two districts and seven science teachers participated in both cohorts.

The districts spanned urbanicity categories, with the National Center for Education Statistics categorizing three districts as rural or remote towns, three districts as suburban, and one district as urban. To be eligible, schools could not have sent the current class of sixth graders to missions in the prior, fifth grade year to ensure the study would measure the effect of the exposure of students to missions instead of the incremental effect of additional missions. The study paid for the mission for transportation to and from the CLC.

The study aims to answer the below questions (RQs) which speak to the outcomes in the theory of change. We preregistered the study on the Registry of Efficacy and Effectiveness Studies (#24480) and ex ante designate confirmatory and exploratory research questions. Research questions that analyze the effects of the missions "immediately" after students experience them (i.e., within approximately two weeks) are considered confirmatory. Research

questions that analyze the effects of the mission approximately six months after the missions as well as heterogeneity analyses and teacher perceptions of student engagement are exploratory. We choose one measure from each outcome domain as confirmatory. These distinctions preclude the need for multiple hypothesis testing on the confirmatory analyses. Exploratory analyses are designed to guide future research on Challenger Center missions and OST STEM programs and should be confirmed through additional studies.

Confirmatory:

1. What is the effect of Challenger Center missions on student-reported self-efficacy in science, interest in science, and STEM identity/interest in STEM careers immediately after the mission?
2. What is the effect of Challenger Center missions on student science elective course-taking in seventh grade?
3. What is the effect of Challenger Center missions on STEM OST participation during the study period?

Exploratory:

1. What is the effect of Challenger Center missions on teacher perceptions of student engagement in science immediately after the mission?
2. What is the effect of Challenger Center missions on student-reported self-efficacy in science, interest in science, and STEM identity/interest in STEM careers six months after the mission?
3. How do the above effects vary by student and teacher background characteristics?

### *4.2 Study Design*

We fielded a block-clustered randomized controlled trial. To maximize sample recruitment, we worked with districts and used their preferred level of blocking to assign treatment status. In all but one district, we blocked at teacher-level and randomly assigned class-sections to either experience the mission or act as the control group. One district preferred teacher-level randomization, and we randomized teachers and all their classes to either the missions or the control condition. In this case, the district is the block.[1] As stated earlier, two districts participated in both cohorts. In order to avoid the possibility that the field trips would affect student assignment to teachers or classes, we rerandomized in the second cohort.

Due to equity concerns, school districts requested that classrooms randomized to the control condition be given a supplementary STEM experience. In all but two districts, control classrooms were provided with a planetarium show at the CLC. Two districts in cohort 2 had control classrooms experience the missions after all data collection activities were complete. In most cases, the planetarium show was provided to the control classes as the treatment classes participated in the mission to minimize school disruptions and transportation costs. Though experiencing the planetarium show may affect the student dispositions that we are measuring, our theory of change hypothesizes that the active component of the missions would drive effects. We therefore expect that a planetarium experience would have a minimal effect on student outcomes.

[1] In our pre-registration, we envisioned only teacher-level randomization. Due to recruitment challenges, we approached districts to see if they would be amenable to class-level randomization and all but one district agreed. Class level randomization allowed us to field a better powered study with fewer participants.

### *4.3 Study Procedures*

Figure 2 depicts the timeline of the study's research activities. We administered three surveys to students: before the treatment (baseline), "immediately" (within approximately two weeks) after the mission (endline), and approximately six months after the mission (follow-up). In all survey waves, participating teachers set aside class time for students to complete our surveys. Teachers completed their own surveys assessing student engagement in class at baseline and endline. All districts, except one, completed the surveys electronically. One district completed paper surveys. Student surveys took fifteen minutes or less to complete and teacher surveys took five minutes or less to complete.

The first cohort completed baseline surveys, randomization, the missions, and endline surveys in the second semester of the 2024-2025 school year. Their six-month follow-up survey occurred in the fall of the 2025-2026 school year. The second cohort completed their baseline surveys, randomization, missions, and endline surveys between October 2025 and February 2026. Six-month follow-up surveys were completed between April and May 2026.

For all classes randomized to treatment, local CLCs provided teachers with the aforementioned pre-mission activities. Implementation surveys indicated that all teachers completed the Crew Manifest, 90 percent completed the pre-mission, and 80 percent completed the student job application. CLC centers were also encouraged to provide post-mission activities. However, uptake was low with only nine percent of teachers reporting completion.

### *4.4 Participants*

Final recruitment resulted in the randomization of 1,761 students. Our study consists of three analytic samples: (1) surveys of students immediately after the mission; (2) surveys of students approximately six months after the mission; and (3) the analysis of seventh grade science

elective course taking. Fewer students took the six-month surveys, 72.5% of the randomized students, due to mobility. Only four of the seven districts provided opportunities for students to choose science electives in seventh grade, representing 66.4 percent of the randomized students. Following What Works Clearinghouse (WWC, 2022) guidelines, we only analyze students in classes at the time of randomization and do not include any students who joined the classes later in the year (i.e., *joiners*).

Table 1 provides descriptives for the sample of 1,451 students who completed surveys immediately after the mission. Tables A.1 and A.3 in the online appendix provide analogous statistics for our other two analytic samples. Students, on average, were 11.75 years old, as is typical for sixth graders. Approximately half were female; one-third identified as Hispanic; about half identified as White; one-fifth identified as Black; and 30 percent identified as various races in smaller numbers, which we combine into an Other category. At baseline, approximately 14 percent participated in an afterschool or summer STEM program (our measure of OST participation) and students also rated themselves about at the midpoint of the Likert scale in our surveys. We discuss these measures in more detail in the next section.

### *4.5 Measures*

We field a comprehensive set of measures meant to investigate the impact of Challenger Center on a variety of student domains listed in the Theory of Change in Figure 1. We chose one measure per domain for our confirmatory analyses. Table 2 details the survey-based measures, the source of the measures, an example item in the scale, and the internal consistency in our analytic samples. For all measures, except our measure of STEM identity, we average the item responses (e.g., in one through five Likert scale) that compose each subscale and only include students who responded to all items in the subscale. In all cases, internal consistencies are greater

than the 0.6 threshold recommended by WWC (WWC, 2022) and are as high as 0.89 for some scales. These statistics suggest that these measures are precise enough to support inferences about program effects. The Single-Item Measure of STEM Identity is a one-item measure and therefore cannot have a measure of internal consistency. McDonald et al. (2019) provide evidence of criterion and predictive validity.

We measure student self-efficacy in their ability to do science activities with a four-item subscale of the Self-Efficacy for Learning and Doing Science Scales (Porticella, Phillips, and Bonney, 2017). One, five-item subscale of the STEM Semantics Survey (Tyler-Wood, Knezek, and Christensen, 2010) measures attitudes towards science content (e.g., whether they find science interesting and exciting). We measure student attitudes towards STEM careers (e.g., whether they find STEM careers interesting and exciting) with a second, five-item subscale of the STEM Semantics Survey and whether students would consider a career in science with a five-item subscale of the Student Attitudes Towards STEM Survey (Faber, Townsend, and Weibe, 2013). We measure students' self-efficacy science coursework with a separate, four-item subscale of the Student Attitudes Towards STEM Survey. Finally, we leverage the Single-item Measure of STEM Identity (McDonald et al., 2019) to measure students' STEM identity. All the above measures take the form of Likert scales that range from five to seven options.

In addition to student self-report attitudinal measures, we include two self-reported measures of their choice to pursue more STEM learning opportunities. The first is whether students participate or plan to participate in any organized STEM afterschool or summer activities. These activities include camps with math, science, robotics, engineering, invention, or computer projects. We ask about afterschool and summer activities separately and combine them into one variable that indicates a student participated or planned to participate in at least one of the two

types of activities. Both cohorts were asked retrospectively about their summer activities prior to the mission at baseline. In the six-month follow-up surveys, we asked students in cohort 1 whether they participated in these activities in the summer after their mission and students in cohort 2 whether they plan to participate in these activities in the coming summer. We asked both at baseline and in the six-month follow-up surveys about their current afterschool activities. The final measure is whether students chose a science elective in seventh grade. Four of the seven districts allow for a choice of electives in seventh grade, one of which can be a science elective. We obtain an indicator for whether a student chose a science elective as part of the administrative data from those districts. In all cases, the behavioral measures are binary measures for either participating in the OST program or elective or not.

Finally, to complement the student self-reports, we asked teachers to judge each class section of their science course on their engagement using the Math and Science Engagement Scales (Wang et al., 2016). With only 91 class sections, this analysis is underpowered; however, we included them to provide an in-class student outcome that was not self-report and consider it an exploratory analysis.

### *4.6 Analysis Models*

The randomization of classes (or teachers) to the Challenger Center mission allows us to, in expectation, attribute any differences in student outcomes to the missions. We formally use ordinary least squares models of the following form to estimate student level outcomes:

$$Y_{icth} = \beta_0 + \beta_1 Mission_{cth} + \boldsymbol{X_{icth}\beta_2} + \boldsymbol{W_{cth}\beta_3} + \beta_4 Y^b_{icth} + \gamma_b + \varepsilon_{icth} \quad (1)$$

where $Y_{icth}$ represents the student-level outcome, *i*, in class section, *c*, of teacher, *t*, in cohort, *h*. $Mission_{cth}$ is an indicator for a class section being randomly assigned to the Challenger Center mission; $\boldsymbol{X_{icth}}$ are student-level background characteristics, $\boldsymbol{W_{cth}}$ are classroom section aggregates

of student background characteristics, $Y_{icth}^{b}$ is a baseline measure of the outcome, where available, $\gamma_b$ are block fixed effects to account for the randomization scheme, and $\varepsilon_{icth}$ is a stochastic, student-level error term. Student-level covariates included age, gender, race/ethnicity, an indicator for special education status, an indicator for English language learner status, an indicator for free or reduced priced lunch status, an indicator for whether the student reports any previous interaction with Challenger Center programming, and fifth grade statewide math and English language arts scores standardized by district-cohort which were all collected from district administrative records.

We explored two types of analysis when looking at overall effects. Our preferred models include all teachers and classes in one model, and we define the blocks as teachers-year cells where class-level randomization occurred and defined the block as the district-year cell where teacher-level randomization occurred. Standard errors are clustered by class-year to account for the correlation of student experiences in a class in a given cohort. In our second approach, we estimate separate models by randomization scheme and cluster our standard errors at the respective level of randomization (class-year or teacher-year). We then take a fixed-effects meta-analytic average of the two estimates. We present this second approach as a robustness check.

When examining heterogeneity by student characteristics, we include in equation 1 an interaction between $Mission_{cth}$ and the student characteristic of interest. Our estimate of the effect of students in the subgroup of interest is the sum of the coefficients on $Mission_{cth}$ and the interaction term. We present point estimates of all continuous measures in effect sizes that take the form of Hedges' $g$. We use linear probability models to analyze binary outcomes, as they are more interpretable.

Similarly, models analyzing teacher reports of engagement took the following form:

$$Y_{cth} = \beta_0 + \beta_1 Mission_{cth} + \boldsymbol{X_{th}\beta_2} + \boldsymbol{W_{cth}\beta_3} + \beta_4 Y^b_{icth} + \gamma_b + \varepsilon_{icth} \quad (2)$$

where $Y_{cth}$ now represents teacher *t's* judgment of class section *c* in cohort *h*. $Mission_{cth}$ is an indicator for a class section being randomly assigned to the Challenger Center mission; $\boldsymbol{X_{th}}$ are teacher-level background characteristics, $\boldsymbol{W_{cth}}$ are classroom section aggregates of student background characteristics (described above), $Y^b_{cth}$ is a baseline measure of the engagement outcome, $\gamma_b$ are block fixed effects to account for the randomization scheme, and $\varepsilon_{cth}$ is a stochastic, class section-level error term. Teacher-level covariates included gender, years of experience, race/ethnicity, and an indicator for self-reported certification in science. We cluster standard errors at the teacher level to account for the multiple class ratings per teacher. Estimates are expressed as effects sizes in the form of Hedges' *g*. Given that these analyses are underpowered, we do not explore heterogeneity by teacher or student characteristics.

Our data include some missing baseline survey data (if a student was absent on the day the surveys were given) and background characteristics from the district administrative data. We employ multiple chained imputation (Raghunathan et al., 2001) and include an indicator for treatment status to aid the imputation, as recommended by WWC (2022). We only impute baseline characteristics and do not impute outcome variables. For each sample, we include students (or classes) for which we have all outcomes (i.e., complete-case analysis).

## 5. *Results*

### *5.1 Baseline Balance and Attrition*

Table 1 shows that all but one of the 18 background and baseline variables were balanced between treatment and control conditions. Treatment students expressed poorer attitudes towards STEM careers at baseline, with that difference significant to the 5 percent level. With nearly 20 statistical tests of baseline balance, one would expect one test to yield a statistically significant

results by chance and, and all effects sizes are within WWC boundaries that allow for correction by their inclusion in the models. Our preferred models include all covariates to correct for any imbalance and increase the precision of our estimates. Further, because the treatment group rated themselves lower on attitudes towards science content, correcting for the imbalance would likely increase point estimates, making unconditional estimates a lower bound.

Table 3 presents the attrition statistics for the endline survey sample, which includes student surveys taken after the mission. Student, class, and teacher level attrition were within WWC (2022) cautious boundaries both overall and differentially between treatment and control contrast. At the student level we experienced 17.6 percent overall attrition and a 2.7 percent differential attrition, at the class level we experienced 1.1 percent overall and 2.2 differential attrition, and at the teacher level we experienced 3.1 percent overall and 4.5 percent differential attrition.

Tables A.2 through A.4 show the balance and attrition for the follow-up survey given about six months after the missions and the sample of students with electives choices. The follow-up survey sample has similar balance to the endline survey sample and student responses are once again within WWC cautious boundaries. Slightly more control attrition puts the class and teacher level attrition within the WWC optimistic boundary.[2] The science elective sample has more imbalance, potentially due to the smaller sample size, though attrition is within WWC cautious boundaries for students and classes and the optimistic boundary for teachers. Given the level of imbalance, the elective outcomes should be interpreted cautiously.

[2] As an OST program that occurred during the middle of the year, WWC optimistic boundaries are likely to hold.

### *5.2 Main Results*

Table 4 details the estimated treatment effects of the missions on student attitudes at endline (confirmatory outcomes) and follow-up after about six months (exploratory outcomes) and teacher reports of class engagement (exploratory outcome). We present estimates from unconditional models and those that include student- and class-level covariates.

At endline we find that students report that the missions produced positive and statistically significant effects on two of the six confirmatory student outcomes. Immediately after the mission, students who experienced the missions rate their self-efficacy in science activities higher by 0.118 SD ($p < 0.05$) and their attitudes towards STEM careers higher by 0.159 SD ($p < 0.01$). We did not observe statistically significant treatment effects in student reports of attitudes towards science content, interest in science careers, self-efficacy in science coursework, or STEM identity, though most of the observed differences were consistent with the expected direction of the treatment. Point estimates are generally stable when including covariates, with the exception of attitudes towards STEM careers. That point estimate increases when including covariates, which is expected given that treatment students rated themselves lower than control students on this dimension at baseline. Estimates of the effect of the mission on teacher reports of science class engagement are also positive but not statistically significant. Point estimates become about one third the size after controlling for teacher and class level characteristics. The lack of significance of the results is not surprising as we were not powered to detect effects of this size.

To put our effects into context, at endline the control and treatment group rated themselves 3.4 and 3.5 on a five-point Likert scale, with a three indicating they neither agree nor disagree with statements on their self-efficacy to engage in science activities and a four agreeing with

those statements. The Challenger missions therefore moved students from the midpoint between these two sentiments to closer to agreeing with those sentiments. Likewise at endline, the control and treatment groups rated themselves an average of 4.3 and 4.5 on a seven-point Likert scale where a rating of one indicated sentiments such as STEM jobs were boring, unexciting, and mundane and seven indicated sentiments such as STEM jobs were interesting, exciting, and fascinating. Again, the Challenger missions moved student ratings from close to the midpoint of that scale closer to the more positive end of the scale. Collectively, these results indicate that the two-hour mission experience nudged students from more neutral responses to responses with a more positive valence.

The significant treatment effects we found immediately after the missions faded and lost significance about six months later in follow-up survey. Point estimates for self-efficacy in science activities and attitudes towards STEM careers are about a third to half as large as those from the endline surveys. Other estimates, which were not significant at endline, also diminished in magnitude. Once again, most point estimates are stable with the inclusion of covariates.

Table 5 presents the results of our two confirmatory outcomes on further engagement with STEM. The estimate on student self-report of participating in an afterschool or summer STEM activity is in the unconditional model and attenuates slightly when controlling for covariates; in all models, no point estimate is statistically significant. Effect estimates on seventh grade science elective taking is small in magnitude and statistically insignificant in both specifications. There is little evidence that the missions affected the behavior of students.

Overall, the few short-term positive effects on attitudes we see on endline surveys do not persist after six months, nor do they translate to differences in observed behavior.

### *5.3 Heterogeneity Analyses*

We probe whether the missions were differentially effective for groups that are traditionally underrepresented in STEM coursework and careers, which include girls and students who are a part of racial/ethnic minority communities. We also look at effects by baseline ratings of the outcome measures to understand whether missions build on high baseline attitudes or whether they could inspire students with lower baseline attitudes.

Given the number of statistical tests possible, we limit our analyses to confirmatory outcomes. Even with this limitation, we performed a considerable number of tests. For this reason, and because we are not powered ex ante to detect differences by subgroup, we consider these analyses to be exploratory. In interpreting the effects in Table 6, we focus on general patterns of statistical significance to guide future research on effects of OST programs.

Looking at the two outcomes for which we found positive overall effects, the missions potentially increased self-efficacy in science activities for students who identified as Black by 0.303 SD ($p< 0.05$) and the highest estimates on attitudes towards STEM careers are for students who identify as Hispanic (0.236 SD; $p<0.01$), Black (0.220 SD; $p < 0.10$), and girls (0.184 SD; $p < 0.05$). Attitudes towards STEM careers may be particularly malleable by the missions. We detected effects for boys, non-Hispanic students, White students, and students who scored above and below the median at the 10 percent level or less.

Looking at the outcomes for which we saw no overall effects, we only detected effects on attitudes towards science content in students from underrepresented communities. Girls (0.132 SD, $p < 0.10$), Hispanic students (0.222 SD, $p < 0.01$), and Black (0.255 SD, $p < 0.05$) experience larger effects while their peers saw quantitatively small and statistically insignificant effects. We see few differential effects in the remainder of the outcomes except self-efficacy in science coursework where Hispanic students saw a 0.152 SD ($p < 0.05$) effect and students who

rated themselves in the top half the baseline distribution saw a 0.129 SD effect ($p < 0.10$). The only estimate that is negative and statistically significant, is the effect of the mission on girls' OST participation. We estimate the mission decreased their propensity to participate in these activities by seven percentage points ($p < 0.05$). However, given the number of statistical tests we performed and that this is the sole statistically significant negative result, we do not place a great deal of weight on this finding.

Overall, our results provide suggestive evidence that missions could be particularly effective with populations that are historically underrepresented in STEM coursework and careers. However, there is little evidence that effects could differ by baseline reports of the outcomes, indicating that missions are equally (in)effective for all students regardless of their starting point.

### *5.4 Robustness Check*

Our main models combine two randomization strategies, class-section randomization in six of the seven districts and teacher-level randomization in the seventh district. As a robustness check, we estimate effects separately by randomization strategy and generate an overall effect by calculating a fixed-effect meta-analytic average (WWC, 2022). The intuition is that each randomization scheme will produce unbiased estimates of the effects of the mission which can be considered pulls from the distribution of possible mission effects. A fixed-effect meta-analytic average combines the two estimates to provide an estimate of the overall, underlying effect, giving more weight to the more precisely measured effect.

Table 7 presents the results on our confirmatory outcomes for our models that include covariates. We present results from models analogous to equation 1 run separately by randomization scheme, the meta-analytic average of those effects, and the effects generated by our preferred specification which we presented above. Overall, results from the meta-analytic

averages are largely similar to results from our preferred specification. Looking at the two confirmatory analyses for which we detect significant effects, point estimates from our preferred specification and the meta-analytic average are identical for the attitudes towards STEM careers measure. For the self-efficacy in science activities measure, the meta-analytic average is slightly attenuated compared to our preferred specification. Despite the smaller standard errors in the meta-analytic average, the p-value is 0.054. In this case, the districts that underwent class-level randomization produced a smaller and more precise effect estimate compared to the district that underwent teacher level randomization, which had a larger but less precise effect estimate.[3]

Overall; however, the robustness check leads to similar inferences regarding the effect of the missions on student outcomes.

## 6. *Discussion*

This study serves as the first causal evaluation of Challenger Center missions and adds to a scant body of causal evidence of the effectiveness of STEM OST programs. We conducted a block-clustered RCT to assess whether participation in Challenger Center missions, a two-hour experiential, space-themed OST experience, can affect sixth-grade students' STEM attitudes and beliefs. We found that the Challenger Center missions improved students' self-efficacy in science activities by 0.12 SD and attitudes towards STEM careers by 0.16 SD immediately following the missions. These effects, however, faded after six months and are no longer statistically significant. We also did not find effects on student reports of OST participation in afterschool or summer STEM programs nor on their propensity to choose seventh grade science

[3] The district with teacher randomization produced qualitatively larger effect estimates on four of the eight outcomes. This district participated in both cohorts with 366 total students in the endline analytic sample. An analysis of district specific effects in the six districts that comprise the class randomization estimate in Table 7 show variation in effect estimates across district, though they are very imprecisely estimated. The available evidence therefore suggests natural variation in district-level effects. Given the small sample sizes and imprecise estimates, we do not hypothesize drivers of that possible heterogeneity.

electives. Our heterogeneity analyses put forth suggestive evidence that effects immediately after the mission were more pronounced for girls, students who identified as Hispanic, and students who identified as Black, all of which are traditionally underrepresented in STEM coursework and careers.

One framework which may help interpret our results is Hidi and Renninger's (2006) four-phase model of interest development, for which researchers have provided some empirical evidence of support (Knogler et al., 2015; Rotgans & Schmidt, 2017). In this framework, external stimuli spark a student's interest in a subject or task, called triggered situational interest. The external stimuli can take the form of groupwork, puzzles, social situations, or formal instruction. If the stimuli are fleeting, the situational interest may stop there. If these experiences are supported with external stimuli over time, the student may shift to maintained situational interest and might begin to internalize the interest. Project based learning, tutoring, and other meaningful activities are examples of continued external stimuli. Eventually, the student may begin to have positive dispositions towards the subject or task, accumulate stored knowledge and value, and opt-in to learning when given a choice. This stage of more self-generated interested is called emerging individual interest. External stimuli still play an important role in this stage, and over time as predispositions becomes more relatively enduring, the student may embark on sustained long-term endeavors as the self-generated interest grows. In this last stage, the well-developed individual interest stage, external stimuli can still support this self-generated process.

Applying this framework to our results suggests that the two-hour Challenger Center mission experience is likely a novel external stimulus that provides students with triggered situational interest. After the mission is complete, students likely largely return to traditional STEM instruction. At this stage, because interest is largely maintained externally and students have not

developed an internal drive to seek out more experiences, the effects wane without a sustained novel external stimulus.

Several aspects of our results are in line with this framework. First, the survey measures for which we see effects are conceptualized as short-term outcomes in the Theory of Change in Figure 1. Absent the intervention, the control group reports neutral attitudes with respect to their self-efficacy in doing science activities and attitudes towards STEM career. After practicing STEM activities in an authentic setting, treatment student reports of self-efficacy in doing science activities are moved in the more positive direction. After stepping into a career role and observing other career roles, treatment students express a more positive attitude towards those roles. However, we do not detect effects on measures of longer-term outcomes such as attitudes towards science content, interest in pursuing science careers, self-efficacy in science coursework, and STEM identity. This pattern of results is in line with triggered situational interest stopping short of students internalizing the interest to potentially move those more distal attitudinal measures. Measures of STEM OST participation and science elective choices, which can be seen as more internally initiated behavior, are also unaffected. Additionally, without continued novel learning experiences these short-term effects fade.

The implication for Challenger Center and other OST programs is that their novel settings and learning experiences can move student attitudes and dispositions. However, to be effective in the longer-term, OST programs must provide these experiences for a more extended period. Lauer et al.'s (2006) meta-analysis that shows that effects were larger for math and reading OST programs lasting more than 45 hours and Allen et al.'s (2009) study finding larger gains for programs that lasted at least four weeks supports this notion. While summer programs and afterschool programs may provide a sufficient dosage of experiences to begin to trigger more

lasting individual interest, shorter duration OST programs such as field trips might need to be supplemented with other experiences to maintain any short-term effects they impart. Rotgans and Schmidt (2017) argue that recurring external triggers should vary over time and be organized around one topic of interest to maintain continuity. Knogler et al. (2015) concur that stimuli should revolve around a shared topic and add that each new activity should have their own appeal and tied to the learner's interest. A continuity of learning experiences over time are likely more effective at promoting individual interest.

Specific to Challenger Center, our implementation survey noted that post-mission activities are rarely used. A more uniform implementation of these activities could be a first step to maintain the short term-effects seen in this study. Additionally, CLC virtual programming could follow field trips in an attempt to sustain effects. This virtual programming could serve as avenues to provide varied activities and experiences that revolve around the core theory of action of Challenger Center while minimizing transportation costs. Another avenue of extending interest would be for providers of short-term OST experiences to work together to provide students with several different OST experiences that, collectively, could have lasting effects. Key to this approach is that the OST programs have a shared theory of action so that varied the activities build on the same topic and target the same short-, medium-, and long-term outcomes. In this way, the programs provide the varied activities around the same topic that builds on triggered student interest.

This study, however, shows that even very short-term immersive activities can impact student self-efficacy in science activities and attitudes towards STEM careers. This can then be the foundation for further activities that build on the spark of interest created by OST programs such as Challenger Center.

### 7. *Limitations*

There are several limitations to this study. One threat to internal validity is spillover between treatment and control students. At the teacher level, if any teacher, against our advice, had shared pre- and post-lesson plans with control classrooms (or teachers), it could affect student outcomes. We see this possibility as low and even in cases where materials were shared with control classes or teachers, we hypothesize they would have had limited utility, since those plans are meant solely to prepare students for and debrief them on the mission. At the student level, treatment students may have talked to control students about the inspiration they received from the mission. This is more likely, though we again hypothesize that this type of communication is unlikely to move attitudes and outcomes in an appreciable manner given that the main component of the treatment was its experiential aspect of being in an immersive environment that allowed students to step into real-life STEM roles.

A more germane limitation is that all but the seventh-grade science elective course-taking outcomes came from self-reported surveys. Though we note that we developed all student constructs with questions from validated measures that are commonly used in the field, we acknowledge the possibility of social-desirability bias in responses and differential survey effort. Surveys, however, remain the best available tool to measure student dispositions and attitudes, which are difficult to measure observationally, especially at scale.

***References***

Allen, P. J., Chang, R., Gorrall, B. K., Waggenspack, L., Fukuda, E., Little, T. D., & Noam, G. G. (2019). From quality to outcomes: A national study of afterschool STEM programming. International Journal of STEM Education, 6(37), 1–21. https://doi.org/10.1186/s40594-019-0191-2

Allen, P. J., & Noam, G. G. (2024). *A systematic review of STEM and social-emotional development in out-of-school time programs: Executive summary.* Institute for the Study of Resilience in Youth (NSF Award #2115868).

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus: How families shape children's engagement and identification with science. *American Educational Research Journal, 49*(5), 881–908. https://doi.org/10.3102/0002831211433290

Barab, S., & Dede, C. (2007). Games and immersive participatory simulations for science education: An emerging type of curricula. *Journal of Science Education and Technology, 16*(1), 1–3. https://doi.org/10.1007/s10956-007-9043-9

Barton, A.C., Tan, E., & Rivet, A. (2008). Creating hybrid spaces for engaging school science among urban middle school girls. *American Educational Research Journal, 45*(1), 68–103. https://doi.org/10.3102/0002831207308641

Cantrell, P., & Sudweeks, R. (2009). Technology task autonomy and gender effects on student performance in rural middle school science classrooms. *Journal of Computers in Mathematics and Science Teaching, 28*(4), 359–379.

Challenger Center. (2026). *Home.* Retrieved September 25, 2026, from https://challenger.org/

Chen, C.-H., & Howard, B. (2010). Effect of live simulation on middle school students' attitudes and learning toward science. *Educational Technology & Society, 13*(1), 133–139.

Crawford, L., & Huscroft-D'Angelo, J. (2015). Mission to space: Evaluating one type of informal science education. *Electronic Journal of Science Education, 19*(1).

Dabney, K. P., Tai, R. H., Almarode, J. T., Miller-Friedmann, J. L., Sonnert, G., Sadler, P. M., & Hazari, Z. (2012). Out-of-school time science activities and their association with career interest in STEM. *International Journal of Science Education, Part B, 2*(1), 63–79. https://doi.org/10.1080/21548455.2011.629455

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. Educational Psychologist, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Jarvis, T., & Pell, A. (2002). Effect of the Challenger experience on elementary children's attitudes to science. *Journal of Research in Science Teaching, 39*(10), 979–1000. https://doi.org/10.1002/tea.10055

Kim, H. (2011). Inquiry-based science and technology enrichment program: Green earth enhanced with inquiry and technology. *Journal of Science Education and Technology, 20*(6), 803–814. https://doi.org/10.1007/s10956-011-9334-z

Kisida, B., Bowen, D. H., & Greene, J. P. (2016). Measuring critical thinking: Results from an art museum field trip experiment. *Journal of Research on Educational Effectiveness, 9*(S1), 171–187. https://doi.org/10.1080/19345747.2015.1086915

Knogler, M., Harackiewicz, J. M., Gegenfurtner, A., & Lewalter, D. (2015). How situational is situational interest? Investigating the longitudinal structure of situational interest. *Contemporary Educational Psychology, 43*, 39–50. https://doi.org/10.1016/j.cedpsych.2015.08.004

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*(2), 275–313.

Lunce, L. M. (2006). Simulations: Bringing the benefits of situated learning to the traditional classroom. *Journal of Applied Educational Technology, 3*(1), 37–45.

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education, 95*(5), 877–907. https://doi.org/10.1002/sce.20441

McCombs, J.S., Whitaker, A.A., and Yoo, P.Y. (2017). The value of out-of-school time programs. *RAND*. Retrieved from: https://www.rand.org/pubs/perspectives/PE267.html

Meluso, A., Zheng, M., Spires, H. A., & Lester, J. (2012). Enhancing 5th graders' science content knowledge and self-efficacy through game-based learning. *Computers & Education, 59*(2), 497–504. https://doi.org/10.1016/j.compedu.2011.12.019

National Assessment Governing Board. (2008). *Science framework for the 2009 National Assessment of Educational Progress.* U.S. Department of Education. https://www.nagb.gov

National Center for Education Statistics. (2024). *NAEP report card: Science.* U.S. Department of Education, Institute of Education Sciences. Retrieved September 11, 2026, from https://www.nationsreportcard.gov/reports/science/2024/

National Research Council. (2015). *Identifying and supporting productive STEM programs in out-of-school settings.* The National Academies Press. https://doi.org/10.17226/21740

National Science Board. (2026). *The state of U.S. science and engineering 2026* (NSB-SEP-2026-1). Science and engineering indicators. National Science Board, National Science Foundation. https://ncses.nsf.gov/indicators

Newmann, F. M. (Ed.). (1992). *Student engagement and achievement in American secondary schools.* Teachers College Press.

Nugent, G., Barker, B., White, A., & Grandgenett, N. (2011). The impact of robotics competitions on youth STEM learning, attitudes and 21st century workplace skills. In T. Bastiaens & M. Ebner (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2011* (pp. 3614–3619). Association for the Advancement of Computing in Education (AACE).

Raybourn, E. M. (2014). A new paradigm for serious games: Transmedia learning for more effective training and education. *Journal of Computational Science, 5*(3), 471–481. https://doi.org/10.1016/j.jocs.2013.08.005

Redmond, A., Thomas, J., High, K., Scott, M., Jordan, P., & Dockers, J. (2011). Enriching science and math through engineering. *School Science and Mathematics, 111*(8), 399–408.

Rotgans, J. I., & Schmidt, H. G. (2017). The relation between individual interest and knowledge acquisition. British Educational Research Journal, 43(2), 350–371. http://www.jstor.org/stable/44954831

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education, 58*(1), 136–153. https://doi.org/10.1016/j.compedu.2011.07.017

Saw, G. K. (2020). Leveraging social capital to broaden participation in STEM. *Policy Insights from the Behavioral and Brain Sciences, 7*(1), 35–43. https://doi.org/10.1177/2372732219895997

Saw, G. K., Swagerty, B., Brewington, S., Chang, C.-N., & Culbertson, R. (2019). Out-of-school time STEM program: Students' attitudes toward and career interests in mathematics and science. *International Journal of Evaluation and Research in Education, 8*(2), 356–362. https://doi.org/10.11591/ijere.v8i2.18702

Sonnert, G., Reid, T., Sunbury, S., & Sadler, P. M. (2025). How do formal and informal science learning experiences during high school shape students' career interest and STEM identity? *International Journal of STEM Education, 12,* Article 55. https://doi.org/10.1186/s40594-025-00568-w

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science, 312*(5777), 1143–1144. https://doi.org/10.1126/science.1128690

Thisgaard, M., & Makransky, G. (2017). Virtual learning simulations in high school: Effects on cognitive and non-cognitive outcomes and implications on the development of STEM academic and career choice. *Frontiers in Psychology, 8,* Article 805. https://doi.org/10.3389/fpsyg.2017.00805

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0* (WWC 2022008REV). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Handbooks

Whitesell, E. R. (2016). A day at the museum: The impact of field trips on middle school science achievement. *Journal of Research in Science Teaching, 53*(7), 1036–1054. https://doi.org/10.1002/tea.21322

Wills, S., Devonshire, E., Leigh, E., Rosser, E., Shepherd, J., & Vincent, A. (2007). Encouraging role based online learning environments. In *ICT: Providing choices for learners and learning: Proceedings ascilite Singapore 2007* (pp. 1093–1098). https://doi.org/10.65106/apubs.2007.2870

Young, J. R., Ortiz, N. A., & Young, J. L. (2017). STEMulating interest: A meta-analysis of the effects of out-of-school time on student STEM interest. *International Journal of Education in Mathematics, Science and Technology, 5*(1), 62–74. https://doi.org/10.18404/ijemst.61149

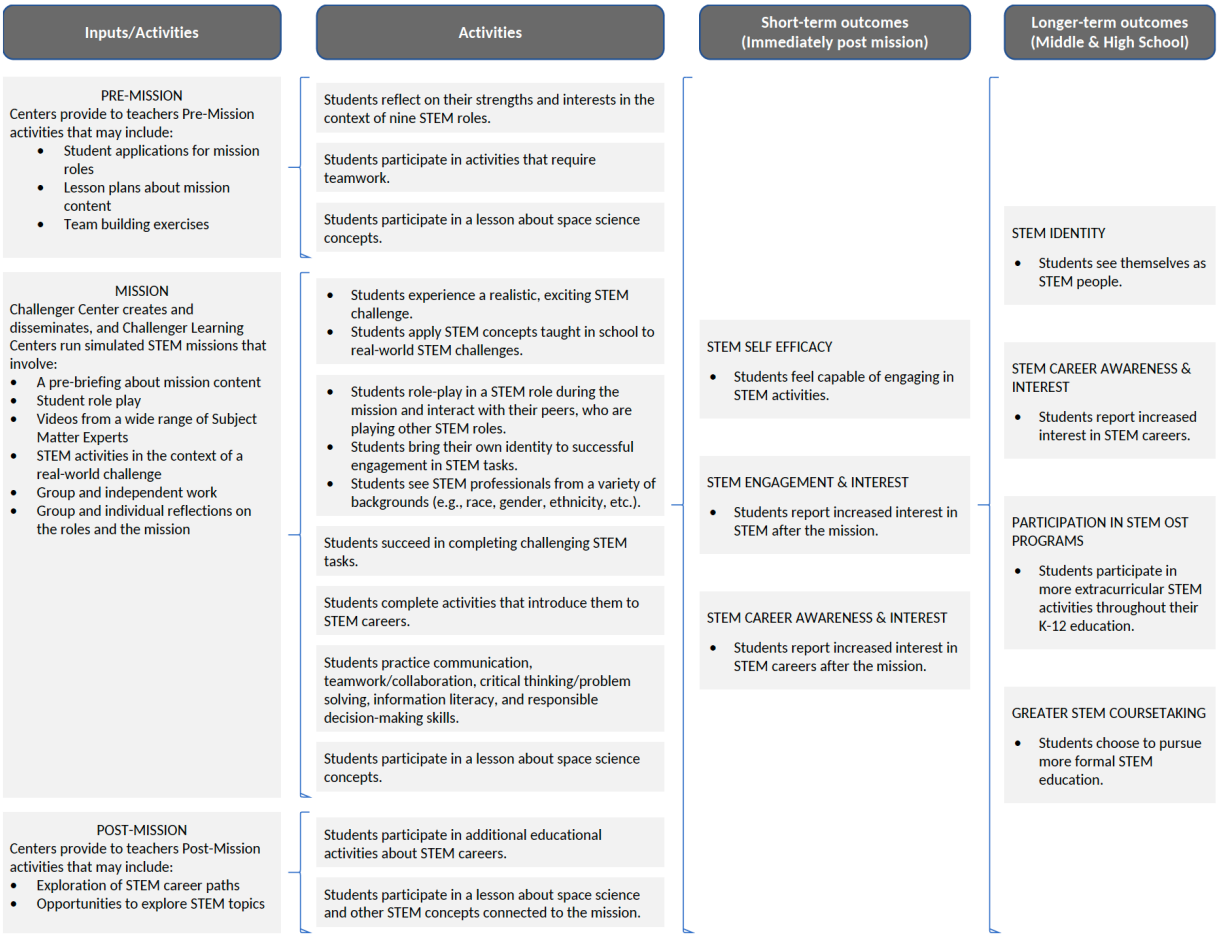
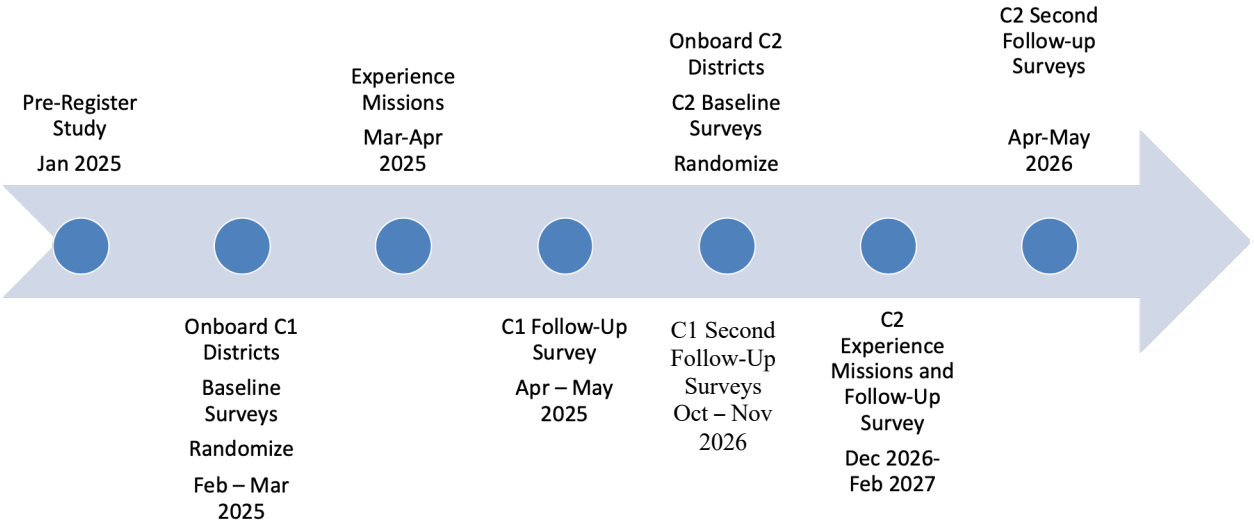



**Figure 1: Challenger Center Theory of Change**




Note: C1 = Cohort 1; C2 = Cohort 2

**Figure 2: Timeline of the research activities**

**Table 1: Descriptive Statistics and Balance: Endline Survey Sample**

| Characteristic | N (Treat) | Treat Average | N (Control) | Control Average | Hedges g/ Cox Index (T-C Difference) | P-value (T-C Difference) |
|---|---|---|---|---|---|---|
| Demographics | | | | | | |
| Age (Years) | 741 | 11.75 | 681 | 11.76 | -0.03 | 0.51 |
| Female | 737 | 0.5 | 678 | 0.48 | 0.06 | 0.39 |
| Hispanic | 741 | 0.31 | 681 | 0.38 | -0.03 | 0.71 |
| White | 741 | 0.51 | 678 | 0.5 | -0.05 | 0.56 |
| Black | 741 | 0.2 | 678 | 0.17 | 0.08 | 0.60 |
| Other | 741 | 0.28 | 678 | 0.33 | 0.00 | 0.97 |
| Special Education | 687 | 0.13 | 633 | 0.16 | -0.19 | 0.14 |
| Free/Reduced Priced Lunch | 687 | 0.73 | 633 | 0.73 | 0.06 | 0.48 |
| English Language Learner | 687 | 0.13 | 633 | 0.16 | -0.03 | 0.82 |
| Baseline Variables | | | | | | |
| Afterschool or Summer STEM Activity | 721 | 0.13 | 665 | 0.15 | -0.11 | 0.27 |
| Self-Efficacy in Science Activities | 712 | 3.56 | 655 | 3.57 | -0.03 | 0.61 |
| Self-Efficacy in Science Coursework | 696 | 3.45 | 647 | 3.43 | 0.04 | 0.5 |
| Science Career Interest | 698 | 3.07 | 646 | 3.02 | 0.06 | 0.27 |
| Attitudes Towards Science Content | 692 | 4.53 | 641 | 4.54 | 0.00 | 0.95 |
| Attitudes Towards STEM Career | 684 | 4.38 | 638 | 4.54 | -0.11 | 0.02* |
| STEM Identity | 700 | 3.56 | 651 | 3.65 | -0.04 | 0.39 |
| Fifth Grade Math (standardized) | 640 | 0.05 | 594 | 0.08 | -0.03 | 0.67 |
| Fifth Grade ELA (standardized) | 637 | 0.1 | 591 | 0.05 | 0.05 | 0.45 |

Notes: Samples change due to missing data. Hedges *g*/ Cox indices and p-values are derived from regression models regressing the covariate denoted in the row headers on an indicator for treatment and block fixed effects. Standard errors are clustered by class. * indicates $p < 0.05$

**Table 2: Survey Measures by Domain**

| | Domain | Source | Example | Cronbach's Alpha |
|---|---|---|---|---|
| 1 | Self-Efficacy in Science Activities | Self-Efficacy for Learning and Doing Science Scales (Porticella, Phillips, and Bonney, 2017) | Compared to other people my age, I think I can do science activities pretty well | 0.60-0.68 |
| 2 | Self-Efficacy in Science Coursework | Student Attitudes Towards STEM Survey (Faber, Townsend, and Weibe, 2013) | I am sure I can do advanced work in science | 0.68-0.73 |
| 3 | Interest in Science Careers | | I would consider a career in science | 0.82-0.89 |
| 4 | Attitudes Towards Science Content | STEM Semantics Survey (Tyler-Wood, Knezek, and Christensen, 2010) | To me, science is boring | 0.78-0.84 |
| 5 | Attitudes Towards STEM Careers | | To me a job in STEM is boring | 0.81-0.85 |
| 6 | STEM Identity | Single-item measure of STEM identity (McDonald et al., 2019) | Venn diagrams as representation of self and scientist | N/A |
| 7 | Teacher Ratings of Class Science Engagement | Math and Science Engagement Scales (Wang et al., 2016) | What proportion of your class puts effort into learning science? | 0.87-0.89 |

**Table 3: Student, Teacher, and Class Level Attrition, Endline Survey Sample**

| | Treatment | Control | Total |
|---|---|---|---|
| Student Level | | | |
| Randomized | 903 | 858 | 1761 |
| Sample | 756 | 695 | 1451 |
| Attrition (Count) | 147 | 163 | 310 |
| Attrition (percent) | 16.3 | 19.0 | 17.6 |
| Differential Attrition | 2.7 | | |
| Class Level | | | |
| Randomized | 46 | 45 | 91 |
| Sample | 46 | 44 | 90 |
| Attrition (Count) | 0 | 1 | 1 |
| Attrition (percent) | 0.0 | 2.2 | 1.1 |
| Differential Attrition | 2.2 | | |
| Teacher Level | | | |
| Randomized | 24 | 22 | 46 |
| Sample | 24 | 21 | 1 |
| Attrition (Count) | 0 | 1 | 1 |
| Attrition (percent) | 0.0 | 4.5 | 2.2 |
| Differential Attrition | 4.5 | | |

Note: Teacher counts are greater than the number of teachers because randomization in six of the seven cohorts was class section within teacher.

**Table 4: Effect of Missions on Student Attitudes, Endline and Follow-up Surveys**

| | Endline | | Follow-Up | |
|---|---|---|---|---|
| Self-Efficacy in Science Activities | 0.106+ | 0.118* | 0.054 | 0.046 |
| | (0.055) | (0.051) | (0.057) | (0.038) |
| Attitudes Towards Science Content | 0.068 | 0.065 | 0.071 | 0.045 |
| | (0.053) | (0.050) | (0.056) | (0.051) |
| Attitudes Towards STEM Careers | 0.087 | 0.159** | 0.010 | 0.080 |
| | (0.053) | (0.052) | (0.056) | (0.054) |
| Interest in Science Careers | 0.042 | 0.045 | -0.012 | -0.031 |
| | (0.057) | (0.050) | (0.062) | (0.042) |
| Self-Efficacy in Science Coursework | 0.074 | 0.067 | 0.014 | -0.003 |
| | (0.056) | (0.046) | (0.055) | (0.048) |
| STEM Identity | -0.004 | -0.015 | -0.023 | -0.008 |
| | (0.051) | (0.050) | (0.059) | (0.052) |
| Teacher Reports of Class Engagement | 0.308 | 0.120 | N/A | N/A |
| | (0.120) | (0.097) | | |
| Covariates | | ✓ | | ✓ |

Notes: Estimates are expressed as Hedges g. Each cell represents and estimates from a regression of the outcome, as indicated by the row header, on an indicator for being randomly assigned to treatment and block fixed effects. Data collection survey wave is indicated by column headers. Covariates are additionally included in models as indicated by the checkmark. Covariates include: student age, gender, race/ethnicity, special education status, free/reduced priced lunch status, English language learner status, baseline summer and afterschool science participation status, a self-report indicator of completing any Challenger Center related activity prior to the study, fifth grade math and English language arts scores (standardized), baseline measure of the outcome, and class aggregates of these variables. Standard errors are clustered at the class-year level. N = 1,451 for endline survey outcomes and 1,276 for follow-up survey outcomes. + indicates $p < 0.10$; * $p < 0.05$; ** $p < 0.01$

**Table 5: Effect of Missions on Student STEM Activities and Science Electives**

| | | |
|---|---|---|
| Afterschool or Summer STEM Activity | -0.034 | -0.021 |
| | (0.021) | (0.022) |
| Seventh Grade Science Elective | 0.005 | -0.002 |
| | (0.019) | (0.020) |
| Covariates | | ✓ |

Notes: Each cell represents and estimates from a linear probability model where the outcome, as indicated by the row header, on an indicator for being randomly assigned to treatment and block fixed effects. Afterschool and summer STEM activities were collected as part of the follow-up surveys Science elective data came from district administrative data. Covariates are additionally included in models as indicated by the checkmark. Covariates include: student age, gender, race/ethnicity, special education status, free/reduced priced lunch status, English language learner status, baseline summer and afterschool science participation status, a self-report indicator of completing any Challenger Center related activity prior to the study, fifth grade math and English language arts scores (standardized), and class aggregates of these variables. Standard errors are clustered at the class-year level. N = 1,276 for afterschool and summer STEM activities and 1,170 for science elective outcomes. + indicates $p < 0.10$

**Table 6: Heterogeneity Analysis, Confirmatory Outcomes**

| | Female | Male | Hispanic | Non-Hispanic | Black | White | Other | Above Baseline Median | Below Baseline Median |
|---|---|---|---|---|---|---|---|---|---|
| Self-Efficacy in Science Activities | 0.122+ | 0.113+ | 0.119 | 0.118+ | 0.303* | 0.059 | 0.104 | 0.138+ | 0.146* |
| | (0.065) | (0.069) | (0.084) | (0.061) | (0.133) | (0.065) | (0.089) | (0.078) | (0.073) |
| Attitudes Towards Science Content | 0.132+ | 0.007 | 0.222** | -0.008 | 0.255* | 0.015 | 0.023 | 0.069 | 0.091 |
| | (0.073) | (0.068) | (0.08) | (0.061) | (0.113) | (0.066) | (0.077) | (0.078) | (0.078) |
| Attitudes Towards STEM Careers | 0.184* | 0.136+ | 0.236** | 0.122+ | 0.220+ | 0.171* | 0.096 | 0.199* | 0.168+ |
| | (0.079) | (0.075) | (0.088) | (0.064) | (0.123) | (0.073) | (0.095) | (0.076) | (0.086) |
| Interest in Science Careers | 0.04 | 0.05 | 0.053 | 0.042 | 0.179 | 0.031 | -0.019 | 0.053 | 0.053 |
| | (0.068) | (0.065) | (0.097) | (0.057) | (0.118) | (0.063) | (0.097) | (0.079) | (0.082) |
| Self-Efficacy in Science Coursework | 0.097 | 0.039 | 0.152* | 0.027 | 0.135 | 0.041 | 0.067 | 0.129+ | 0.009 |
| | (0.068) | (0.061) | (0.075) | (0.055) | (0.118) | (0.056) | (0.081) | (0.07) | (0.067) |
| STEM Identity | -0.014 | -0.016 | 0.035 | -0.042 | 0.143 | -0.012 | -0.12 | -0.019 | 0.000 |
| | (0.069) | (0.076) | (0.089) | (0.061) | (0.125) | (0.063) | (0.096) | (0.086) | (0.074) |
| N | 689 | 726 | 487 | 935 | 268 | 735 | 433 | 569-665 | 662-774 |
| Afterschool or Summer STEM Activities | -0.070* | 0.023 | -0.028 | -0.026 | -0.026 | -0.009 | -0.041 | N/A | N/A |
| | (0.031) | (0.025) | (0.025) | (0.056) | (0.056) | (0.023) | (0.029) | | |
| N | 597 | 669 | 382 | 890 | 256 | 660 | 352 | | |
| Seventh Grade Science Electives | -0.018 | 0.014 | -0.039 | 0.020 | 0.044 | -0.033 | -0.003 | N/A | N/A |
| | (0.031) | (0.035) | (0.031) | (0.027) | (0.038) | (0.030) | (0.031) | | |
| N | 572 | 598 | 390 | 780 | 330 | 463 | 374 | | |

Notes: Effects on afterschool and summer STEM activities and seventh grade electives are expressed in percentage points. All other effects are expressed as Hedges *g*. Each cell represents and estimates from a regression of the outcome, as indicated by the row header, on an indicator for being randomly assigned to treatment, block fixed effects and the following covariates: student age, gender, race/ethnicity, special education status, free/reduced priced lunch status, English language learner status, baseline summer and afterschool science participation status, a self-report indicator of completing any Challenger Center related activity prior to the study, fifth grade math and English language arts scores (standardized), baseline measure of the outcome, and class aggregates of these variables. Standard errors are clustered at the class-year level. + indicates $p < 0.10$; * $p < 0.05$; ** $p < 0.01$

**Table 7: Robustness Check, Effects by Randomization Scheme and Meta-Analytic Average**

| | Class Randomization | Teacher Randomization | Meta-Analytic Average | Preferred Specification |
|---|---|---|---|---|
| Self-Efficacy in STEM Activities | 0.055 | 0.249* | 0.085+ | 0.118* |
| | (0.048) | (0.113) | (0.044) | (0.051) |
| Attitudes Towards STEM Content | 0.016 | 0.201* | 0.058 | 0.065 |
| | (0.053) | (0.098) | (0.047) | (0.050) |
| Attitudes Towards STEM Careers | 0.136* | 0.242* | 0.159** | 0.159** |
| | (0.057) | (0.110) | (0.051) | (0.052) |
| Interest in STEM Careers | -0.025 | 0.118 | -0.003 | 0.045 |
| | (0.047) | (0.109) | (0.043) | (0.050) |
| Self-Efficacy in STEM Coursework | 0.022 | 0.053 | 0.027 | 0.067 |
| | (0.048) | (0.104) | (0.043) | (0.046) |
| STEM Identity | -0.038 | 0.037 | -0.079 | -0.015 |
| | (0.057) | (0.109) | (0.052) | (0.050) |
| Afterschool or Summer Activities | -0.032+ | 0.060 | -0.025 | -0.021 |
| | (0.018) | (0.063) | (0.017) | (0.022) |
| Seventh Grade Science Elective | -0.056 | -0.008 | -0.021 | -0.002 |
| | (0.038) | (0.024) | (0.020) | (0.020) |

Notes: Effects on afterschool and summer STEM activities and seventh grade electives are expressed in percentage points. Each cell represents and estimates from a regression of the outcome, as indicated by the row header, on an indicator for being randomly assigned to treatment and block fixed effects. Covariates include: student age, gender, race/ethnicity, special education status, free/reduced priced lunch status, English language learner status, baseline summer and afterschool science participation status, a self-report indicator of completing any Challenger Center related activity prior to the study, fifth grade math and English language arts scores (standardized), baseline measure of the outcome, and class aggregates of these variables. Standard errors are clustered at the class-year level in class randomization and our preferred specification and teacher-year in the teacher randomization models. N = 1,451 for endline survey outcomes and 1,276 for follow-up survey outcomes. + indicates $p < 0.10$; * $p < 0.05$; ** $p < 0.01$

*Online Appendix*

**Table A.1: Descriptive Statistics and Balance: Follow-up Survey Sample**

| Characteristic | N (Treat) | Treat Average | N (Control) | Control Average | Hedges g/ Cox Index (T-C Difference) | P-value (T-C Difference) |
|---|---|---|---|---|---|---|
| Demographics | | | | | | |
| Age (Years) | 656 | 11.77 | 616 | 11.76 | 0.01 | 0.78 |
| Female | 652 | 0.48 | 614 | 0.46 | 0.06 | 0.39 |
| Hispanic | 656 | 0.28 | 616 | 0.32 | -0.07 | 0.45 |
| White | 656 | 0.54 | 615 | 0.50 | -0.05 | 0.61 |
| Black | 656 | 0.21 | 615 | 0.19 | 0.14 | 0.30 |
| Other | 656 | 0.25 | 615 | 0.31 | -0.06 | 0.52 |
| Special Education | 611 | 0.13 | 572 | 0.17 | -0.25 | 0.03* |
| Free/Reduced Priced Lunch | 611 | 0.73 | 572 | 0.70 | 0.08 | 0.30 |
| English Language Learner | 611 | 0.11 | 572 | 0.12 | -0.01 | 0.94 |
| Baseline Variables | | | | | | |
| Afterschool or Summer STEM Program | 628 | 0.13 | 595 | 0.14 | -0.07 | 0.45 |
| Self-Efficacy in Science Activities | 618 | 3.56 | 584 | 3.57 | -0.02 | 0.69 |
| Self-Efficacy in Science Coursework | 604 | 3.45 | 571 | 3.43 | 0.03 | 0.58 |
| Science Career Interest | 609 | 3.03 | 572 | 2.99 | 0.05 | 0.45 |
| Towards Science Content | 601 | 4.58 | 571 | 4.55 | 0.02 | 0.67 |
| Towards STEM Career | 596 | 4.37 | 566 | 4.56 | -0.13 | 0.01* |
| STEM Identity | 609 | 3.52 | 578 | 3.58 | -0.03 | 0.54 |
| Fifth Grade Math (standardized) | 572 | 0.05 | 542 | 0.11 | -0.06 | 0.50 |
| Fifth Grade ELA (standardized) | 569 | 0.10 | 540 | 0.05 | 0.05 | 0.43 |

Notes: Samples change due to missing data. Hedges *g* and p-values are derived from regression models regressing the covariate denoted in the row headers on an indicator for treatment and block fixed effects. Standard errors are clustered by class. * indicates $p < 0.05$

**Table A.2: Student, Teacher, and Class Level Attrition, Follow-Up Survey Sample**

| | Treatment | Control | Total |
|---|---|---|---|
| Student Level | | | |
| Randomized | 903 | 858 | 1761 |
| Sample | 656 | 620 | 1276 |
| Attrition (Count) | 247 | 238 | 485 |
| Attrition (percent) | 27.4 | 27.7 | 27.5 |
| Differential Attrition | 0.3 | | |
| Class Level | | | |
| Randomized | 46 | 45 | 91 |
| Sample | 46 | 42 | 88 |
| Attrition (Count) | 0 | 3 | 3 |
| Attrition (percent) | 0 | 6.7 | 3.3 |
| Differential Attrition | 6.7 | | |
| Teacher Level | | | |
| Randomized | 24 | 22 | 46 |
| Sample | 24 | 20 | 44 |
| Attrition (Count) | 0 | 2 | 2 |
| Attrition (percent) | 0 | 9.1 | 4.3 |
| Differential Attrition | 9.1 | | |

Note: Teacher counts are greater than the number of teachers because randomization in six of the seven cohorts was class section within teacher.

**Table A.3: Descriptive Statistics and Balance: Science Elective Sample**

| Characteristic | N (Treat) | Treat Average | N (Control) | Control Average | Hedges g/ Cox Index (T-C Difference) | P-value (T-C Difference) |
|---|---|---|---|---|---|---|
| Demographics | | | | | | |
| Age (Years) | 586 | 11.76 | 584 | 11.78 | -0.03 | 0.56 |
| Female | 586 | 0.51 | 584 | 0.47 | 0.09 | 0.30 |
| Hispanic | 586 | 0.30 | 584 | 0.36 | 0.07 | 0.44 |
| White | 586 | 0.39 | 584 | 0.40 | -0.12 | 0.34 |
| Black | 586 | 0.32 | 584 | 0.24 | 0.07 | 0.68 |
| Other | 586 | 0.29 | 584 | 0.35 | 0.07 | 0.52 |
| Special Education | 586 | 0.10 | 584 | 0.19 | -0.48 | 0.00** |
| Free/Reduced Priced Lunch | 586 | 0.77 | 584 | 0.75 | 0.17 | 0.10+ |
| English Language Learner | 586 | 0.11 | 584 | 0.12 | 0.18 | 0.14 |
| Baseline Variables | | | | | | |
| Afterschool or Summer STEM Program | 546 | 0.14 | 555 | 0.13 | 0.07 | 0.54 |
| Self-Efficacy in Science Activities | 534 | 3.53 | 547 | 3.55 | -0.02 | 0.70 |
| Self-Efficacy in Science Coursework | 516 | 3.42 | 539 | 3.41 | 0.01 | 0.90 |
| Science Career Interest | 519 | 3.05 | 538 | 3.05 | 0.00 | 0.99 |
| Towards Science Content | 513 | 4.49 | 524 | 4.57 | -0.06 | 0.30 |
| Towards STEM Career | 504 | 4.35 | 522 | 4.59 | -0.17 | 0.00** |
| STEM Identity | 522 | 3.60 | 542 | 3.83 | -0.12 | 0.05* |
| Fifth Grade Math (standardized) | 549 | -0.04 | 552 | 0.03 | -0.07 | 0.48 |
| Fifth Grade ELA (standardized) | 549 | 0.03 | 552 | -0.03 | 0.06 | 0.47 |

Notes: Samples change due to missing data. Hedges *g* and p-values are derived from regression models regressing the covariate denoted in the row headers on an indicator for treatment and block fixed effects. Standard errors are clustered by class. * indicates $p < 0.05$

**Table A.4: Student, Teacher, and Class Level Attrition, Science Elective Sample**

| | Treatment | Control | Total |
|---|---|---|---|
| Student Level | | | |
| Randomized | 903 | 858 | 1761 |
| Sample | 586 | 584 | 1170 |
| Attrition (Count) | 317 | 274 | 591 |
| Attrition (percent) | 35.1 | 31.9 | 33.6 |
| Differential Attrition | 3.2 | | |
| Class Level | | | |
| Randomized | 46 | 45 | 91 |
| Sample | 30 | 30 | 60 |
| Attrition (Count) | 16 | 15 | 31 |
| Attrition (percent) | 34.8 | 33.3 | 34.1 |
| Differential Attrition | 1.4 | | |
| Teacher Level | | | |
| Randomized | 24 | 22 | 46 |
| Sample | 17 | 14 | 31 |
| Attrition (Count) | 7 | 8 | 15 |
| Attrition (percent) | 29.2 | 36.4 | 32.6 |
| Differential Attrition | 7.2 | | |

Note: Teacher counts are greater than the number of teachers because randomization in six of the seven cohorts was class section within teacher.