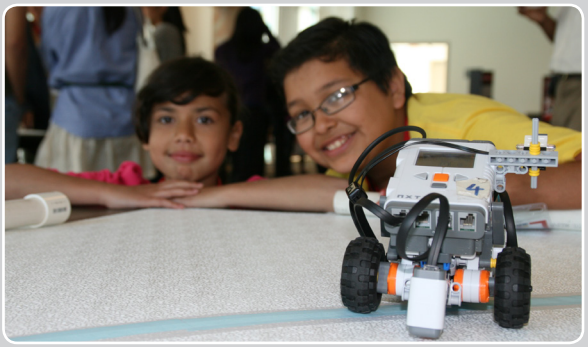
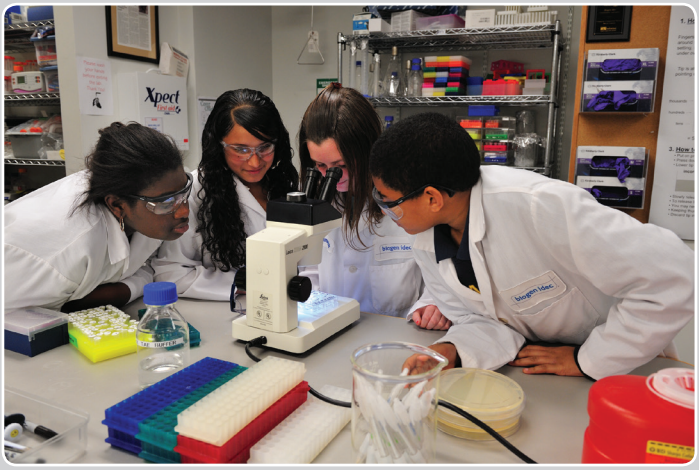




Closing Inspiration and Achievement Gaps in STEM with Volunteer-Led Apprenticeships

SEPTEMBER 27, 2018

PREPARED FOR:
Citizen Schools
308 Congress St., 5th Floor
Boston, MA 02210

SUBMITTED BY:
Alyssa Rulf Fountain
Melissa Velez
Beth Gamse
Abt Associates
10 Fawcett Street
Cambridge, MA 02138

Table of Contents

1. Introduction

Citizen Schools was awarded an Investing in Innovation (i3) grant in 2012 with the goal of expanding and developing innovative practices that could serve as models of best practices. Exploratory analyses from a prior national evaluation of the Citizen Schools expanded learning time (ELT) model suggested that the Citizen Schools ELT model had a positive impact on math achievement for selected subgroups (Rulf Fountain et al, 2016). The i3 grant funding built on the earlier study by focusing explicitly on program elements most directly related to student outcomes in science, technology, engineering, and math (STEM). The i3 grant was intended to better understand impacts on student outcomes of one core component of Citizen Schools' ELT model – STEM apprenticeships.

Citizen Schools provides additional learning opportunities for middle school students by expanding the school day–not only to support and extend the instruction provided during the regular school day, but to engage students in hands-on, project-based learning. Citizen Schools uses community volunteers, the majority of whom have STEM-related backgrounds and jobs, to co-teach apprenticeships with Citizen Schools staff (AmeriCorps members called 'Teaching Fellows'). The Citizen Schools apprenticeships are explicitly designed to increase student exposure, engagement, and interest in STEM subjects. The apprenticeship experience, in particular, is intended to address what seems to be a persistent and pervasive apathy toward STEM learning.

The vast majority of American students are neither prepared nor sufficiently engaged to become science, technology, engineering, and mathematics (STEM)-literate citizens or innovative STEM professionals (National Research Council, 2007; Carnegie Corporation and the Institute for Advanced Study, 2009; National Science Board [NSB], 2010; President's Council of Advisors on Science and Technology [PCAST], 2010). Evidence from the 2012 Program for International Student Assessment (PISA) assessment places the US in the bottom third in science (20th of 34 OEPD nations), and bottom fourth in mathematics achievement (27th of 34) (PISA, 2012).[1] Scores from the National Assessment of Educational Progress (NAEP) indicate that less than one-third of U.S. eighth graders show proficiency in mathematics and science, with African-American, Hispanic, and Native American students consistently underperforming compared to their white peers (National Center for Education Statistics [NCES], 2013), and under 9 percent of US 15-year old students were top performers (level 5 or 6) as measured by PISA (PISA, 2013). These results clearly indicate that American students are ill-prepared for advanced scientific training or the more rigorous STEM courses necessary to pursue post-secondary education and/or careers in the STEM fields.

The challenge of developing STEM-literate citizens and building the STEM professional pipeline in the United States is multifaceted, however, and extends beyond lack of academic preparation or achievement. Evidence about students' interest in science—which often predicts students' pursuit of science-related careers—is a critical part of the puzzle. Too many secondary school students report that they dismiss STEM career possibilities because they neither know people who work in STEM areas nor understand what such people do (Tai et al., 2006; Lemelson, 2010). The interest gap is particularly severe among

[1] Evidence from the most recent TIMMS assessments suggest somewhat better results for 4th and 8th grade students; American 4th graders' scores were in the top third, and 8th graders' scores in the top half for both math and science, respectively (see http://nces.ed.gov/timss/results11.asp).

girls and minorities who are far less likely to pursue postsecondary coursework or graduate with STEM degrees than their white and/or male counterparts (Higher Education Research Institute Research Brief, 2010). Increasing students' interest in STEM is an essential step in increasing their subsequent pursuit of STEM education and careers as well as the general competency expected of U.S. citizens in the 21st century workforce (NSB, 2010, 2014).

In today's technological and global society, STEM disciplines are viewed as fundamental to the nation's economic growth and prosperity. Employment opportunities in STEM fields have increased at a faster rate than in non-STEM fields (Government Accounting Office [GAO], 2006). Additionally, even professions in agricultural and law fields historically considered as distinct from STEM-related, are increasingly requiring scientific and technological proficiency (PCAST, 2010). As American students continue to underperform in math and science compared to their international peers, concerns have arisen about America's economic and technological competitiveness. Further, a large share of U.S. science degrees are awarded to people born abroad (Borjas, 2005), and America's dependence on foreign-born and foreign-trained scientists is on the rise (Xie & Achen, 2009). Such trends have sparked political interest in uncovering means to prepare and engage students in STEM fields more effectively. In 2013, a report from the White House Office of Science and Technology Policy called for "a concerted and inclusive effort to ensure that the STEM workforce is equipped with the skills and training needed to excel in these fields" to sustain the capable and flexible workforce necessary to compete in the global marketplace (National Science and Technology Council, 2013).

How can we, as a nation, better engage our students—and our citizenry—in the pursuit of knowledge and learning about STEM subjects? There are numerous large-scale efforts underway, which include: projects to improve the quality of the teaching force (e.g., 100Kin10, Robert Noyce Scholarship Program[2]); efforts to increase the relevance of instructional materials (e.g., the Common Core State Standards and Next Generation Science Standards); presidential initiatives (Educate to Innovate, FIRST, First in the World, America Competes, Youth Career Connect); and a wide array of statewide and community-based efforts (e.g., Massachusetts High Tech Council, local Community-based Science Projects sponsored by the National Science Teachers Association).

Policy interest has extended to the role of professionals in providing introductions to STEM content, experiences, or learning to K-12 students. For example, the National Academies of Science issued a report about integration of STEM into K-12 education, and convened an invited workshop summarizing best practices in informal science, and the GAO issued a report on connections between STEM education and the workforce.[3] In 2013, President Obama announced the launch of the Corporation for National and Community Service (CNCS) STEM AmeriCorps initiative, to build student interest in science and engineering, by mobilizing AmeriCorps national service members in nonprofit organizations to work with STEM professionals and inspire young people to excel in STEM education (CNCS, 2013). CNCS subsequently expanded this effort (CNCS, 2014). In 2013, another ambitious national effort was

---

[2] See http://www.100kin10.org/ and http://nsfnoyce.org/, respectively for information about these two initiatives, each of which is designed to improve the quality (and quantity) of STEM teachers.

[3] See http://sites.nationalacademies.org/DBASSE/BOSE/DBASSE_086989, and http://www.nap.edu/catalog.php?record_id=18612 for information on the National Academies efforts, and http://www.gao.gov/assets/670/663079.pdf for the GAO's report]

launched; Citizen Schools' US2020 initiative is a national network focused on preparing our students for STEM-related careers by matching girls, underrepresented minorities, and low-income children with STEM mentors.[4] By providing students with a broader range of exposure to STEM content that is more relevant to their lives, from a more diverse set of adults than their teachers alone (including young adults in STEM-related careers), the hope is to provide greater opportunities to larger numbers of students to learn more about STEM subjects, become more interested in STEM content, and ultimately, to pursue additional STEM education and career possibilities.

As we grapple with the challenge of bolstering student interest in and preparation for further secondary and post-secondary STEM education and careers, research on youth development more generally may offer important and applicable lessons about how to most effectively engage students. The youth development literature has found, for example, that hands-on project learning (often called inquiry learning) and peer-to-peer interactions have positive impacts on metacognitive development, academic outcomes, and student motivation (Flick, 1993; Haury & Rillero, 1994). Hands-on learning often enables students to work together in groups and, in turn, develop social skills. Such opportunities for students to talk through course materials with peers have been found to help students learn in different ways and retain information more effectively (Johnson & Johnson, 1986). Additionally, the research in this area indicates that use of hands-on activities in the classroom positively influences students' attitudes about the content they are learning (e.g., Gerstner & Bogner, 2010; Randler & Hulde, 2007).

The research also suggests that opportunities for adolescents to have meaningful engagements and supportive relationships with adults can influence a range of outcomes, including educational performance, mental health, and problem behavior (DuBois & Silverthorn, 2005; Eby et al., 2008; Rhodes & DuBois, 2008). The presence of positive adult role models, particularly in the form of mentor relationships, has been shown to have benefits for adolescents across academic and socioemotional domains (Coleman, 1988; Jekielek, Moore, Hair, & Scarupa, 2002; Karcher, 2008; Werner & Smith, 1992). A meta-analysis of research about mentoring found that mentoring programs particularly benefit at-risk youth (DuBois, Holloway, Valentine, & Cooper, 2002), and an impact study of the Big Brothers/Big Sisters program found that mentored students showed greater scholastic competence, higher attendance rates and grades than those without mentors (Tierney, Grossman, & Resch, 2000).

The research base about youth development provides useful insights for thinking about how best practices for adolescent engagement could apply to programs designed to spur educational and ultimately career interests in STEM. One common element across many current STEM initiatives is expanding the number and types of adults with whom students interact about STEM careers and learning. More specifically, many programs have sought to incorporate mentoring relationships, whether the mentors serve as role models, coaches, informal or formal educators, or as representatives of individuals who work in diverse STEM content areas, to support engagement with the STEM fields. Given that mentoring relationships can positively influence academic and developmental outcomes for students, the use of apprenticeships with a purposeful emphasis on STEM may also have the potential to provide similar benefits to participating students.

[4] See http://us2020.org/stem-mentoring/.

1.1 Citizen Schools Expanded Learning Time (ELT) Programming

1.1.1 The Citizen School ELT Program Model

The Citizen Schools Expanded Learning Time (ELT) model is defined by two core components: Apprenticeships and Academic Support, both of which are built into a lengthened school day. The Apprenticeships are the cornerstone of the Citizen Schools ELT program model; they are designed to engage student interest in a broad array of subject areas and experiences by leveraging expertise and commitment from local communities. Generally, apprenticeships are complemented by two distinct types of academic support. Structured homework time is generally offered for an hour each program day; it includes one-on-one goal setting and tutoring, and is known as AIM, or "aspire, invest, and make the grade." Academic League includes targeted academic support in either math or English/Language Arts (at each school's discretion); it is offered twice a week for between 30 and 90 minutes.

The chief mechanism by which the above program components are delivered is through a "second shift" of educators hired to support the schools' extended day, as Teaching Fellows who are embedded with the schools. Teaching Fellows generally work full-time, are funded by AmeriCorps, and are expected to serve for two years. The requirements have become more explicit over time: Teaching Fellows are required to have had some prior experience working with children, have earned some college credit, and to be U.S. citizens. In practice, most have a college degree. Teaching Fellows spend their mornings engaged in varied activities (e.g., providing support for the school, preparing for the afternoon programming). ELT programming generally begins around 3:00 PM, and concludes by 6:00 pm. Although partner schools do not necessarily adhere to the same universal schedule network-wide, schools typically schedule activities across both program components each week.

Through its systematic and but flexible model, Citizen Schools provides consistency to students and their staff, and takes the burden of additional teaching away from first shift educators in their partner schools. Citizen Schools' ELT model also establishes clear expectations against which it can be evaluated. It is important to note that the implementation and impact of the overall model have been examined in a prior national evaluation (see Rulf Fountain, Gamse, Velez, Hillard & Cropper, 2016). We turn next to a discussion of the apprenticeships in particular, which represent the focus of this evaluation. The i3 grant allowed Citizen Schools to examine whether the STEM-focused apprenticeships, in particular, served to engage students in STEM learning, and increase their interest in STEM subjects.

Illustrative Weekly Schedule for Students

- Monday: 60 minutes of homework support, followed by 90-minute math Academic League lesson on fractions
- Tuesday: 60 minutes of homework support, followed by 90-minute *Robotics* Apprenticeship taught by team of Citizen Teachers from Google
- Wednesday: 60 minutes of homework support, followed by 30 minutes of refresh/review on fractions, then 60 minutes of team-building exercises
- Thursday: 60 minutes of homework support, followed by 90-minute Mock Trials Apprenticeship taught by Citizen Teachers from local law firm
- Friday: No Citizen Schools programming typically offered (time used for staff professional development and planning)

1.1.2 Citizen School Apprenticeships

The apprenticeship experience is a distinguishing feature of the Citizen Schools ELT program model. Apprenticeships appeal to middle school students because they have the autonomy to choose their own topics of interest at a time when students do not typically have much choice about what they experience in school. Simultaneously, it appeals to teachers and administrators whose schools can then offer a much

more diverse set of learning experiences to their students, and to community-minded volunteers who can contribute their time, enthusiasm for learning, and expertise. Students can connect with adult volunteers who are interested in teaching a skill or content area about which they are passionate and have clear practice knowledge. The adult volunteers, called Citizen Teachers, are quite diverse in their backgrounds, experiences, and content knowledge, and that diversity is reflected in the breadth of topics about which they teach. Citizen Schools offers apprenticeships in such areas as robotics, mock trials, poetry, dance, documentary film-making, understanding math puzzles, investing, and many other subjects. The apprenticeships, which are typically co-taught by Citizen Fellows and Citizen Teachers, consist of approximately10, 90-minute sessions during a semester, that culminate in a showcase called the WOW! during which students "teach back" to friends, family, and community members to illustrate what they learned.

Students take four apprenticeships each year, two in the fall and two in the spring semester. They learn about possible apprenticeship topics at the beginning of each semester via an Apprenticeship Fair, where Citizen Teachers present short pitches, and students rank their top choices. Citizen Schools staff then assign students based on a combination of student preference and apprenticeship availability, and they try to match students to their top apprenticeship choices each semester. Citizen Schools also tries to maintain a student-staff ratio in the apprenticeships such that there are 24 or fewer students in each one (with at least two adults, one Teaching Fellow and at least one Citizen Teacher), a number that can flexibly accommodate hands-on learning with adequate adult supervision and support.

The Citizen Teachers can use curricula developed and refined by the Citizen Schools national program team to teach their respective apprenticeships (the organization maintains a library organized by broad topic area), and in some cases, Citizen Teachers rely upon their own material or materials/syllabi from other sources. Generally, the Citizen Schools organization provides orientation and training to new Citizen Teachers in the weeks before each semester begins. Additionally, all new Citizen Fellows are required to participate in a summer institute for an introduction to Citizen Schools as well as a WOW-specific training. The expectation is that training sessions are required for both new Citizen Teachers and for Citizen Fellows, although unanticipated obstacles in recruiting, hiring, and placing new Fellows and Citizen Teachers can disrupt the typically required participation in training.

Over the past several years, Citizen Schools has endeavored to increase both the number and range of STEM-focused apprenticeships, responding to increased demand and recognizing the importance of engaging middle school students in STEM subjects through a combination of hands-on learning and exposure to enthusiastic ambassadors for professional careers in STEM fields. This study examines whether the STEM-focused apprenticeships, in particular, serve to engage students in STEM learning, and increase their interest in STEM subjects more so than apprenticeships in other topics.

1.1.3 Organization of the Report

The remainder of this report presents study findings on the implementation and outcomes of the evaluation. Chapter 2 describes the study, including its design, the research questions, the study sample, and data collection and analysis. Chapter 3 summarizes the findings about apprenticeship implementation and impact, and Chapter 4 provides a brief discussion of the findings.

2. Study Design and Methods

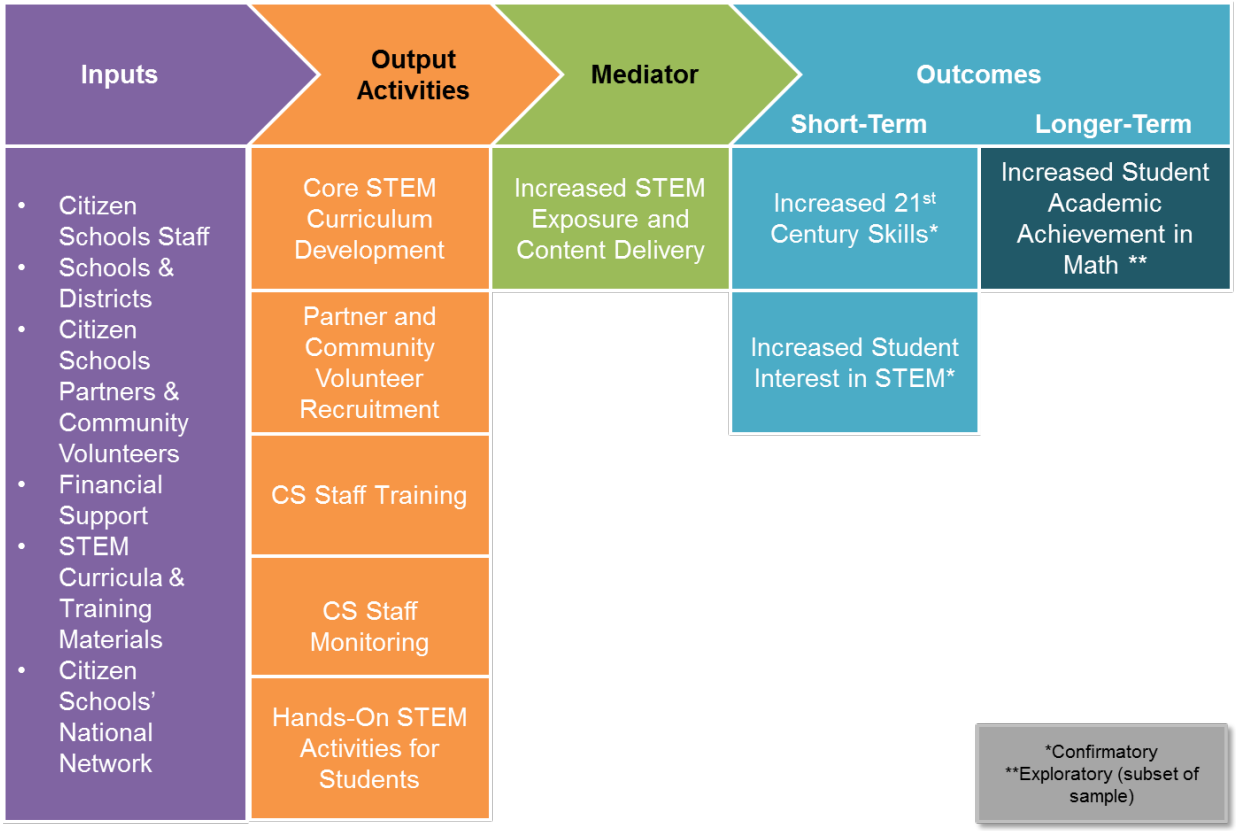
The evaluation of Citizen Schools' Expanded Learning Time (ELT) apprenticeships, funded by the Investing in Innovation (i3) Program, seeks to learn whether and how student participation in apprenticeships leads to improved student engagement and skills. Specifically, the evaluation asks whether participation in science, technology, engineering, and mathematics (STEM)-focused apprenticeships (above and beyond other non-STEM apprenticeships in which students may participate) leads to improved 21st Century and Inquiry Skills (hereafter referred to as 21st Century Skills for brevity) in STEM and STEM engagement for middle school students. Additionally, in accordance with the expectations for all i3-funded evaluations, this study examines the fidelity of implementation, to learn whether and how the apprenticeships have been implemented as intended, or how well participating sites have implemented the STEM apprenticeship model. Program impact is assessed using a quasi-experimental design (QED) to compare students who participate in at least one (and up to four) STEM-focused apprenticeship at Citizen Schools campuses to those who participate in non-STEM-focused apprenticeships. Fidelity of implementation is assessed using records on STEM vs. non-STEM apprenticeship designations, number of students in each apprenticeship, Citizen Teacher and Citizen Fellow attendance at training events, student attendance records, and data routinely collected by Citizen Schools about each session of the apprenticeships. This chapter describes the study's logic model, design, research questions, study sample, data collection measures and administration, and approach to data analysis.

2.1 Logic Model, Study Design, and Research Questions

2.1.1 Logic Model

The Citizen Schools ELT logic model illustrates the inputs, outputs, mediator, and expected outcomes, moving from left to right in Exhibit 2.1. The inputs represent the elements that together comprise the Citizen Schools ELT program model, including the school/district, the Citizen Schools local staff, partners and community volunteers, materials (e.g., curricula), and other supports (e.g., financial support, access to national staff). The outputs (activities/personnel encompassed in the model) include a second shift of ELT staff, establishing partnerships with community members and organizations, training, supporting, and monitoring Citizen Schools staff, and project-based learning experiences for students through apprenticeships. If the model is implemented with fidelity, the theorized short-term outcomes would then include increased student 21st Century Skills and STEM interest. The theorized longer-term outcomes include increased student achievement, as captured by state standardized math test scores. For this study, we focus primarily on examining short-term outcomes, reflecting the duration of the intervention (within a single academic year).

Exhibit 2.1: Citizen Schools Logic Model



2.1.2 Study Design

The Citizen Schools i3 study evaluates both how well the Citizen Schools apprenticeship model, as portrayed in the logic model above, is implemented, and whether participation in the program improves student short and longer-term outcomes. The study purposefully includes an implementation component, to help us understand whether the program model is being carried out with fidelity.

The impact component of the study is designed to assess short and longer-term student outcomes. To estimate the effect of participating in STEM apprenticeships on those outcomes, the study relies on a pre-post QED design to control for as many reasonable alternative hypotheses as possible, the most common of which are: 1) time-varying effects that could influence the treatment and control sample; and 2) characteristics of schools that affect outcomes. Analyses of short-term and longer-term outcomes address the first category of alternative hypotheses using a group of comparison students from the same schools as treatment students. Comparison students help control for changes in a school or region that could affect outcomes. In addition, models use school fixed effects to address the second category of alternative hypotheses mentioned above: characteristics of schools that could potentially affect outcomes. For example, within a given school, students' parental motivation presumably affects pre- as well as post-intervention outcomes. Including school fixed effects in the models allows the estimation of the effect of STEM apprenticeship participation over and above what might be expected given the effects of school characteristics.

2.1.3 Research Questions

The evaluation's research questions, presented below, follow the left-to-right pathway depicted by the logic model.

Outputs: Implementation

The implementation study allows the evaluation team to determine whether the Citizen Schools apprenticeship model is being carried out with fidelity. The implementation research question is:

How well do participating sites implement the Citizen Schools apprenticeship model?

Student Outcomes

Research questions for the short and longer-term outcomes (non-academic and academic) portion of the study were delineated as confirmatory and exploratory prior to analysis. Following Schochet (2008), the study's confirmatory analyses test central research questions and hypotheses of the study; they are driven by the theory of change and are limited in number. Exploratory analyses help to generate hypotheses, but may be limited because of smaller samples sizes.

Short-Term Non-Academic Outcomes

The impact of participation in at least one STEM apprenticeship on STEM interest and is assessed using both **confirmatory** and **exploratory** analyses. Research questions are listed below.

Confirmatory

- What is the effect of participating in at least one STEM apprenticeship on middle school students' 21st Century Skills after one year, compared to participation in no STEM apprenticeships?
- What is the effect of participating in at least one STEM apprenticeship on middle school students' STEM interest after one year, compared to participation in no STEM apprenticeships?

Exploratory

- What is the effect of participating in at least one STEM apprenticeship on middle school students' 21st Century Skills after one year, compared to participation in no STEM apprenticeships, **by gender**?
- What is the effect of participating in at least one STEM apprenticeship on middle school students' 21st Century Skills after one year, compared to participation in no STEM apprenticeships, **by grade**?
- What is the effect of participating in at least one STEM apprenticeship on middle school students' STEM interest after one year, compared to participation in no STEM apprenticeships, **by gender**?
- What is the effect of participating in at least one STEM apprenticeship on middle school students' STEM interest after one year, compared to participation in no STEM apprenticeships, **by grade**?

Longer-Term Academic Outcomes

All analyses conducted with student academic outcomes are considered **exploratory** as these are longer-term outcomes where impacts may not be observed within the year that the study was conducted. Further, prior to the study's start, we were aware that we would not receive achievement data from all schools/districts. As a result, we anticipated that these analyses might have smaller sample sizes, and thus less power, than the short-term outcomes.

Exploratory

- What is the effect of participating in at least one STEM apprenticeship on middle school students' math standardized test scores after one year, compared to students who participate in no STEM apprenticeships?
- What is the effect of participating in at least one STEM apprenticeship on middle school students' math standardized test scores after one year, compared to students who participate in no STEM apprenticeships, **by gender**?
- What is the effect of participating in at least one STEM apprenticeship on middle school students' math standardized test scores after one year, compared to students who participate in no STEM apprenticeships, **by grade**?

2.2 Study Sample

Citizen Schools uses an iterative, multi-step process utilizing mutual "goodness of fit" measures and intensive, ongoing communication with district and school administrative staff to select ELT campuses. All 21 campuses implementing ELT during the 2015-16 school year, serving 2,835 students in five regions, were included in the study. However, as illustrated in Exhibit 2.2, four schools were unable to provide survey data, and five were not able to provide achievement data and were thus removed from analyses.

Exhibit 2.2. Characteristics of Study Sample Schools

	Grade Levels Served	Survey Data	Achievement Data	Number of Apprenticeships Offered	Number of Students in Apprenticeships		Number of Students in School (all grade levels)
Region A					Fall	Spring	
School A	6	X	X	18	84	75	423
School B	6	X	X	18	110	99	368
School C	6	X	X	34	159	138	677
School D	6	X	X	32	187	202	607
School E	6			16	56	55	611
Region B							
School F	6,7,8	X	X	20	68	46	953
School G	6,7,8	X	X	10	53	43	223
School H	6,7,8		X	19	75	66	386
School I	6,7,8	X	X	18	87	66	461
Region C							
School J	6	X	X	32	76	65	511
School K	6,7,8	X	X	49	213	197	844
School L	6	X	X	28	115	93	366
School M	6,7	X	X	72	246	242	551
Region D							
School N	6,7,8			18	58	43	540
School O	6,7,8	X		25	96	84	154
School P	6,7,8	X	X	34	137	93	165
School Q	6,7,8	X	X	31	111	95	151
School R	6,7,8			50	220	158	462
School S	6,7,8	X	X	30	116	98	157
School T	6,7	X	X	28	120	113	236
Region E							
School U	6,7,8	X		32	189	105	891
Total	21	17	16	614	2576	2176	9737

Exhibit Reads: Survey and achievement data were available from 6th grade students participating in the i3 evaluation in School A. The school offered 18 apprenticeships for 84 students in the fall and 75 students in the spring. The total enrollment in the school across all grades in the 2015-16 school year was 423 students.

Notes:
School names have been replaced by letters to present school data anonymously. Schools are organized by region and cells shaded in red indicate that either pre- or post-data were not received from the school.

Source: Citizen Schools communications; Citizen Schools administrative records; and surveys from the 2015-16 school year, and achievement datasets from spring 2015 and spring 2016; Common Core of Data Enrollments from 2015-16.

2.2.1 Student Sample

Non-academic outcomes sample

Non-academic outcomes, including 21st Century Skills and STEM interest, were assessed using pre- and post-student surveys administered in the fall and spring (see more detail on measures in the following

section). Students were eligible for inclusion in the analytic sample if they: 1) had parental consent to participate in the study; 2) had completed a pre- and post-survey; 3) were not exempt from participating in apprenticeships; and 4) had non-missing values for gender and grade. Treatment status was assigned to students who participated in at least one STEM apprenticeship during the 2015-16 school year. Students in the comparison sample had participated in **no** STEM apprenticeships during that school year. Citizen Schools provided the study team with information about each student's specific apprenticeships and each apprenticeship's designation as STEM-focused or not. The Abt team assessed the designations to ensure that those apprenticeships categorized as STEM-focused included sufficient content to merit a STEM designation.

A key aspect of finalizing the non-academic analytic sample was ensuring that treatment and comparison students were similar enough on their baseline 21st Century Skills and STEM interest that they could be reasonably considered fair counterparts of each other. To assess the comparability of the treatment and comparison students, the study team compared the an average baseline standardized effect sizes adjusted for school blocks[5] of 21st Century Skills and the STEM interest composite measures for treatment and comparison students (see Appendix A.1 for model specifications). Any baseline differences larger than .25 standard deviations would be considered meaningful enough that they might affect the analysis (Ho et al., 2007). This approach is consistent with guidelines outlined by the U.S. Department of Education's What Works Clearinghouse (U.S. Department of Education, 2017). Exhibit 2.3 presents the average baseline differences between treatment and comparison students and indicates that the two groups had almost identical baseline values for both outcomes.[6] As such, analyses proceeded with all eligible students in the sample (and no matching strategy was needed to balance the sample). Note that because some students skipped portions of the survey, the analytic sample sizes for the two outcomes are slightly different (1,008 for 21st Century Skills and 1,040 for STEM interest).

Exhibit 2.3. Baseline Equivalence of Treatment and Comparison Students in Non-Academic Sample

	Treatment			Comparison			
Measure	N	Mean	Unadjusted SD	N	Mean	Unadjusted SD	Difference in Effect Size Units
21st Century Skills	838	4.89	1.22	170	4.83	1.16	0.04
STEM Interest	860	3.49	0.56	180	3.49	0.56	-0.01

EXHIBIT READS: Treatment and comparison students scored 4.89 and 4.83, respectively, on the baseline 21st Century Skills scale. The difference between treatment and comparison students adjusted for school blocks included in the analytic model is 0.04 effect size units.

Source: Baseline Citizen Schools student surveys from 2015.

Sample: 1,008 Citizen Schools students eligible for inclusion in the 21st Century Skills sample and 1,040 Citizen Schools students eligible for inclusion in the STEM interest sample.

Exhibit 2.4 presents the characteristics of the non-academic outcomes analytic samples. For ease of presentation, the table includes the characteristics of students with either 21st Century Skills or STEM

[5] The study team took this approach given that the allocation of students to treatment and comparison groups varied widely by school.

[6] The difference in effect size is adjusted by including school dummies that account for unequal assignment to treatment and comparison groups across schools.

interest outcome data, although the samples, as noted above, differ slightly (see Appendix Exhibit A.2 for characteristics of students by outcome sample). The table indicates that over 80 percent (82.8 percent) of the non-academic analytic sample had participated in a STEM apprenticeship during the 2015-16 school year. Although the overall analytic sample is almost perfectly divided between males and females, the latter, who by middle school typically report less interest in STEM (Desy, Peterson, & Brockman, 2011), comprise a smaller percentage of the treatment group (45.7 percent) than the comparison group (62.2 percent). Further, the overall analytic sample is comprised primarily of 6th (63.8 percent) and 7th (27.1 percent) grade students, although the comparison sample included a higher proportion of 8th grade students compared to the treatment sample (23.8 and 6.1 percent, respectively). Analytic models, described in more detail below, will control for these differences in student characteristics.

Exhibit 2.4. Characteristics of Non-Academic Analytic Student Sample

	Treatment		Comparison		All students	
	N	Percent	N	Percent	N	Percent
Gender						
Female	407	45.7	115	62.2	522	48.5
Male	484	54.3	70	37.8	554	51.5
Grade						
6th	580	65.1	106	57.3	686	63.8
7th	257	28.8	35	18.9	292	27.1
8th	54	6.1	44	23.8	98	9.1
All students	891	82.8	185	17.2	1076	

EXHIBIT READS: Females comprised 45.7 percent of the treatment sample, 62.2 percent of the comparison sample, and 48.5 percent of the sample overall.

Source: Citizen Schools student surveys and administrative records.

Sample: 1,076 Citizen Schools students in the 21st Century Skills or STEM interest analytic sample.

Similar to the non-academic student sample, the academic student sample was comprised of students for whom the study had parental consent, who were not exempt from apprenticeships, and had non-missing gender and grade values. In addition, this set of analyses is limited to students with math achievement scores from both 2014-15 and 2015-16 school years. Recall from Exhibit 2.2 that the number of schools for which we obtained achievement data were slightly different. As a result, a separate analytic sample was constructed for the achievement sample, beginning with assessing baseline differences between treatment and comparison students. Exhibit 2.5 indicates that, like the survey sample, students in the two groups were similar on their baseline achievement profiles, with baseline differences below .25 standard deviations. Analyses therefore proceeded using all eligible students.

Exhibit 2.5. Baseline Equivalence of Treatment and Comparison Students in Academic Sample

	Treatment			Comparison			
Measure	N	Mean	Unadjusted SD	N	Mean	Unadjusted SD	Difference in Effect Size Units
2015 Math Scores	927	-0.34	0.94	228	-0.50	0.88	0.09

EXHIBIT READS: Treatment and comparison students scored 0.34 and 0.50 standard deviations below the state average respectively, at baseline. The difference between treatment and comparison students adjusted for school blocks included in the analytic model is 0.09 effect size units.

Source: Spring 2015 Citizen Schools student achievement data.

Sample: 1,155 Citizen Schools students eligible for inclusion in the academic outcomes sample.

Exhibit 2.6 suggests that the academic outcome student sample was very similar to the non-academic outcome student sample. About 80 percent of students (80.3 percent) were in the treatment sample, indicating that they had participated in at least one STEM apprenticeship during the 2015-16 school year. Similar to the non-academic sample, there were higher proportions of females in the comparison than treatment sample (65.4 versus 44.3 percent) and a higher proportion of 8th grade students (36.4 versus 11 percent). Analytic models control for both student characteristics.

Exhibit 2.6. Characteristics of Academic Analytic Student Sample

	Treatment		Comparison		All students	
	N	Percent	N	Percent	N	Percent
Gender						
Female	411	44.3	149	65.4	560	48.5
Male	516	55.7	79	34.7	595	51.5
Grade						
6th	551	59.4	89	39.0	640	55.4
7th	274	29.6	56	24.6	330	28.6
8th	102	11.0	83	36.4	185	16.0
All students	927	80.3	228	19.7	1155	

EXHIBIT READS: Females comprised 44.3 percent of the treatment sample, 65.4 percent of the comparison sample, and 48.5 percent of the sample overall.

Source: Citizen Schools student achievement, survey, and administrative data.

Sample: 1,155 Citizen Schools students in the academic analytic sample.

2.3 Data Collection Measures and Administration

Data collection for the study began in the fall of 2015 and concluded by the spring of 2016. Below, we describe each data collection element (data collection instruments can be found in Appendix Volume II). Note that although several data sources were obtained directly from Citizen Schools, findings were not subject to the approval of the organization, and the study team independently conducted key aspects of the evaluation (e.g., assessing baseline equivalence, impact analysis, and reporting as described in greater detail below).

2.3.1 Implementation Data

Throughout the 2015-16 school year, Citizen Schools collected several different types of data about implementation, including records on the number of STEM curriculum units developed, the number of students attending each apprenticeship, Citizen Teacher and Teaching Fellow attendance rates during trainings, student attendance records, and an electronic weekly checklist about each apprenticeship, called weekly trackers. Additional information about each implementation measure is provided below, and Appendix Exhibit A.3 provides more detailed information about all implementation data measures.

STEM Curriculum Units: Citizen Schools calculated the number of STEM curriculum units developed and provided this information to the study team.

Number of Students in Apprenticeships: Citizen Schools provided the study team with an Excel spreadsheet containing the number of students enrolled in each apprenticeship in the fall and spring in study schools.

New Teaching Fellow Attendance at Summer Institute Citizen Schools provided the study team with an Excel spreadsheet containing the attendance records for each new Teaching Fellow assigned to the study schools at a five-day summer training institute.

Teaching Fellow Attendance at WOW Training: Citizen Schools provided the study team with an Excel spreadsheet containing the attendance records for each Teaching Fellow assigned to the study schools at WOW trainings. These trainings occur once per semester.

New Citizen Teacher Attendance at Training: Citizen Schools provided the study team with an Excel spreadsheet containing the attendance records for each new Citizen Teacher assigned to study schools at a required training.

Student Attendance Rates: Citizen Schools provided the study team with an Excel spreadsheet containing student attendance records for each apprenticeship students participated in throughout the 2015-16 school year.

Weekly Trackers: Teaching Fellows, who co-teach apprenticeships with Citizen Teachers, completed online weekly trackers after each apprenticeship week to monitor the delivery of curriculum. The trackers included information on each week's amount/type of contact between the Citizen Teacher and Teaching Fellow, areas of success and concern, and Teaching Fellows' judgement about the degree to which the Citizen Teacher was: prepared for the lesson, able to execute the lesson effectively, and was in command of lesson content. The weekly tracker also included Teaching Fellows' ratings about the Citizen Teachers' rapport with students, the degree to which students were engaged, student behavior, the degree to which students benefitted from the week's apprenticeship, and whether the Teaching Fellow would like to meet with a supervisor to discuss any issues that arose during the weekly apprenticeship.

2.3.2 Surveys

Citizen Schools students were administered either paper or online surveys in the Fall of 2015 and the Spring of 2016, at each school's preference. Schools with limited or no access to computer labs were mailed paper surveys directly from the study team and then were asked to return the completed surveys directly to Abt. Schools with access to computers administered surveys online using Citizen Schools' existing survey portal. De-identified, raw data files were then transferred to the study team to be combined with the paper survey data.

Surveys asked three sets of questions: 1) 21st Century Skills questions; 2) STEM interest questions; and 3) Citizen Schools specific questions, which were identical across the two survey waves. Twenty-first Century Skills were recorded using the 21st Century and Inquiry Skill Assessment (LaBanca, Oh, Lorentson, Jia, & Sibuma, 2014). This Likert-scale instrument is comprised of 12 items measured on a scale from 1 = Extremely Poor to 7 = Excellent. The scale asks students to indicate how well they, for example, identify "real world challenges or problems in STEM." To assess STEM interest, we used a Likert-scale instrument developed by Weinburgh and Steele (2000) designed specifically for middle school students called the Modified Attitudes Towards Science Inventory (mATSI). The 25-item measure asks students to indicate their agreement/disagreement with statements such as "I have a real desire to learn science." Responses were recorded on a scale from 1 = Strongly Disagree to 5 = Strongly Agree. Composite scores for each scale were calculated by averaging non-missing values across the scales' items for those students who had answered at least four-fifths of questions. This resulted in one average 21st Century Skills metric (baseline $\alpha = .92$; follow-up $\alpha = .95$) and one average STEM interest metric (baseline $\alpha = .87$; follow-up $\alpha = .88$) at baseline and follow-up.

Finally, the surveys included five Citizen Schools-specific questions measured on Likert scales. These questions asked students about their general interest in a STEM career, their overall experience with Citizen Schools, and plans to graduate from high school, attend college, and graduate from college. These items were of particular interest to Citizen Schools as they are items that Citizen Schools has measured for several years and use to measure student impact annually and to inform programming; they are not included in the present evaluation.

2.3.3 Achievement Data

The study team obtained students' math standardized test scores from the Spring of 2015 and Spring of 2016 from a subset of districts and schools. Test scores were sent first to Citizen Schools as per their Memoranda of Understanding (MOUs) with each school/district, and then transferred to Abt after de-identification. To enable the merging of achievement scores across states that use different standardized tests, prior to combining data, student test scores were converted into a standardized z-score metric.[7]

2.4 Implementation Data Analysis

As outlined in greater detail in Appendix Exhibit A.3, each element of the implementation data was coded to indicate whether or not a baseline level of fidelity was met. For example, to measure whether a sufficient number of Citizen Teachers had been recruited to provide the student-teacher ratio intended by Citizen Schools (i.e., fewer than 24 students), the number of students in each apprenticeship was recorded. Those apprenticeships with fewer than 24 students were categorized as meeting the teacher-student ratio criterion. These apprenticeship-level metrics were then averaged across the full set of apprenticeships implemented throughout the year to determine whether at least 95 percent of apprenticeships had met the fidelity threshold. A similar process was used across all implementation data elements and then each element was summed to produce one composite implementation measure summarizing whether the intervention as a whole met fidelity benchmarks. Additional information and results from the implementation data are presented in Chapter 3.

2.5 Non-Academic and Academic Outcomes Data Analysis

All outcomes, whether non-academic or academic, were modelled using an identical strategy (see Appendix A.4 for formal model specifications). All models included school fixed effects to account for characteristics of schools, along with the students' baseline measure, gender, and grade information. The impact of participating in a STEM apprenticeship on each outcome was modeled using an indicator for whether or not a student was in the treatment group (i.e., participated in at least one STEM apprenticeship during the 2015-16 school year). Note that because the confirmatory outcomes are in different domains (21st Century Skills and STEM interest), the study team did not need to adjust for multiple comparisons.

[7] For each test subject, the z-score is calculated by subtracting the average scaled score for all students in that grade and year in the state from each student's score, and dividing by the standard deviation in that grade and year. For example, if a student's math score is identical to the statewide average for that year, the math z-score for that student would be zero. Positive and negative scores would indicate that the student scored above and below the state average, respectively. Note that for schools in one school district included in the achievement analytic sample, the study team received scores on a districtwide test rather than a statewide test. For that state, the study team used information available about other districts in the state that had also used that particular test to calculate the z-score.

Similarly, no adjustments were made to the exploratory analyses given that these are preliminary. Additional information and results from the outcomes analyses are included in Chapter 3.

2.5.1 Interpreting the Results

The goal of these analyses is to estimate the impacts among students who participated in STEM apprenticeships and to test whether their outcomes differ from what would be expected had they not participated in these apprenticeships, often referred to as the counterfactual. Given that the counterfactual cannot be observed, the study uses a statistical technique to estimate the counterfactual using data from students who did not participate in STEM apprenticeships. If differences are observed, there is evidence that those differences can be attributed to the effects participating in at least one STEM apprenticeship.

2.6 Study Limitations

2.6.1 Study Sample

As highlighted above, although every school in Citizen Schools' network was eligible to participate in the study, four of the 21 schools provided no survey data and five schools provided no achievement data. As a result, findings are only generalizable to those schools that provided data. Further, because the schools providing survey and achievement data did not directly overlap, findings from the two samples are not directly comparable. Recall also that the samples were limited to students for whom we had parental consent, were not exempt from apprenticeships, and had non-missing baseline, demographic, and outcome variables, and are therefore results are only generalizable to this subset of participants. To the extent that students with this information were qualitatively different than those without this information, findings may not be applicable to all students in the Citizen Schools network.

2.6.2 Implementation

There are two primary limitations of the implementation data. First, there was limited information about Citizen Schools' staff trainings in terms of content and quality. Further, some of the trainings were mandatory while others were voluntary and the study team was only able to collect information about attendance at mandatory trainings, making it difficult to know how much total training staff received. Together, these limitations make it challenging to determine how heavily the trainings may have affected implementation. Second, determining which apprenticeships were to be counted as STEM vs. non-STEM was somewhat ambiguous. The study team based its apprenticeship determinations on brief descriptions provided by Citizen Schools; however, some descriptions were vague, which meant that the categorization of STEM vs non-STEM may be somewhat imperfect. Also, in theory, students were assigned to their first, second, or third choice apprenticeships, but depending on the popularity of apprenticeships and the mix of interested students, this was not always feasible, potentially affecting the number of STEM apprenticeships available to students. Similarly, some districts were in highly STEM-friendly areas of the country (e.g., Silicon Valley) while others were in regions with less access to STEM expertise. Specifically, in some regions Citizen Schools had a higher proportion of corporate and community partners focused on STEM than other regions. The STEM apprenticeship options (both number and desirability) varied between districts and even between schools.

2.6.3 Outcomes

The outcomes analyses face two primary limitations. First, the breadth and diversity of the roughly 400 apprenticeships offered by Citizen Schools (with topics ranging from robotics to video game design to students' financial futures and yoga), means that there is not one single measure that could apply equally

well to all of them. Instead, more general measures of STEM interest and skills were selected, with the intention of capturing the attitudes and engagement that the STEM-focused apprenticeships were hypothesized to change. However, by choosing more general rather than more customized measures, it is possible that the outcome metrics were not sensitive enough to detect more fine-grained outcomes that may well be associated with particular apprenticeship experiences. Second, note that although analyses modeling the relationship between participation in at least one STEM apprenticeship and student outcomes include numerous controls to increase their rigor, because the study relies on a quasi-experimental design, it is possible models may omit unobservable characteristics that affect outcomes.

3. Findings

How schools implement the Citizen Schools ELT apprenticeships is critical to the effectiveness of the program model. This chapter summarizes findings about ELT apprenticeship implementation, drawing from multiple data sources that measure key aspects of the program model, including curriculum development, volunteer recruitment, Citizen Schools staff and volunteer training and monitoring, and hands-on STEM activities for students. The chapter also examines the apprenticeships themselves, by exploring the weekly challenges and successes experienced by those who delivered program content. The discussion on implementation is followed by an assessment of program impacts using a rigorous, QED design to compare students who participated in at least one STEM apprenticeship to those who participated in no STEM apprenticeships. Impacts on student short-term outcomes are discussed, including 21st Century Skills and STEM engagement, and longer-term outcomes, math achievement.

3.1 Implementation Fidelity Results

As described in Chapter 2, implementation measures assess whether the ELT apprenticeship model was carried out with fidelity. Overall program fidelity was measured across the full sample of participating Citizen Schools ELT campuses using five key components, described in further detail below. The study team also analyzed the content of the *Weekly Trackers*, one component of the overall fidelity measure, to provide additional information about student engagement and behavior, Citizen Teacher preparedness, and areas of success and concern. The implementation section of the chapter first describes the measurement of overall program fidelity and the results, and then describes information derived from the *Weekly Tracker*.[8]

3.1.1 Overall Program Fidelity

The overall program fidelity measure was created to reflect key activities that theoretically underpin successful implementation of the Citizen Schools ELT model. Five key components were identified, including the creation of STEM curriculum, recruitment of sufficient partner and community volunteers, successful training of Teaching Fellows and Citizen Teachers, curriculum delivery, and provision of hands-on activities for students (including those with a STEM focus). Given the variability in training offered to (and taken up by) Teaching Fellows and Citizen Teachers, the training component was measured using indicators for attendance at trainings for new Teaching Fellows, WOW trainings for all Teaching Fellows, and trainings for new Citizen Teachers, respectively. Meeting overall program fidelity expectations would mean that each component (and indicator, where applicable) meets the established fidelity thresholds. Exhibit 3.1 presents information on each indicator and component used in the overall program fidelity measure, specifying whether or not the threshold was met (see also Appendix A.3 for additional information about data sources and an outline of the components and their final threshold scores).

[8] Results for this section rely on a slightly smaller sample size than the sample used for the fidelity metric. The fidelity metric counted all trackers submitted to Citizen Schools across all weeks. In just over 1 percent of trackers, two submissions were received for a given date, but appear to have been submitted by different people as the records were not exact duplicates. For the fidelity results, one of the two records was randomly selected to be removed so that the resulting descriptives do not double count information for a given week.

Exhibit 3.1 Overall Program Fidelity Measurement and Results

Key Component	Metric	Fidelity Threshold	Fidelity Threshold Met?
1: Core STEM Curriculum Development	Citizen Schools' records from the end of the 2015-16 school year on STEM vs. non-STEM apprenticeship designations assess whether STEM curriculum units were developed	Development of 21+ STEM curriculum units.	Yes
2: Partner and Community Volunteer Recruitment	The number of students attending each apprenticeship serves as a proxy for whether the organization recruited a sufficient number of Citizen Teachers	95% or more of the apprenticeships had fewer than 24 students	No
3: Citizen School Staff Training			No
Indicator 1	Teaching Fellow attendance at a 5-day summer institute training for new Teaching Fellows (i.e., Teaching Fellow 1s)	86% or more of Teaching Fellow 1s attended 4+ days of training	*No*
Indicator 2	Teaching Fellow attendance at a WOW! training offered each semester of teaching	86% or more of Teaching Fellows attended WOW! trainings each semester they taught	*Yes*
Indicator 3	Citizen Teacher attendance at trainings for new Citizen Teachers (i.e., Citizen Teacher 1s)	86% or more of Citizen Teacher 1s attended training	*Yes*
4: Citizen School Staff Monitoring	Completion of *Weekly Apprenticeship Trackers* by Citizen Schools staff after each week of the apprenticeship to monitor curriculum delivery	86% or more of apprenticeships submitted weekly trackers 9+ times per apprenticeship[1]	Yes
5: Hands-On Activities for Students (including those with a STEM focus)	Apprenticeship attendance records reflect whether students participated in hands-on STEM activities	86% or more of students attended 9+ weeks of apprenticeships	No
		Overall	No

EXHIBIT READS: The threshold was met for the core STEM curriculum development component of the fidelity measure, indicating that over 21 STEM curriculum units were developed. This component was measured using Citizen School records on STEM apprenticeship designations.
Notes:
[1]The weekly tracker data used to calculate fidelity for this component was missing for one school in the spring that exited the network at the end of the school year.
Source: Citizen Schools administrative records from the 2015-16 school year.

As demonstrated in Exhibit 3.1, fidelity thresholds were not met for the three components: partner and community volunteer recruitment, Citizen Schools staff training, and sufficient student participation in hands-on activities. Fidelity thresholds were met for curriculum development, two of the three indicators for Citizen Schools staff training, and for Citizen Schools staff monitoring. Based on an a priori determination that thresholds would have to be met or exceeded across *all* of the indicators listed above, the overall fidelity threshold was not met.

Data on the number of students attending each apprenticeship, a proxy for whether Citizen Schools had recruited a sufficient number of Citizen Teachers, indicate that a large majority (87 percent) of apprenticeships served fewer than 24 students. Yet the average number of students per apprenticeship varied widely by school, ranging from roughly 12 to 26 (see Appendix B.1). Only three schools, all in one region, averaged more than 24 students per apprenticeship, and the school-level averages in those three

cases were just over the threshold (either 25 or 26 students per apprenticeship). As a result, although the specified threshold was not met for this component, the results do not suggest that apprenticeships in study schools were substantially over-enrolled, nor that there is a pervasive issue with partner and community volunteer recruitment.

Likewise, although the Citizen Schools staff training component did not meet its threshold, defined by all three indicators meeting fidelity, note that only one of the three indicators did not meet fidelity standards—the measure of first-year Teaching Fellow attendance at the summer institute training. The attendance data for this group of new Teaching Fellows indicate that ***each*** individual who did not attend at least four days of the summer institute was deemed a "late hire," who had joined the program on or after July 29th when the summer institute had already been completed. Indeed, three-quarters of the schools in the sample that hired new Teaching Fellows had at least one late hire, and in one-quarter of schools at least of half of the new Teaching Fellows were hired late (see Appendix B.1). As a result, across the sample, 71 percent of first year Teaching Fellows attended the summer institute for at least four days. Note, however, that Teaching Fellow WOW! attendance and first-year Citizen Teacher training attendance were high and met fidelity thresholds.

The fidelity measure also captured whether students participated in hands-on activities, as measured by apprenticeship attendance records. Results indicate that attendance rates across apprenticeships were lower than expected; 72 percent of students attended at least nine weeks of all apprenticeships they participated in. The school-by-school attendance rates (for nine or more sessions) ranged from 59 to 86 percent of students, and these rates also varied by region (see Appendix B.1). School-by-school semester attendance rates for students attending at least nine apprenticeship sessions indicate rates that ranged from 67 to 92 percent. However, students didn't always maintain high attendance *across* semesters, which was the required metric for the fidelity measure.

Together, closer examination of the fidelity component data suggests that the program was close to meeting targets for partner and community volunteer recruitment, but faced challenges in timely hiring of Teaching Fellows as well as in ensuring stable student attendance in apprenticeships throughout the school year. Note, also, that while the threshold was high for both partner and community volunteer recruitment (86 percent), the 95 percent threshold (for maintaining enrollment under 24 students per apprenticeship) was even more stringent. Also important to recognize is that the fidelity measure revealed implementation successes in the development of core STEM curriculum, Teaching Fellow attendance at WOW! trainings, new Citizen Teacher attendance at trainings, and completion of the *Weekly Tracker*. The next section on fidelity uses the latter data source to provide additional information about apprenticeship implementation.

3.1.2 Weekly Apprenticeship Tracker Findings

Citizen Schools collects weekly trackers from Teaching Fellows who co-teach apprenticeships with Citizen Teachers to monitor the delivery of curriculum. Teaching Fellows are expected to complete a tracker each week for any apprenticeship they teach during the typical 10-week session.[9] The apprenticeship tracker collects information on Teaching Fellows' perceptions, each week, on the following:

[9] Apprenticeships are expected to be held for 10 weeks. On average, 9 weekly trackers were submitted per apprenticeship, although the number submitted ranged from 1 to 11.

- amount/type of contact between the Citizen Teacher and Teaching Fellow during the week;
- areas of success and concern;
- the degree to which the Citizen Teacher was prepared for the lesson, able to execute the lesson effectively, and was in command of the lesson content; and
- information about Citizen Teacher rapport with students, the degree to which students were engaged, student behavior, the degree to which students appeared to benefit from the week's apprenticeship, and whether the Teaching Fellow would like to meet with a supervisor to discuss any issues that arose during the weekly apprenticeship.

Below, we present results, by theme, averaged across all weekly tracker records and apprenticeships.[10] Appendix B.2 presents additional detail on all results below.

Areas of Success

Teaching Fellows were asked to identify areas of success, with a particular focus on student engagement and Citizen Teacher preparation and practices for each week of a given apprenticeship (see Exhibit 3.2). Overall, the Teaching Fellows perceived that apprenticeships covered an appropriate amount of content and maintained high student attendance in most weeks (73 and 70 percent, respectively), while high student engagement/enthusiasm and coverage of the 21st Century Skill learning objectives were perceived to be areas of success in somewhat fewer apprenticeship weeks (65 and 59 percent, respectively).

Respondents generally perceived that apprenticeships covered the appropriate amount of content.

Exhibit 3.2 Areas of Success

Identify the areas of success this week (check all that apply)	N	Percent
Apprenticeship session covered an appropriate amount of content	5109	73%
High student attendance	5109	70%
High student engagement/enthusiasm	5109	65%
Apprenticeship session achieved all or most of the students' 21st century skill learning objectives	5109	59%

EXHIBIT READS: Across all apprenticeships, the apprenticeship session covered an appropriate amount of content in 73 percent of all weeks.
Notes:
Weekly Tracker responses were aggregated across all apprenticeships and weeks (N=5109, Missing=0)
Responses do not sum to 100% because multiple responses were permitted.
Source: Apprenticeship Weekly Tracker Q5 (Identify areas of success this week (check all that apply))

Areas of Concern

Teaching Fellows were also asked to identify areas of concern, with a particular focus on student engagement and Citizen Teacher preparation and practices for each week of a given apprenticeship (see Exhibit 3.3). Overall, Teaching Fellows reported few areas of concern in their apprenticeships. Only three topics (student behavior, low student engagement/enthusiasm, and low attendance) were characterized as concerns, for less than one-fifth of all

Student behavior, low student engagement/enthusiasm, and low attendance were characterized as concerns in less than one-fifth of all apprenticeship weeks

[10] The study team also examined results by first averaging across all weeks within an apprenticeship and then across apprenticeships. However, results were not substantively different than those presented.

apprenticeship weeks (affecting 19, 12, and 11 percent, respectively). Other topics were perceived as concerns for less than 7 percent of apprenticeship weeks.

Exhibit 3.3 Areas of Concern

Identify the area(s) of concern with the Citizen Teacher (CT) relationship or the quality of the apprenticeship this week (check all that apply)	N	Percent
Student behavior issues negatively affected apprenticeship	5109	19%
Low student engagement/enthusiasm	5109	12%
Low student attendance	5109	11%
Apprenticeship session achieved only some or none of the 21st century skill learning objectives	5109	7%
CT(s) was/were unprepared	5109	4%
Apprenticeship session covered very little or none of the intended curriculum content	5109	3%
Concern for student safety	5109	2%

EXHIBIT READS: Across all apprenticeships, student behavior issues negatively affected apprenticeships in 19 percent of all weeks.
Notes:
Weekly Tracker responses were aggregated across all apprenticeships and weeks (N=5109, Missing=0)
Responses do not sum to 100% because multiple responses were permitted.
Source: Apprenticeship Weekly Tracker Q6 (Identify areas of concern with the CT relationship or the quality of the apprenticeship this week (check all that apply))

Citizen Teacher Performance

Questions focusing on Citizen Teachers' performance during the apprenticeship used a Likert scale ranging from one to four, where higher numbers indicated a more positive response (Exhibit 3.4). Respondents reported that Citizen Teachers were prepared and delivered well-executed lessons, exhibited a high level of understanding and demonstrated a good rapport with the students (i.e., means for these items were all above the scales' midpoints of 2.5).

Respondents perceived that Citizen Teachers were prepared and delivered well-executed lessons and demonstrated command of the content.

Exhibit 3.4 Citizen Teacher Performance

	N	Mean
Degree to which…		
Citizen Teacher (s) (CTs) was/were prepared for week's lesson	5107	3.4
Citizen Teacher (s) delivered a well-executed lesson	5107	3.2
Ratings of…		
Citizen Teacher (s)'s command of content in given week's lesson	5107	3.5
Citizen Teacher (s)'s rapport with the students	5107	3.2

EXHIBIT READS: Across all weeks of all apprenticeships, the average degree to which respondents indicated Citizen Teachers were prepared for a week's lesson was 3.4
Notes:
Weekly Tracker responses were aggregated across all apprenticeships and weeks (N=5107, Missing=2)
Responses were on a Likert Scale. Q7 response options ranged from: 1 = not at all prepared to 4 = extremely well prepared; Q8 response options ranged from: 1 = not at all well executed to 4 = extremely well executed; Q9 response options ranged from: 1 = does not at all understand the content to 4=the highest level of understanding; Q10 response options ranged from: 1 = poor rapport to 4 = extremely good rapport.
Source: Apprenticeship Weekly Tracker Q7 (To what degree was/were CT(s) prepared for this week's lesson (e.g., planned for students' level of understanding, planned to include hands-on apprenticeship centered activities?)); Q8 (To what degree did the CT(s) deliver a well-executed lesson (e.g., had materials ready, appropriately gauged students attention and interest, covered an appropriate amount of content given students' level of understanding)?); Q9 (How would you rate the CT(s)'s command of the content in this week's lesson (e.g., answering student questions, providing varied and real life examples to illustrate concepts)?); Q10 (How would you rate the CT(s)'s rapport with the students (e.g., has connection with all or most students in apprenticeship, knows everyone's name)?)

Student Engagement and Behavior

Teaching Fellows reported on student engagement and behavior during apprenticeships, using a similar Likert scale that ranged from one to four where higher numbers indicated a more positive response. As Exhibit 3.5 demonstrates, on average, respondents reported that students were engaged across all weeks. Note also that although some Teaching Fellows highlighted student behavior as an area of concern, it was generally not perceived to have a negative impact on apprenticeship learning.

Exhibit 3.5 Student Engagement and Behaviors

	N	Mean
Degree to which students were engaged	5101	3.1
Degree to which student behavior negatively impacted apprenticeship learning	4909	1.7

EXHIBIT READS: Across all weeks of all apprenticeships, the average degree to which students were engaged was 3.1.
Notes:
Weekly Tracker responses were aggregated across all apprenticeships and weeks; Q11 (N=5101, Missing=6); Q12 (N=4909; Missing=198)
Responses were on a Likert Scale. Q11 response options ranged from: 1=not at all engaged to 4=extremely engaged; Q12 response options ranged from: 1=no negative impact to 4=extremely negative impact.
Source: Apprenticeship Weekly Tracker Q11 (How would you rate student engagement (e.g., participating in lesson, doing what was asked of them)?); Q12 (Please rate the extent to which student behavior negatively impacted apprenticeship learning)?)

Student and Apprenticeship Outcomes

Questions about the perceived student and apprenticeship outcomes used a five-point scale, with higher numbers indicating more positive responses. As Exhibit 3.6 demonstrates, on average, respondents reported that the teacher teams were effective in meeting objectives for a given week and delivering a high-quality apprenticeship, and that students benefited from a week's lesson (i.e., both means were above the scales' midpoints of 3).

Exhibit 3.6 Student and Apprenticeship Outcomes

	N	Mean
Degree to which teacher team (Citizen Teacher(s) and Teacher Lead (TL) was effective in meeting objectives for given week and delivered high-quality apprenticeship	4952	3.9
Degree to which students benefitted from week's lesson	4906	3.8

EXHIBIT READS: Across all weeks of all apprenticeships, the average degree to which the teacher team was effective in meeting objectives for a given week and delivered high-quality apprenticeship was 3.9.
Notes:
Weekly Tracker responses were aggregated across all apprenticeships and weeks; Q13 (N=4906, Missing=203); Q14 (N=4952; Missing=157)
Responses were on a Likert Scale. Q13 response options ranged from: 1 = not at all beneficial to 5 = extremely beneficial; Q14 response options ranged from: 1 = poor to 5 = excellent.
Source: Apprenticeship Weekly Tracker Q13 (To what degree do you think students benefitted from this week's lesson (e.g., were able to build on or apply knowledge/understanding from earlier apprenticeship sessions)?); Q14 (Overall, how effective was the teaching team (CT(s) and TL) in meeting your objectives for this week and delivering a high-quality apprenticeship?)

3.2 Impact of Participation in STEM Apprenticeships on Student Outcomes

This section of the chapter examines the relationship between participation in at least one STEM apprenticeship during the 2015-16 school year and student non-academic and academic outcomes. As described in Chapter 1, prior research suggests that opportunities for adolescents to have meaningful relationship with adults can influence a range of outcomes, including attitudes and achievement. Participation in apprenticeships, where students are exposed to positive role models, can be one way to facilitate mentoring relationships and thereby improve student outcomes. This section first examines the confirmatory, short-term outcomes for the overall sample (depicted in the logic model in Chapter 2),

including 21st Century Skills and STEM interest. The chapter then examines exploratory outcomes for short-term measures before progressing through the logic model to explore the hypothesized distal or longer-term outcomes—the impact of participation in STEM apprenticeships on academic achievement.

The impact models presented in this section leverage (1) data from comparison students to control for variation that affect outcomes similarly across students, and (2) school fixed effects to control for stable characteristics of schools that affect outcome measures. The statistical models estimate the impact of participation in at least one STEM apprenticeship on non-academic and academic outcomes by comparing the observed outcomes for treatment students and the counterfactual outcomes (e.g., regression-adjusted estimates of what would have occurred in the absence of participation in STEM apprenticeships). These models control for alternative hypotheses that might explain observed differences between treatment and comparison students, and, therefore represent strong, quasi-experimental analyses.

Key Findings

Key findings related to student outcomes are summarized below:

- Treatment and comparison students scored above the mid-point of the 21st Century Skills and STEM interest scales at baseline and follow-up.
- There were no significant differences between treatment and comparison students' follow-up 21st Century Skills or STEM interest scores.
- Treatment and comparison students scored below the state/district average on math standardized test scores at baseline and follow-up, likely reflecting the types of schools that the Citizen Schools network targets—schools that tend to have below-average levels of academic achievement.
- Overall, there were no significant differences between treatment and comparison students' math achievement.
- Participation in at least one STEM apprenticeship had a marginally significant impact on male students, an effect roughly equivalent to three months of math growth in a school year.
- Participation in STEM apprenticeships had a significant impact on 6th grade students, relative to 8th grade students, an effect roughly equivalent to four months of math growth in a school year.

3.2.1 Student Non-Academic (Short-Term) Outcomes

Confirmatory Analyses

As described in Chapters 1 and 2, participation in STEM apprenticeships during the school year is hypothesized to affect students' 21st Century Skills and STEM interest. To explore these relationships, the study draws upon surveys administered to Citizen Schools students in the Fall of 2015 and Spring of 2016. Exhibit 3.7 presents unadjusted baseline and follow-up averages by treatment condition for the confirmatory outcomes; it indicates that averages were almost identical between treatment and comparison groups for both outcomes, and above the mid-point for each scale (4 for 21st Century Skills and 3 for STEM interest). Similar to baseline results, follow-up averages are almost identical between treatment and comparison students, and differ only slightly from the baseline scores. Citizen Schools students may have had above average 21st Century Skills and STEM interest at baseline, and thus have less room to change these over time. Note, also, that 7th and 8th grade students who participated in the evaluation may have participated in Citizen Schools and experienced STEM apprenticeships in previous school years.

Exhibit 3.7 also presents results from models examining the impact of participation in at least one STEM apprenticeship on 21st Century Skills and STEM Interest. These impacts were assessed using analytic models that controlled for pre-test, gender, grade, and school. Recall that the confirmatory outcomes are in different domains (21st Century Skills and STEM interest); as a result, it was not necessary to adjust for multiple comparisons within domain.[11] Results from the confirmatory models are in line with the descriptive findings: there were no significant impacts of participation in at least one STEM-focused apprenticeship on either non-academic outcome. Appendix C.1 presents additional information on the confirmatory impact models.

Exhibit 3.7 Mean Unadjusted Scores and Impact Results for Non-Academic Outcomes

	Treatment					Comparison					Impact Estimates		
Measure	N	Baseline Mean	Unadjusted Baseline SD	Follow-up Mean	Unadjusted Follow-up SD	N	Baseline Mean	Unadjusted Baseline SD	Follow-up Mean	Unadjusted Follow-up SD	Impact Estimate	Impact Standard Error	p-value
21st Century Skills	838	4.9	1.2	4.9	1.3	170	4.8	1.2	5.0	1.2	-0.04	0.11	0.69
STEM Interest	860	3.5	0.6	3.5	0.6	180	3.5	0.6	3.5	0.5	0.01	0.05	0.89

EXHIBIT READS: Treatment students scored 4.9 on the 21st Century Skills scale at baseline and follow-up. These means are unadjusted. Students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.04 points lower on the 21st Century Skills scale than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant. Estimated impacts represent the adjusted differences between treatment and comparison students after controlling for pre-test, gender, grade, and school or region fixed effects.

Notes:

Responses on the 21st Century Skills Scale ranged from 1 = Extremely Poor to 7 = Excellent. Responses on the STEM Interest scale ranged from 1 = Strongly Disagree to 5 = Strongly Agree.

Source: Citizen Schools Student Surveys and administrative records from the 2015-16 school year.

Exploratory Analyses

Recall that analyses of the short-term outcomes by subgroup were included as part of the study's exploratory outcomes. Exploratory analyses of non-academic outcomes included examining impacts by gender and grade to learn whether, for example, males in the treatment condition experienced different effects of participation in STEM apprenticeship(s). These outcomes are limited by less statistical power, as the sample is diminished by exploring outcomes by subgroups. Similar to the approach taken for the confirmatory analyses, no were adjustments made for multiple comparisons, as those are preliminary. As presented in Exhibit 3.8, results indicate no differential impacts by gender or grade. Appendix C.1 includes additional information about the exploratory analyses for the non-academic outcomes.

[11] The confirmatory analyses represent the most rigorous tests of the study's primary research questions and hypotheses; exploratory analyses examine variation in outcomes across subgroups of interest, and also serve to generate hypotheses to be tested more rigorously in the future.

Exhibit 3.8 Impact of Participation in STEM Apprenticeships on Non-Academic Outcomes

	21st Century Skills		STEM Interest	
	N	Estimated Impact	N	Estimated Impact
Gender				
Female	486	0.03	504	-0.05
Male	522	-0.15	536	0.08
Grade				
6th	650	0.04	667	0.03
7th	265	-0.29	276	-0.06
8th	93	-0.07	97	0.02

EXHIBIT READS: Students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.03 points higher on the 21st Century Skills scale than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant.

Notes:

† marginally significant at p < .10 level

* statistically significant at p < .05 level

** statistically significant at p < .01 level

*** statistically significant at p < .001 level

Estimated impacts represent the adjusted differences between treatment and comparison students after controlling for pre-test, gender, grade, and school or region fixed effects.

Responses on the 21st Century Skills Scale ranged from 1 = Extremely Poor to 7 = Excellent. Responses on the STEM Interest scale ranged from 1 = Strongly Disagree to 5 = Strongly Agree.

Source: Citizen Schools Student Surveys and administrative records from the 2015-16 school year.

3.2.2 Student Longer-Term Outcomes (Exploratory)

The chapter now turns to an examination of the impact of participation in at least one STEM apprenticeship on more distal outcomes—student achievement, as measured by standardized mathematics tests administered by districts and states. Recall that all analyses conducted with student academic outcomes are considered exploratory, reflecting the hypothesized longer-term nature of outcomes for which impacts may not be observed within the year that the study was conducted.

Data for the impact analysis include student achievement data from four different states, each of which has its own math achievement test and associated scores. For ease of interpretation and analysis, the impact estimates are presented as z-scores which have been scaled to student-level state means and standard deviations for a given grade, subject, and year, and can thus be interpreted as the score for a student relative to the statewide average.[12] Using z-scores as the analysis metric also allows for the estimated impact of participation in at least one STEM apprenticeship to be interpreted as an effect size, or a proportion of the student-level standard deviation.

[12] For each test subject, the z-score is calculated by subtracting the average scaled score for all students in that grade and year in the state from each student's score, and then dividing by the standard deviation in that grade and year. For example, if a student's math score is identical to the statewide average for that year, the math z-score for that student would be zero. Positive and negative scores would indicate that the student scored above and below the state average, respectively. Note that for schools in one school district included in the achievement analytic sample, the study team received scores on a districtwide test rather than a statewide test. For that state, the study team used information available about other districts in the state that had also used that particular test to calculate the z-score.

Unadjusted Baseline and Follow-up Averages

Exhibit 3.9 presents summary statistics on math achievement z-scores for students in the treatment and comparison groups in the spring of 2015 (baseline) and spring of 2016 (follow-up). Students in both groups scored below the statewide average: the treatment group mean is approximately one-third of a standard deviation below the statewide average at baseline, and the comparison group mean is half a standard deviation below the statewide average at baseline. Average math achievement scores at follow-up were higher than at baseline in both groups, likely reflecting expected growth between years, although still below statewide averages. These scores reflect the types of schools that the Citizen Schools network targets—schools that tend to have below-average levels of academic achievement.

Exhibit 3.9 Mean Unadjusted Baseline and Follow-up Scores for Academic Outcomes, by Treatment Status

	Treatment					Comparison				
Measure	N	Baseline Mean	Unadjusted Baseline SD	Follow-up Mean	Unadjusted Follow-up SD	N	Baseline Mean	Unadjusted Baseline SD	Follow-up Mean	Unadjusted Follow-up SD
Math Achievement	927	-0.34	1.2	-0.29	1.3	228	-0.50	1.22	-0.42	1.2

EXHIBIT READS: Treatment students scored 0.34 standard deviations below the state average at baseline and 0.29 standard deviations below the state average at follow-up. These means are unadjusted.

Notes:

Math achievement scores are presented as z-scores scaled to student-level state means and standard deviations for a given grade, subject, and year.

Source: Student-level achievement data from spring 2015 and spring 2016 from schools in the Citizen Schools ELT Network.

Adjusted Baseline and Follow-up Averages

The impact of participation in at least one STEM apprenticeship on math achievement was assessed using an analytic model that controlled for pre-test (math achievement in spring of 2015), gender, grade, and school. These controls allow us to measure the impact of the STEM apprenticeships apart from factors affecting all students of the same gender, grade, or school. Exhibit 3.10 shows that there were no significant impacts of being in the treatment condition on math achievement.[13] Appendix C.2 presents additional information on the exploratory academic impact models.

[13] Had impact estimates been significant, they would have been roughly equivalent to one month of growth in math in a school year. Growth in math was calculated using effect-size benchmarks aggregated from various national standardized tests (Bloom et al., 2008; p. 16). Average yearly gains in math for middle school students were divided by nine to estimate how much growth occurs in each month of school. These monthly estimates were converted to the number of months of growth in the present study.

Exhibit 3.10 Impact of Participation in STEM Apprenticeships on Academic Outcomes, Overall

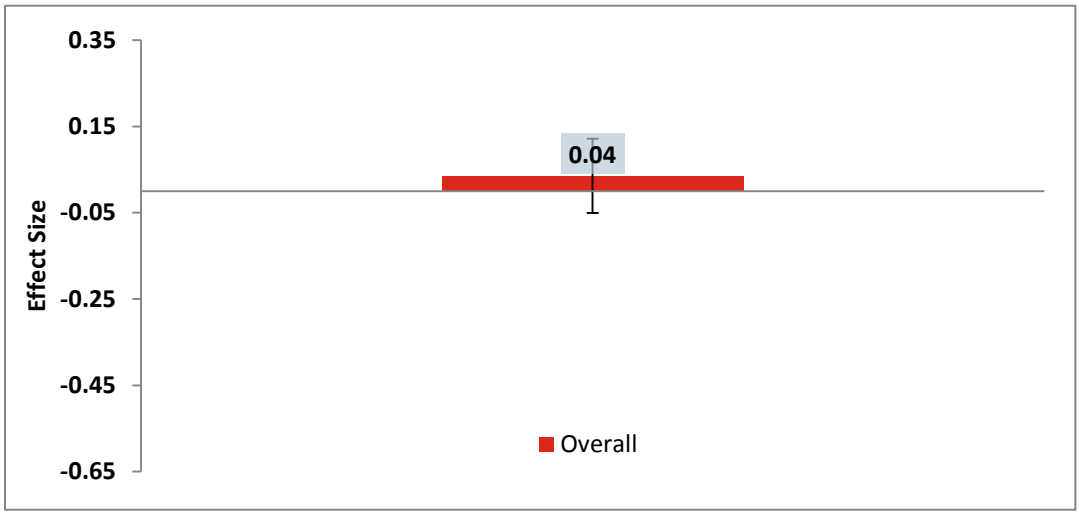


EXHIBIT READS: Students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.04 standard deviations higher on a standardized math test than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant. The upper confidence limit of the impact is 0.12, the lower confidence limit is -0.05.

Notes:

† marginally significant at $p < .10$ level

* statistically significant at $p < .05$ level

** statistically significant at $p < .01$ level

*** statistically significant at $p < .001$ level

Source: Student-level achievement data from spring 2015 and spring 2016 from schools in the Citizen Schools ELT Network.

Sample: Standardized math test scores from 1,155 Citizen Schools ELT students.

Additional exploratory analyses of the academic outcome included examining impacts by gender and grade to learn whether, for example, males in the treatment condition experienced different effects of participation in STEM apprenticeship(s). Recall that because these subgroup analyses divide the sample into smaller groups of students, and statistical power hinges on the number of students in a subgroup, they generally have less power than the full-sample analyses. Nonetheless, Exhibits 3.11 and 3.12 suggest differential impacts by gender and grade. Participation in at least one STEM apprenticeship had a significant impact on 6th grade students, an effect roughly equivalent to four months of math growth in a school year. This finding is particularly noteworthy given that 6th grade is a foundational year for Citizen Schools programming, and the one grade served consistently across all participating campuses in the study. Exploratory analyses conducted as part of the national evaluation of Citizen Schools similarly revealed impacts by grade, finding a marginally significant impact of participating in Citizen Schools on math achievement outcomes for 7th graders (Rulf Fountain et al., 2016). Participation in at least one STEM apprenticeship also had a marginally significant impact on male students, an effect roughly equivalent to three months of math growth in a school year.

Exhibit 3.11. Impacts of Participation in STEM Apprenticeships on Academic Outcomes, by Gender

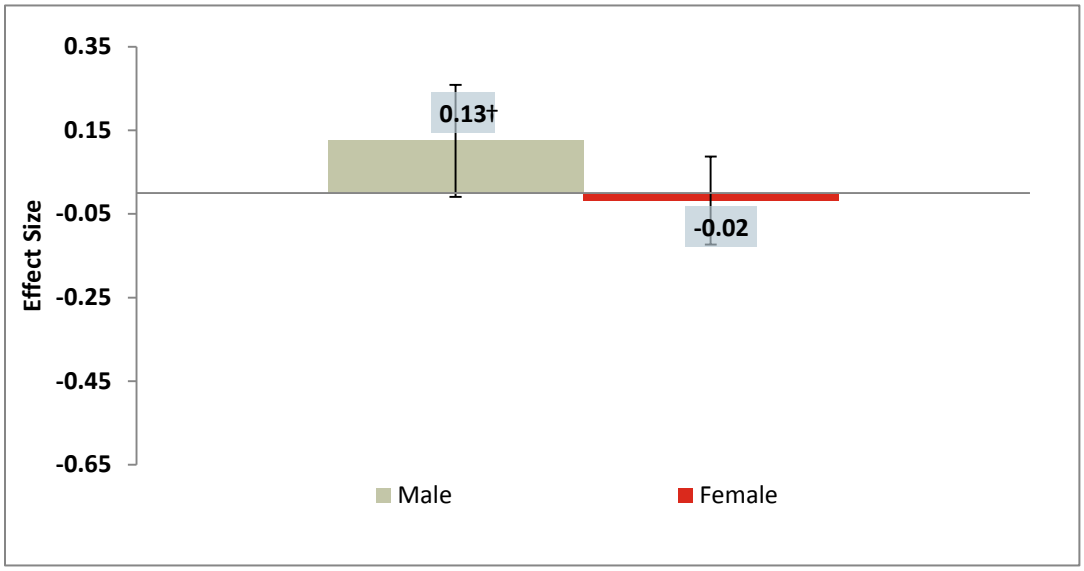


EXHIBIT READS: Male students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.13 standard deviations higher on a standardized math test than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was marginally significant. The upper confidence limit of the impact is 0.26, the lower confidence limit is -0.01.

Notes:

† marginally significant at $p < .10$ level
* statistically significant at $p < .05$ level
** statistically significant at $p < .01$ level
*** statistically significant at $p < .001$ level

Source: Student-level achievement data from spring 2015 and spring 2016 from schools in the Citizen Schools ELT Network.

Sample: Standardized math test scores from 1,155 Citizen Schools ELT students, 595 males and 560 females.

Exhibit 3.12. Impacts of Participation in STEM Apprenticeships on Academic Outcomes, by Grade

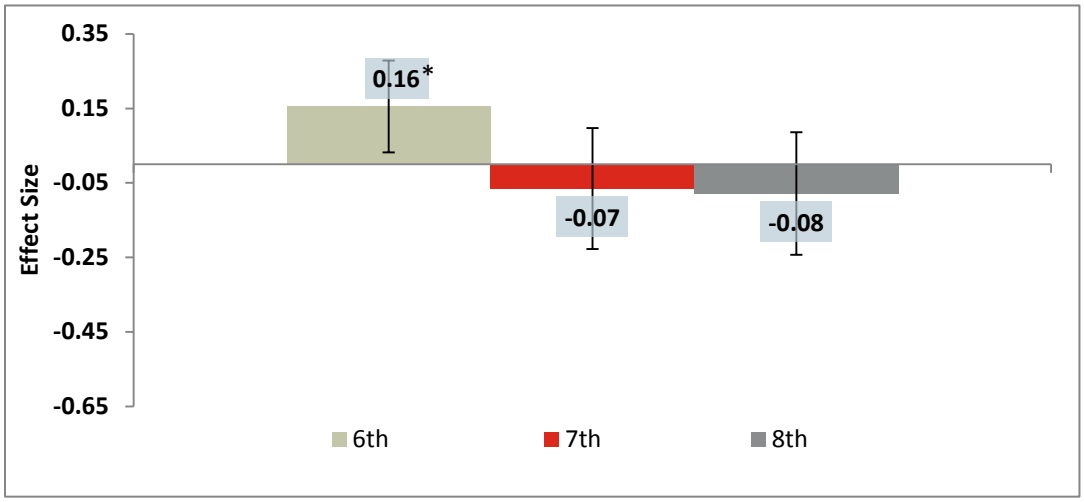


EXHIBIT READS: Sixth grade students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.16 standard deviations higher on a standardized math test than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was statistically significant. The upper confidence limit of the impact is 0.28, the lower confidence limit is 0.03.

Notes:

† marginally significant at p < .10 level
* statistically significant at p < .05 level
** statistically significant at p < .01 level
*** statistically significant at p < .001 level

Source: Student-level achievement data from spring 2015 and spring 2016 from schools in the Citizen Schools ELT Network.

Sample: Standardized math test scores from 1,155 Citizen Schools ELT students, 640 6th graders, 330 7th graders, and 185 8th graders.

3.3 Summary

As the Citizen Schools logic model illustrates, the program is hypothesized to operate through its outputs (activities) to affect short-term outcomes first, by providing students with STEM exposure and experiences that may alter their STEM outlook and skillsets, and that may also influence academic achievement more distally. To assess this spectrum of activities, we first examined how well the ELT apprenticeship model was carried out with fidelity, as successful implementation is a precursor to expected student outcomes. Overall, the program met fidelity thresholds for two of the five measured components, suggesting that there were potential implementation challenges, particularly in the timely recruitment of new Teaching Fellows and in sustaining high student attendance throughout the school year. Analysis of data recording weekly apprenticeship implementation, however, included positive reports from Teaching Fellows in the areas of student engagement, benefits of the apprenticeship to students, and performance of the Citizen Teachers.

Analyses examining student short-term outcomes indicate that students in both conditions scored above the mid-point on both the 21st Century Skills and STEM interest scales at baseline, and that follow-up scores did not differ significantly between treatment and comparison students. It is possible that students had less room to increase over time given their baseline scores, and/or that the 7th and 8th grade students in the sample had had previous exposure to STEM apprenticeships in previous school years.

Results for the distal academic achievement outcome were similar to prior quasi-experimental research on the relationship between after-school ELT programming and academic achievement that have not found consistently positive, significant impacts of ELT on achievement, including a recent evaluation conducted on the national network of Citizen Schools programming (Rulf Fountain et al., 2016). However, similar to the national evaluation of Citizen Schools, exploratory analyses revealed suggestive positive results, including a statistically significant positive impact on 6th graders' math achievement scores, and a marginally significant positive impact for male students.

4. Conclusion

This report summarizes the findings from an evaluation of Citizen Schools' apprenticeships funded by an Investing in Innovation (i3) grant. The study focuses on the 2015-16 school year, during which Citizen Schools worked with 21 schools across several urban school districts to provide project-based apprenticeships as part of Citizen Schools' expanded learning time (ELT) model. Unlike earlier studies that examined the Citizen Schools ELT program model as a whole, this study focuses specifically on one central element of Citizen Schools' programming: apprenticeships. In participating campuses, Citizen Schools works with local community volunteers, called Citizen Teachers, who work with Citizen Fellows to teach 10-week apprenticeships; in the apprenticeships, civic-minded professionals from the local business and professional communities, engage students in project-based learning experiences about which they have specialized knowledge. Over the past half-dozen years, Citizen Schools has focused more specifically on refining its apprenticeship offerings. According to stakeholder feedback, apprenticeships represent one of the strongest program elements and are appreciated by students, school faculty and staff, and local community members. The positive reinforcement about apprenticeships led the organization to focus on streamlining its apprenticeship processes and offerings, addressing aspects that were perceived to be implemented inconsistently across regions and/or schools, and increasing the number and quality of apprenticeships about STEM-related subjects.

The Abt study team worked with Citizen Schools to design a study focused primarily on whether and how STEM apprenticeships impacted attitudinal and achievement outcomes of participating students. In close collaboration with Citizen Schools, the Abt study team developed a quasi-experimental study to compare the outcomes of students who participated in at least one STEM-focused apprenticeship to those who participated in apprenticeships focused on other topics. Short-term, attitudinal outcomes included 21st Century Skills and STEM interest; longer-term achievement outcomes included math standardized tests. Additionally, the evaluation examined fidelity of implementation, to learn whether and how the apprenticeships were implemented as intended. The study measured implementation fidelity using organization records about whether apprenticeships were designated as STEM vs. non-STEM, how many students participated in each apprenticeship, Citizen Teacher and Citizen Fellow attendance at training events, student attendance records, and data routinely collected by Citizen Schools about each apprenticeship session.

4.1 Summary of Findings

Fidelity of implementation was measured using five key components: core STEM curriculum development; partner and community volunteer recruitment; Citizen Schools' staff training; Citizen Schools' staff monitoring; and hands-on activities for students. Results indicate that the program met pre-established thresholds for two of the five components: core STEM curriculum development and Citizen School staff monitoring. Closer examination of the fidelity component data for those components that did not meet thresholds revealed that the program was close to meeting the targets for partner and community volunteer recruitment, but that there were challenges in hiring new Teaching Fellows in a timely manner and ensuring consistent student attendance in apprenticeships. The data also revealed successes in the development of core STEM curriculum and other staff trainings, along with positive reports from Teaching Fellows about the quality of Citizen Teachers and levels of student engagement during apprenticeships.

Findings about the impact of participation in at least one STEM-focused apprenticeship on student outcomes revealed mixed, but also promising, results. Analyses indicate that participation in STEM-focused apprenticeships has no impact on students' 21st Century Skills or STEM interest (the confirmatory outcomes). However, it is possible that the selected measures were not the most suitable for assessing changes in students' 21st Century Skills and STEM interest; in fact, because all participating students—whether or not they participated in STEM-focused apprenticeships—scored above the midpoint on baseline assessments, there may well have been limited room (on those measures) for substantial positive growth. Students' prior experience with STEM-rich curricula may be an important factor to consider in future evaluation. It is also possible that the wide diversity in the apprenticeship topics means that the "STEM" focus of the STEM apprenticeships was so variably operationalized that the "intervention" being tested was not a focused and consistent experience. One other important point is that four (of 21) schools did not submit data about the confirmatory (student survey-based) outcomes, so the findings do not apply to the entire network of Citizen Schools campuses participating in the study.

Exploratory findings suggest positive impacts on subgroups for math achievement test scores, including a statistically significant positive impact on 6th graders' math achievement scores, and a marginally significant positive impact for male students. Perhaps the exposure to STEM-related content for 6th grade students represented a more meaningful departure from the types of STEM learning they had experienced in earlier grades and the 7th and 8th graders had already benefited from Citizen Schools apprenticeships in prior years. It is also worth noting that there are roughly twice as many 6th than 7th graders, and almost three times more 6th than 8th graders (640 6th graders, 330 7th graders, and 185 8th graders, respectively). [The numbers of girls and boys are much more balanced (560 females and 595 males, respectively.] Generally, however, the underlying mechanisms that might help explain the exploratory findings are not clear, given the non-significant proximal measures.

As reflected in the Citizen Schools theory of change, the underlying logic is that through participation in STEM-focused apprenticeships, students' 21st Century Skills and their STEM interest would increase, and subsequently, their academic learning (as measured by math achievement test scores) would improve. The exploratory findings about student achievement examine a relatively distal outcome that one might hypothesize would not be influenced over the course of a one-year study. Further, even though the apprenticeships target STEM (and not math alone), the study uses math achievement scores, because they are generally more readily available across the grade levels served by Citizen Schools than are science achievement scores. Here, too, however, the study had incomplete data, as five schools (of 21) did not provide academic achievement data.

4.2 Take-Aways

One clear lesson learned in the Citizen Schools i3 study is that implementing an intervention as intended can be challenging. Despite setting fidelity thresholds at what appeared to be achievable levels, the variability in school- and organization-level processes resulted in challenges to consistent program implementation. Although Citizen Schools did not achieve the overall fidelity standard defined at the outset of the study, they were close. Of the three program components in which fidelity was not achieved, two were missed by small margins. The rigor of fidelity thresholds in a program model that is implemented across a diverse set of school partnerships nationally is another factor to consider for future evaluation.

Another consideration is the ability to reliably retrieve data. One factor that may have contributed to the results of the study is that approximately one-quarter of participating schools did not provide either

survey or student achievement data, despite considerable effort on the part of Citizen Schools staff and the study team. While the available data were sufficient to address the study's key questions, the absence of four or five schools (from a group of 21) means that results from some analyses can only generalize to the schools that did submit data, and Citizen Schools was not able to benefit from analyses of the network as a whole.

Together, the results from the Citizen Schools i3 study have revealed some successes in the areas of implementation and student impact, as well as areas for growth. Additionally, the results open up promising avenues for future investigation—from measuring cumulative effects of multiple years of participation in Citizen Schools to identifying or developing measures targeted specifically for the program. As the Citizen Schools network continues to redefine itself, in particular, moving towards a primary focus on the provision of STEM apprenticeships, the findings in this report can provide useful information for the organization as it plans for the future.

References

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1, 289–328.

Borjas, G. J. (2005). The labor-market impact of high-skill immigration. *American Economic Review*, *95*(2), 56-60.

Carnegie Corporation and the Institute for Advanced Study. (2009). The opportunity equation: Transforming mathematics and science education for citizenship and the global economy. New York, New York.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, S95-S120.

Corporation for National and Community Service. (2013). President Obama announces STEM AmeriCorps to inspire young people's interest in science and technology [Press Release]. Available at: http://www.nationalservice.gov/programs/americorps/stem-americorps. Accessed July 26, 2013.

Corporation for National and Community Service. (2014). National service agency announces expansion of STEM AmeriCorps [Press Release]. Available at: http://www.nationalservice.gov/newsroom/press-releases/2014/national-service-agencyannounces-expansion-stem-americorps. Accessed June 3, 2014.

Desy, E. A., Peterson, S. A., & Brockman, V. (2011). Gender differences in science-related attitudes and interests among middle school and high school students. *Science Educator*, *20*(2), 23.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of Mentoring Programs for Youth: A Meta-Analytic Review. American Journal of Community Psychology, Vol. 30, No. 2, April 2002, 157-197.DuBois, D. L., & Silverthorn, N. (2005). Natural mentoring relationships and adolescent health: Evidence from a national study. *American Journal of Public Health*, 95, 518-524.

Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & DuBois, D. L. (2008). Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior*, *72*(2), 254-267.

Flick, L. B. (1993). The meanings of hands-on science. *Journal of Science Teacher Education, 4*, 1–8.

Government Accounting Office. (2006). Science, technology, engineering, and mathematics trends and the role of federal programs. Statement of Cornelia M. Ashby, Director, Education, Workforce, and Income Security Issues. Available at: http://www.gao.gov/new.items/d06702t.pdf. Accessed June 3, 2014.

Gerstner, S., & Bogner, F. X. (2010). Cognitive achievement and motivation in hands-on and teacher-centered science classes: Does an additional hands-on consolidation phase (concept mapping) optimize cognitive learning at work stations? *International Journal of Science Education, 32*(7), 849-870.

Haury, D. L., & Rillero, P. (1994). Perspectives of hands-on science teaching. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Higher Education Research Institute Research Brief. (2010). Degrees of success. Accessible at www.heri.ucla.edu/nih/HERI_ResearchBrief_OL_2010_STEM.pdf.

Ho, D., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Policy Analysis*, *15*(3), 199–236.

Jekielek, S. M., Moore, K. A., Hair, E. C., & Scarupa, H. J. (2002). Mentoring: A promising strategy for youth development (Child Trends Research Brief). Washington, DC: Child Trends.

Johnson, D. W & Johnson, R. T. (1986). Encouraging student/student interaction. Washington DC: National Association for Research in Science Teaching (ERIC Document Reproduction Service No. ED 266960).

Karcher, M. J. (2008). The Study of Mentoring in the Learning Environment (SMILE). A randomized study of the effectiveness of school-based mentoring. Prevention Science. Karp, T., & Schneider, A. (2011). Evaluation of a K-8 LEGO robotics program. Frontiers in Education Conference (FIE), October, 2011 (pp. T1D-1). IEEE.

LaBanca, F., Oh, Y.J., Lorentson, M., Jia, Y., & Sibuma, B. (2014). Development and Validation of a 21st Century and Inquiry Skill Assessment in Science, Technology, Engineering, and Mathematics for High School Students. In F. LaBanca (Ed.), *A Battery of 21st Century and Inquiry Self Assessment Instruments* (pp. 4-12). Litchfield, CT: Century for 21st Century Skills at EDUCATION CONNECTION.

Lemelson Center. (2010). 2010 Lemelson-MIT Invention Index Reveals Ways to Enhance Teens' Interest in Science, Technology, Engineering and Mathematics in The Classroom and Beyond (press release). Available at: http://web.mit.edu/invent/n-pressreleases/n-press-10index.html.

National Center for Education Statistics (NCES). (2013). Nation's report card: A first look: 2013 mathematics and reading. U.S. Department of Education, Institute of Education Sciences. Available at: http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf

National Research Council. (2007). Rising Above the gathering storm: Energizing and employing America for a brighter economic future. Committee on Prospering in the Global Economy of the 21st Century. Washington, DC: The National Academies Press.

National Science Board. (2010). Preparing the next generation of STEM innovators: Identifying and developing our nation's human capital. Arlington VA: National Science Foundation (NSB-1033).

National Science Board. (2014). Science and Engineering Indicators 2014. Arlington VA: National Science Foundation (NSB 14-01).

National Science and Technology Council. (May 2013). Federal science, technology, engineering, and mathematics (STEM) education 5-year strategic plan: A report from the committee on STEM education. Washington, DC: Executive Office of the President. Available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/stem_stratplan_2013.pdf

Oh, Y.J. (2014). DEV14 STEM21 Academy Evidence Report. Draft.

President's Council of Advisors on Science and Technology. (2010). Report to the President. Prepare and inspire: K-12 education in science, technology, engineering, and math (STEM) for America's future.

Programme for International Student Assessment (2012). United States—Country note: Results from PISA 2012. Available at: http://www.oecd.org/pisa/keyfindings/PISA-2012-results-US.pdf. Accessed June 3, 2014.

Programme for International Student Assessment. (2013). PISA 2012 results in focus: What 15-yearolds know and what they can do with what they know. OECD. Available at: http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf. Accessed June 3, 2014.

Randler, C., & Hulde, M. (2007). Hands-on versus teacher-centered experiments in soil ecology. Research in Science & Technological Education, 25(3), 329-338. Redmond, A., Thomas, J., High, K., Scott, M., Jordan, P. & Dockers, J. (2011). Enriching science and math through engineering. *School Science and Mathematics, 111*, 399–408. doi: 10.1111/j.19498594.2011.00105.x.

Rhodes, J. E., & DuBois, D. L. (2008). Mentoring relationships and programs for youth. *Current Directions in Psychological Science, 17*(4), 254-258.

Rulf Fountain, A., Gamse, B., Velez, M., Hillard, M., and P. Cropper. 2016. *Evaluation of Citizen Schools' Expanded Learning Time Model: Final Report* prepared for Citizen Schools Cambridge., MA: Abt Associates.

Schochet, P. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics, 33*(1), 62–87.

Tai, Robert H., Liu, C.Q., Maltese, A.V., & Fan, X... (2006). Planning Early for Careers in Science. Science 312 (May 26): 1143-1144.

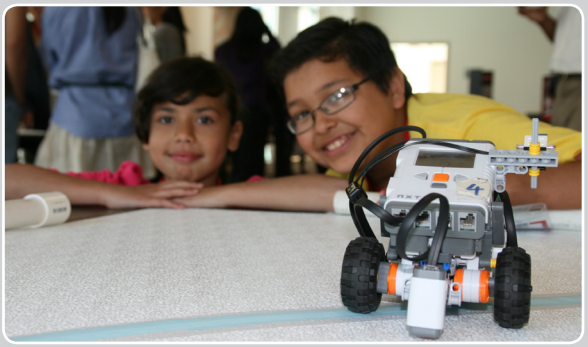
Tierney, J. P., Grossman, J. B., & Resch, N. L. (2000). Making a Difference: An impact study of Big Brothers Big Sisters. Philadelphia: Public/Private Ventures.

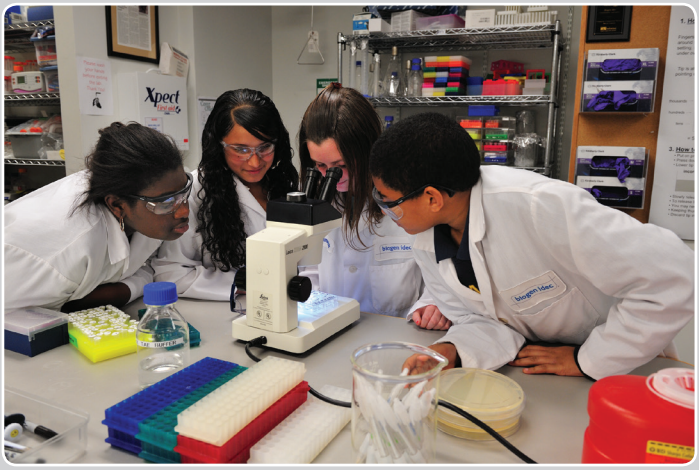
U.S. Department of Education, What Works Clearinghouse. (2017). *Procedures and standards handbook (version 4.0)*. Washington, DC: US Department of Education, Institute for Education Sciences. Available at https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf (accessed November 7, 2017).

Weinburgh, M.E. & Steele, D. (2000). The modified attitudes toward science inventory: Developing an instrument to be used with fifth grade urban students. *Journal of Women and Minorities in Science and Engineering, 6*(1), 87-94.

Werner, E. E., & Smith, R. S. (1992). Overcoming the odds: High risk children from birth to adulthood. Cornell University Press.

Xie, Y., & Achen, A. (2009). Science on the decline? Educational outcomes of three cohorts of young Americans. *Population Studies Center Research Report, 9*, 684.





Closing Inspiration and Achievement Gaps in STEM with Volunteer-Led Apprenticeships: Appendices

SEPTEMBER 27, 2018

PREPARED FOR:
Citizen Schools
308 Congress St., 5th Floor
Boston, MA 02210

SUBMITTED BY:
Alyssa Rulf Fountain
Melissa Velez
Beth Gamse
Abt Associates
10 Fawcett Street
Cambridge, MA 02138

VOLUME I

Appendix A

Appendix A.1 Model Specification for Checking Baseline Equivalence

The model specification for checking baseline equivalence is as follows:

$$Y_{ij} = \beta_0 + \beta_1\left(TRMT_{ij}\right) + \sum_{j=1}^{J-1}\gamma_j\left(Block_j\right) + \varepsilon_{ij}$$

where:

Y_{ij} = outcome measure for student i in school j at baseline (measured in spring 2015).

$TRMT_{ij}$ = indicator variable for students who participate in at least one STEM apprenticeship during the 2015-16 academic year. Specifically, it equals one if a student i in school j participated in at least one STEM apprenticeship and zero otherwise.

$Block_j$ = school dummy variables, modeled by a series of indicator variables for each school.

ε_{ij} = the usual student-level error term.

Appendix A.2 Characteristics of 21st Century Skill and STEM Interest Student Samples

Characteristics of 21st Century Skills Student Sample						
	Treatment		Comparison		All students	
	N	Percent	N	Percent	N	Percent
Gender						
Female	380	45.4	106	62.4	486	48.2
Male	458	54.7	64	37.7	522	51.8
Grade						
6th	550	65.6	100	58.8	650	64.5
7th	236	28.2	29	17.1	265	26.3
8th	52	6.2	41	24.1	93	9.2
All students	838	83.1	170	16.9	1008	

EXHIBIT READS: Females comprised 45.4 percent of the treatment sample, 62.4 percent of the comparison sample, and 48.2 percent of the sample overall.

Source: Citizen Schools student surveys and administrative records from the 2015-16 school year.

Sample: 1008 Citizen Schools students in the 21st Century Skills analytic sample.

Characteristics of STEM Interest Student Sample						
	Treatment		Comparison		All students	
	N	Percent	N	Percent	N	Percent
Gender						
Female	392	45.6	112	62.2	504	48.5
Male	468	54.4	68	37.8	536	51.5
Grade						
6th	565	65.7	102	56.7	667	64.1
7th	242	28.1	34	18.9	276	26.5
8th	53	6.2	44	24.4	97	9.3
All students	860	82.7	180	17.3	1040	

EXHIBIT READS: Females comprised 45.6 percent of the treatment sample, 62.2 percent of the comparison sample, and 48.5 percent of the sample overall.

Source: Citizen Schools student surveys and administrative records from the 2015-16 school year.

Sample: 1040 Citizen Schools students in the STEM interest analytic sample.

Appendix A.3 NEi3 Fidelity Matrix and Threshold Scores

Indicators	Definition	Unit of implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement	Program level score
Key Component 1 = Core STEM Curriculum Development											
STEM curriculum	Development of at least 21 STEM curriculum units	Program	Citizen Schools report	Citizen Schools; end of 2015-16 SY	0=1-10 STEM curriculum units developed= low (score=0) 1=11-20 STEM curriculum units developed= moderate (score=1) 2=21-30 STEM curriculum units developed= high (score=2)	Adequate fidelity = 2 (21-30 STEM curriculum units)			All curriculum units	2015-16	2
Key Component 2 = Partner and Community Volunteer Recruitment											
Recruitment of Citizen Teachers	# of students per apprenticeship does not exceed 24	All apprenticeships (~400)	Citizen Schools report	Citizen Schools; end of fall 2015 and spring 2016 semesters	At the apprenticeship level: 0 = more than 24 students in apprenticeship 1 = less than 24 students in apprenticeship	Adequate fidelity at apprenticeship level = 1	At the program level: 0=less than 85% of apprenticeships have fewer than 24 students; 1=85%-94% of apprenticeships had fewer than 24 students; 2=95% or more of apprenticeships had fewer than 24 students	Adequate fidelity at sample level = 2 (or 95% of apprenticeships had fewer than 24 students)	All apprenticeships	2015-16	1

Indicators	Definition	Unit of implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement	Program level score
Key Component 3 = Citizen Schools staff training											
Indicator 1: Summer institute (TF1)	Attendance at a 5 day summer institute training for new TFs (called TF1s)	TF1	Attendance records	Citizen Schools report; end of summer	0 = Attended less than 4 days; 1=attended 4 or more days	Adequate fidelity = 1	Program: 0= 70% or fewer TF1s attended 4 or more days; 1 = 70-85% attended 4 or more days; 2=86-100% attended 4 or more days	Adequate fidelity at the program level = 2 (86-100% of TFs attend 4 or more days of institute)	All TF1s	2015-16	1
Indicator 2: WOW training	Attendance at training once per semester that fellow is teaching (fellows may teach 1 or 2) apprenticeships over the year).	TF (all)	Attendance records	Citizen Schools report; end of each semester	0 = attended no WOW training 1 = attended all relevant WOW trainings (either 1 or 2 semesters, depending on whether fellow taught an apprenticeship two semesters)	Adequate fidelity for a fellow = score of 1	Program: 0= less than 70% of TFs attended training; 1 = 70-85% of TFs attended training; 2=86-100% of TFs attended training;	Sample level threshold for adequate fidelity = score of 2 (86-100% of TFs receive a score of 1 on their attendance)	All TFs who teach at least one apprenticeship	2015-16	2
Indicator 3: Training for new Citizen Teachers (CT1)	Attendance at trainings	CT1	Attendance records	Citizen Schools report; end of each semester	0 = attended no training 1 = attended training	Adequate fidelity for a CT1 = score of 1	Program: 0= less than 70% of CT1s attended training; 1 = 70-85% of CT1s attended training; 2=86-100% of CT1s attended training	Adequate fidelity for the program = score of 2	All CT1s	2015-16	2
All indicators								Sum of scores across three indicators Threshold for adequate implementation= 6	All TFs and CT1s	2015-16	5

Indicators	Definition	Unit of implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement	Program level score
Key Component 4 = Citizen Schools Staff Monitoring											
Monitoring delivery of curriculum	Completion of weekly trackers about apprenticeship (11 weeks in each semester)	Apprenticeship (~400)	Weekly tracker completed by Citizen Schools staff	Program file after each week	At the apprenticeship level, 0 (low)=report submitted less than nine times per apprenticeship; 1(high) = report submitted 9 or more times per apprenticeship	Adequate Fidelity at Apprenticeship Level = 1	Program: 0= less than 70% of apprenticeships score high; 1 = 70-85% are high; 2=86-100% are high	Adequate fidelity at program level = 2	All apprenticeships	2015-16	2
Key Component 5 = Hands-On Activities for Students (including those with a STEM focus)											
Apprenticeship	Attendance at an apprenticeship	Student (3,000 – 4,000)	Attendance sheets	Citizen Schools report; at the end of each semester	At the student level; 0 = did not attend at least 9 weeks of all apprenticeships 1 = attended 9 or more weeks of all apprenticeships	Adequate fidelity at the student level = 1	At the program level: 0= less than 70% of students have high; 1 = 70-85% are high; 2=86-100% are high	Adequate fidelity = 2	All students	2015-16	1
										TOTAL	11

Appendix A.4 Impact Model Specification

The model specification for assessing overall impacts is as follows (confirmatory analysis). Models for assessing impacts by subgroups (exploratory analyses) were identical to the model presented below, but included an interaction between the treatment indicator and the subgroup of interest (one model with a dummy variable for male and another model with dummy variables for grade, with 8th grade being the omitted category):

$$Y_{ij} = \beta_0 + \beta_1(TRMT_{ij}) + \beta_2(\Pr e_{ij}) + \sum_{j=1}^{J-1} \gamma_j(Block_j) + \sum_{m=1}^{M} \lambda_m X_m + \varepsilon_{ij}$$

where:

Y_{ij} = outcome measure for student i in school j after one year of the intervention (measured in spring 2016).

$TRMT_{ij}$ = indicator variable for students who participate in at least one STEM apprenticeship during the 2015-16 academic year. Specifically, it equals one if a student i in school j participated in at least one STEM apprenticeship and zero otherwise.

Pre_{ij} = is student pretest.

$Block_j$ = school dummy variables, modeled by a series of indicator variables for each school.

X_m = student-level characteristics, including gender and grade.

ε_{ij} = the usual student-level error term.

Appendix B

Appendix B.1 Supplementary Information on Fidelity Measures

	Average Number of Students in Apprenticeships	Percentage of First Year Teaching Fellows Hired Late	Percentage of Students Attending >= 9 Apprenticeship Weeks Across All Apprenticeships	Percentage of Students Attending >= 9 Apprenticeship Weeks, Fall Apprenticeships	Percentage of Students Attending >= 9 Apprenticeship Weeks, Spring Apprenticeships
Region A					
School A	26	60%	75%	86%	87%
School B	25	40%	64%	78%	76%
School C	25	83%	81%	88%	87%
School D	23	17%	66%	82%	75%
School E	17	0%	63%	79%	78%
Region B					
School F	19	0%	64%	77%	74%
School G	19	0%	72%	89%	77%
School H	18	0%	80%	90%	88%
School I	16	33%	74%	80%	89%
Region C					
School J	12	20%	86%	92%	91%
School K	16	10%	69%	84%	80%
School L	18	50%	81%	90%	88%
School M	20	33%	80%	89%	91%
Region D					
School N	12	50%	73%	85%	86%
School O	13	38%	72%	82%	85%
School P	12	0%	69%	81%	77%
School Q	14	13%	80%	89%	86%
School R	14	27%	59%	72%	67%
School S	12	50%	68%	84%	79%
School T	18	33%	64%	76%	81%
Region E					
School U	19	N/A[1]	75%	85%	77%

EXHIBIT READS: There was an average of 26 students per apprenticeship in School A.

Notes:

[1]There were no new Teaching Fellows hired at this school during the 2015-16 school year.

Source: Citizen Schools administrative records from the 2015-16 school year.

Appendix B.2 Weekly Apprenticeship Tracker Results

Areas of Success					
Identify the areas of success this week (check all that apply)	N	Mean	Standard Deviation	Min	Max
High student attendance	5109	0.7	0.5	0	1
High student engagement/enthusiasm	5109	0.7	0.5	0	1
Apprenticeship session covered an appropriate amount of content	5109	0.7	0.4	0	1
Apprenticeship session achieved all or most of the students' 21st century skill learning objectives	5109	0.6	0.5	0	1

EXHIBIT READS: Across all apprenticeships, the proportion of weeks that high student attendance was identified as an area of success was 0.7.
Notes:

Weekly Tracker responses were aggregated across all apprenticeships and weeks (N = 5109, Missing = 0).

Responses do not sum to 100% because multiple responses were permitted.

Source: 2015-16 Apprenticeship Weekly Tracker Q5 (Identify areas of success this week (check all that apply))

Areas of Concern					
Identify the area(s) of concern with the CT relationship or the quality of the apprenticeship this week (check all that apply)	N	Mean	Standard Deviation	Min	Max
Student behavior issues negatively affected apprenticeship	5109	0.2	0.4	0	1
Apprenticeship session achieved only some or none of the 21st century skill learning objectives	5109	0.2	0.3	0	1
Low student attendance	5109	0.1	0.3	0	1
Low student engagement/enthusiasm	5109	0.1	0.3	0	1
Concern for student safety	5109	<0.1	0.1	0	1
CT(s) was/were unprepared	5109	<0.1	0.2	0	1
Apprenticeship session covered very little or none of the intended curriculum content	5109	<0.1	0.2	0	1

EXHIBIT READS: Across all apprenticeships, the proportion of weeks that student behavior issues negatively affected the apprenticeship was 0.2.
Notes:

Weekly Tracker responses were aggregated across all apprenticeships and weeks (N = 5109, Missing = 0)

Responses do not sum to 100% because multiple responses were permitted.

Source: 2015-16 Apprenticeship Weekly Tracker Q6 (Identify areas of concern with the CT relationship or the quality of the apprenticeship this week (check all that apply))

Citizen Teachers Performance					
	N	Mean	Standard Deviation	Min	Max
Degree of CT(s)'s command of given week's lesson	5107	3.5	0.7	1	4
Degree to which CT(s) was/were prepared for week's lesson	5107	3.4	0.7	1	4
Degree to which CT(s) delivered a well-executed lesson	5107	3.2	0.8	1	4
Degree of CT(s)'s rapport with the students	5107	3.2	0.7	1	4

EXHIBIT READS: Across all weeks of all apprenticeships, the average degree of CT's command of a given week's lesson was 3.5.

Notes:

Weekly Tracker responses were aggregated across all apprenticeships and weeks (N = 5107, Missing = 2)

Responses were on a Likert Scale. Q7 response options: 1 = not at all prepared to 4 = extremely well prepared; Q8 response options: 1 = not at all well executed to 4 = extremely well executed; Q9 response options: 1 = does not at all understand the content to 4 = the highest level of understanding; Q10 response options: 1 = poor rapport to 4 = extremely good rapport.

Source: 2015-16 Apprenticeship Weekly Tracker Q7 (To what degree was/were CT(s) prepared for this week's lesson (e.g., planned for students' level of understanding, planned to include hands-on apprenticeship centered activities?)); Q8 (To what degree did the CT(s) deliver a well-executed lesson (e.g., had materials ready, appropriately gauged students attention and interest, covered an appropriate amount of content given students' level of understanding)?); Q9 (How would you rate the CT(s)'s command of the content in this week's lesson (e.g., answering student questions, providing varied and real life examples to illustrate concepts)?); Q10 (How would you rate the CT(s)'s rapport with the students (e.g., has connection with all or most students in apprenticeship, knows everyone's name)?)

Student engagement and behavior					
	N	Mean	Standard Deviation	Min	Max
Degree to which students were engaged	5101	3.1	0.7	1	4
Degree to which student behavior negatively impacted apprenticeship learning	4909	1.7	0.8	1	4

EXHIBIT READS: Across all weeks of all apprenticeships, the average degree to which students were engaged was 3.1.

Notes:

Weekly Tracker responses were aggregated across all apprenticeships and weeks Q11 (N = 5101, Missing = 8); Q12 (N = 4909; Missing = 200)

Responses were on a Likert Scale. Q11 response options: 1 = not at all engaged to 4 = extremely engaged; Q12 response options: 1 = no negative impact to 4 = extremely negative impact.

Source: 2015-16 Apprenticeship Weekly Tracker Q11 (How would you rate student engagement (e.g., participating in lesson, doing what was asked of them)?); Q12 (Please rate the extent to which student behavior negatively impacted apprenticeship learning)?)

Perceived Effects on Students					
	N	Mean	Standard Deviation	Min	Max
Degree to which teacher team (CT(s) and TL) was effective in meeting objectives for given week and delivered high-quality apprenticeship	4952	3.9	1.0	1	5
Degree to which students benefitted from week's lesson	4906	3.8	1.0	1	5

EXHIBIT READS: Across all weeks of all apprenticeships, the average degree to the teacher team was effective in meeting objectives for given a week and delivered a high-quality apprenticeship was 3.9.

Notes:

Weekly Tracker responses were aggregated at the apprenticeship level and then Q13 (N = 4906, Missing = 203); Q14 (N = 4952; Missing = 157)

Responses were on a Likert Scale. Q13 response options: 1 = not at all beneficial to 5 = extremely beneficial; Q14 response options: 1 = poor to 5 = excellent.

Source: 2015-16 Apprenticeship Weekly Tracker Q13 (To what degree do you think students benefitted from this week's lesson (e.g., were able to build on or apply knowledge/understanding from earlier apprenticeship sessions)?); Q14 (Overall, how effective was the teaching team (CT(s) and TL) in meeting your objectives for this week and delivering a high-quality apprenticeship?)

Appendix C

Appendix C.1 Supplementary Information on Analyses of Short-Term Outcomes

Confirmatory Outcomes

Non-Academic Outcome Results

	21st Century Skills			STEM Interest		
Treatment	-0.04	0.69	(0.11)	0.01	0.89	(0.05)
Pre-test	0.11	0.00***	(0.03)	0.20	<.0001***	(0.03)
Grade (8th omitted)						
6th	-0.36	0.03*	(0.17)	-0.07	0.37	(0.07)
7th	-0.47	0.01**	(0.18)	-0.10	0.20	(0.08)
Male	0.04	0.61	(0.08)	0.06	0.12	(0.04)
Intercept	4.99	<.0001***	(0.27)	2.87	<.0001***	(0.14)
School Fixed Effects	√			√		
N	1008			1040		

EXHIBIT READS: Students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.04 points lower on the 21st Century Skills scale than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant.
Notes:
† marginally significant at p < .10 level
* statistically significant at p < .05 level
** statistically significant at p < .01 level
*** statistically significant at p < .001 level

Responses on the 21st Century Skills Scale ranged from 1 = Extremely Poor to 7 = Excellent. Responses on the STEM Interest scale ranged from 1 = Strongly Disagree to 5 = Strongly Agree.
Source: Citizen Schools Student Surveys and administrative records from the 2015-16 school year.

Exploratory Outcomes

Non-Academic Outcome Results, by Gender

	21st Century Skills		STEM Interest	
Treatment	0.03	(0.14)	-0.05	(0.06)
Treatment*Male	-0.18	(0.22)	0.13	(0.1)
Pre-test	0.11***	(0.03)	0.20***	(0.03)
Grade (6th omitted)				
7th	-0.11	(0.11)	-0.03	(0.05)
8th	0.37*	(0.17)	0.07	(0.07)
Male	0.19	(0.2)	-0.05	(0.09)
Intercept	3.75***	(0.37)	2.62***	(0.19)
School Fixed Effects	√		√	
N	1008		1040	

EXHIBIT READS: Female students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.03 points higher on the 21st Century Skills scale than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant.
Notes:
† marginally significant at p < .10 level
* statistically significant at p < .05 level
** statistically significant at p < .01 level
*** statistically significant at p < .001 level

Responses on the 21st Century Skills Scale ranged from 1 = Extremely Poor to 7 = Excellent. Responses on the STEM Interest scale ranged from 1 = Strongly Disagree to 5 = Strongly Agree.
Source: Citizen Schools Student Surveys and administrative records from the 2015-16 school year.

Non-Academic Outcome Results, by Grade

	21st Century Skills		STEM Interest	
Treatment	0.04	(0.14)	0.03	(0.06)
Treatment*7th Grade	-0.32	(0.29)	-0.08	(0.12)
Treatment*8th Grade	-0.11	(0.31)	-0.01	(0.14)
Pre-test	0.11***	(0.03)	0.20***	(0.03)
Grade (6th omitted)				
7th	0.17	(0.27)	0.04	(0.12)
8th	0.45†	(0.25)	0.08	(0.11)
Male	0.04	(0.08)	0.05	(0.04)
Intercept	3.75***	(0.37)	2.58***	(0.19)
School Fixed Effects	√		√	
N	1008		1040	

EXHIBIT READS: Sixth grade students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.04 points higher on the 21st Century Skills scale than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant.
Notes:
† marginally significant at p < .10 level
* statistically significant at p < .05 level
** statistically significant at p < .01 level
*** statistically significant at p < .001 level

Responses on the 21st Century Skills Scale ranged from 1 = Extremely Poor to 7 = Excellent. Responses on the STEM Interest scale ranged from 1 = Strongly Disagree to 5 = Strongly Agree.
Source: Citizen Schools Student Surveys and administrative records from the 2015-16 school year.

Appendix C.2 Supplementary Information on Exploratory Analyses of Student Longer Term Outcomes

Exploratory Outcomes

Impact of Participation in STEM Apprenticeships on Academic Outcomes, Overall

	Math Achievement	
Treatment	0.04	(0.04)
Pre-test	0.76***	(0.02)
Grade (8th omitted)		
6th	-0.18**	(0.06)
7th	-0.22***	(0.06)
Male	0.01	(0.03)
Intercept	0.02	(0.08)
School Fixed Effects	√	
N	1155	

EXHIBIT READS: Students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.04 standard deviations higher on a standardized math test than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant.
Notes:
† marginally significant at p < .10 level
* statistically significant at p < .05 level
** statistically significant at p < .01 level
*** statistically significant at p < .001 level
Source: Student-level achievement data from spring 2015 and spring 2016 from schools in the Citizen Schools ELT Network.

Sample: Standardized math test scores from 1,155 Citizen Schools ELT students.

Impacts of Participation in STEM Apprenticeships on Academic Outcomes, by Gender

	Math Achievement	
Treatment	-0.02	(0.05)
Treatment*Male	0.14†	(0.08)
Pre-test	0.76***	(0.02)
Grade (8th omitted)		
6th	-0.19**	(0.06)
7th	-0.23***	(0.06)
Male	-0.11	(0.08)
Intercept	0.07	(0.08)
School Fixed Effects	√	
N	1155	

EXHIBIT READS: Female students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.02 standard deviations lower on a standardized math test than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant.

Notes:

† marginally significant at p < .10 level
* statistically significant at p < .05 level
** statistically significant at p < .01 level
*** statistically significant at p < .001 level

Source: Student-level achievement data from spring 2015 and spring 2016 from schools in the Citizen Schools ELT Network.
Sample: Standardized math test scores from 1,155 Citizen Schools ELT students, 595 males and 560 females.

Impacts of Participation in STEM Apprenticeships on Academic Outcomes, by Grade

	Math Achievement	
Treatment	-0.08	(0.08)
Treatment*6th Grade	0.23*	(0.10)
Treatment*7th Grade	0.01	(0.12)
Pre-test	0.77***	(0.02)
Grade (8th omitted)		
6th	-0.35***	(0.09)
7th	-0.21*	(0.10)
Male	0.00	(0.03)
Intercept	0.10	(0.09)
School Fixed Effects	√	
N	1155	

EXHIBIT READS: Eighth grade students who participated in at least one STEM apprenticeship during the 2015-16 school year scored, on average, 0.08 standard deviations lower on a standardized math test than would have been expected if they had not participated in at least one STEM apprenticeship. The estimated impact was not statistically significant.

Notes:

† marginally significant at p < .10 level
* statistically significant at p < .05 level
** statistically significant at p < .01 level
*** statistically significant at p < .001 level

Source: Student-level achievement data from spring 2015 and spring 2016 from schools in the Citizen Schools ELT Network.
Sample: Standardized math test scores from 1,155 Citizen Schools ELT students, 640 6th graders, 330 7th graders, and 185 8th graders.

VOLUME II

Appendix A: Weekly Apprenticeship Tracker

Electronic Weekly Apprenticeship Tracker[1]

Campus & Apprenticeship Day: *Please select from dropdown*

Apprenticeship Name: *Please select from dropdown*

Apprenticeship Week Number: *Please select from dropdown*

1. What type of communication with your CT(s) did you have before the apprenticeship this week? (Select all that apply)
   - Email
   - Phone
   - In-person
   - None
2. Approximately how much total time did you spend in communication with your CT(s) this week? ____ minutes
3. Please share an area of success/highlight from this week's apprenticeship (optional open response)
4. Please share any areas of concern around your relationship with the CT or the quality of the apprenticeship (optional open response)

Review the items below; check all responses that apply to this week's apprenticeship

5. Identify areas of success this week (check all that apply):

Student Engagement:

- High student attendance
- High student engagement/enthusiasm

CT Preparation & Practices:

- Apprenticeship session covered an appropriate amount of content
- Apprenticeship session achieved all or most of the students' 21st century skill learning objectives

Other:

- If any, please explain): _____

6. Identify any area(s) of concern with the CT relationship or the quality of the apprenticeship this week (check all that apply):

[1] Note that this instrument was administered online exclusively.

Student Engagement:

- Low student attendance
- Low student engagement/enthusiasm
- Concern for student safety
- Student behavior issues negatively affected apprenticeship

CT Preparation & Practices:

- CT(s) was/were unprepared
- Apprenticeship session covered very little or none of the intended curriculum content
- Apprenticeship session achieved only some or none of the 21st century skill learning objectives

Other:

- If any, please explain: _____

Review the items below; select one response that reflects this week's apprenticeship

7. To what degree was/were CT(s) prepared for this week's lesson (e.g., planned for students' level of understanding, planned to include hands-on apprenticeship centered activities)?
    a. Extremely well prepared
    b. Prepared
    c. Somewhat prepared
    d. Not at all prepared

8. To what degree did the CT(s) deliver a well-executed lesson (e.g., had materials ready, appropriately gauged students' attention and interest, covered an appropriate amount of content given students' level of understanding)?
    a. Extremely well executed
    b. Well executed
    c. Somewhat well executed
    d. Not at all well executed

9. How would you rate the CT(s)'s command of the content in this week's lesson (e.g., answering student questions, providing varied and real life examples to illustrate concepts)?
    a. The highest level of understanding
    b. Understands the content
    c. Has some level of understanding
    d. Does not at all understand the content

10. How would you rate the CT(s)'s rapport with the students (e.g., has connection with all or most students in apprenticeship, knows everyone's name)?

    a. Extremely good rapport
    b. Good rapport
    c. Somewhat good rapport
    d. Poor rapport

11. How would you rate student engagement (e.g., participating in lesson, doing what was asked of them)?

    a. Extremely engaged
    b. Engaged
    c. Somewhat engaged
    d. Not at all engaged

12. Please rate the extent to which student behavior negatively impacted apprenticeship learning?

    a. Extremely negative impact
    b. Negative impact
    c. Somewhat negative impact
    d. No negative impact

13. To what degree do you think students benefitted from this week's lesson (e.g., were able to build on or apply knowledge/understanding from earlier apprenticeship sessions)?

    a. Extremely beneficial
    b. Mostly beneficial
    c. Beneficial
    d. Somewhat beneficial
    e. Not at all beneficial

14. Overall, how effective was the teaching team (CT(s) and TL) in meeting your objectives for this week and delivering a high-quality apprenticeship?

    a. Excellent
    b. Very Good
    c. Good
    d. Fair
    e. Poor

15. Would you like to check in with your CTL? (Select "yes" or "no")

Appendix B: Electronic and Paper Student Survey[2]

- **Please <u>do not</u> put your name on this form. This survey is confidential. No one at your school will see your survey.** Only the people who are doing the study at a company called Abt Associates and Citizen Schools will see your answers.
- **<u>This survey is to be filled out by students themselves.</u>**
- **Please return your survey to your teacher. These surveys will then be sent to the study team at Abt Associates.**
- **Thank you for taking time to tell us your views about your school.**

Welcome! Your school is partnering with Citizen Schools to provide expanded learning time during your school day. Students in your school are being asked to complete this survey. There are no "right" or "wrong" answers to any of the questions. Citizen Schools wants to learn what students think of the Citizen Schools program including your apprenticeships. We estimate that it will take about 15 minutes to complete the questions. Thank you very much for your help!

We would like you to participate in a study by completing this survey, but you do not have to participate. If you decide not to participate, no one at Citizen Schools or your school will be upset. You can decide to participate and change your mind later. You can answer only some of the questions if you'd like. You do not have to answer any questions you do not like.

We will keep your answers private. None of your friends or teachers will see your answers. Only the researchers doing the study will see your answers.

If you have questions about the study, please contact the study director, Alyssa Rulf Fountain of Abt Associates Inc. at (855) 355-1003 (toll-free) or email at alyssa_rulf_fountain@abtassoc.com. For questions about your rights as a participant in this study, contact Katie Speanburg at the Abt Associates Inc. Institutional Review Board at 877-520-6835 (toll-free).

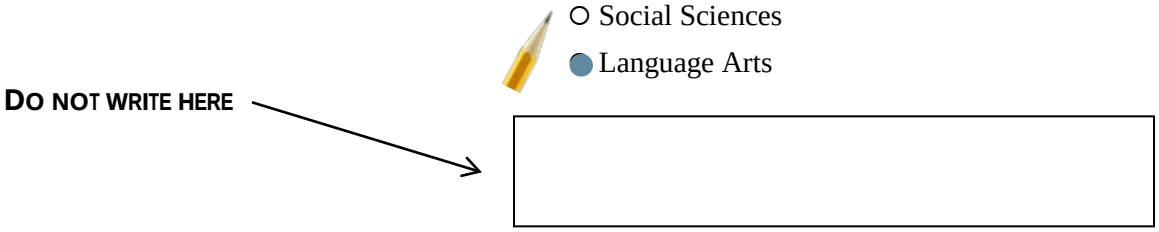
[2] Student surveys were administered electronically and by paper depending on school, but the instrument was identical across administration modes.

Instructions

All questions in this survey ask you to mark just one answer. These questions say, "Fill in one bubble". For example, in the following question you should fill in only one bubble:

What is your favorite subject? *(Mark one box.)*

- ○ Math
- ○ Science
- ○ Social Sciences
- ● Language Arts



APPRENTICESHIP SURVEY • 2015-16 Academic Year

We would like to learn more about you and your experience in your Citizen Schools apprenticeships.

By answering these questions, you are helping us learn about your apprenticeship experience.

First, we would like to know a little about who you are.

What is your date of birth?

(Month) (Day) (Year)

What is your grade?

6th	7th	8th
○	○	○

What is your gender?

Male	Female	Other
○	○	○

Please answer the following questions about the study of science. Please indicate how much you DISAGREE or AGREE with the following statements. Choose only one answer on each line.

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1. Science is useful in helping to solve the problems of everyday life.	○	○	○	○	○
2 Science is something that I enjoy very much.	○	○	○	○	○
3. I would like to do some extra or un-assigned reading in science.	○	○	○	○	○
4. Science is easy for me.	○	○	○	○	○
5. When I hear the word science, I have a feeling of dislike.	○	○	○	○	○
6. Most people should study some science.	○	○	○	○	○
7. Sometimes I read ahead in our science book.	○	○	○	○	○
8. Science is helpful in understanding today's world.	○	○	○	○	○
9. I usually understand what we are talking about in science.	○	○	○	○	○
10. Science teachers make science interesting.	○	○	○	○	○
11. No matter how hard I try, I cannot understand science.	○	○	○	○	○
12. I feel tense when someone talks to me about science.	○	○	○	○	○
13. Science teachers present material in a clear way.	○	○	○	○	○
14. I often think, "I cannot do this," when a science assignment seems hard.	○	○	○	○	○
15. Science is of great importance to a country's development.	○	○	○	○	○
16. It is important to know science in order to get a good job.	○	○	○	○	○
17. I like the challenge of science assignments.	○	○	○	○	○
18. It makes me nervous to even think about doing science.	○	○	○	○	○
19. It scares me to have to take a science class.	○	○	○	○	○
20. Science teachers are willing to give us individual help.	○	○	○	○	○

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
21. It is important to me to understand the work I do in science class.	○	○	○	○	○
22. I have a good feeling toward science.	○	○	○	○	○
23. Science is one of my favorite subjects.	○	○	○	○	○
24. I have a real desire to learn science.	○	○	○	○	○
25. I do not do very well in science.	○	○	○	○	○

On a scale of 1-7, please choose how well you do each of the following (1=Extremely Poor and 7=Excellent). Choose only one answer on each line.

	Extremely Poor (1)	2	3	4	5	6	Excellent (7)
1. Using digital tools to locate information to understand STEM (Science, Technology, Engineering, and Math) concepts. (digital tools: online books and articles, Internet searches, online discussions, gathering information from websites, or on-line databases)	○	○	○	○	○	○	○
2. Identifying necessary information to accomplish a STEM task.	○	○	○	○	○	○	○
3. Using technical writing to clearly communicate STEM topics.	○	○	○	○	○	○	○
4. Identifying “real world” challenges or problems in STEM areas.	○	○	○	○	○	○	○
5. Determining an innovative solution to a STEM challenge (e.g. Digital animation, oil spill, application of nanotechnology).	○	○	○	○	○	○	○
6. Evaluating the quality of an idea for a STEM project.	○	○	○	○	○	○	○
7. Evaluating the validity of data or evidence collected from a STEM project.	○	○	○	○	○	○	○
8. Applying STEM concepts to solve problems in other areas.	○	○	○	○	○	○	○
9. Respectfully working with individuals from other cultures.	○	○	○	○	○	○	○
10. Using reflective practices (e.g. blog journal, discussions with mentor, teacher, or peer) to foster a lifelong learning process.	○	○	○	○	○	○	○

	Extremely Poor (1)	2	3	4	5	6	Excellent (7)
11. Using technology in a responsible way (e.g. protecting digital identity, protecting the rights of others, following licensing laws).	○	○	○	○	○	○	○
12. Conducting a STEM project that has a value to society.	○	○	○	○	○	○	○

Are you interested in a career in science, technology, engineering, or math (STEM)? Select only one answer.

○ No, I am not interested

○ Maybe, I might be interested if I knew more about these types of jobs

○ Yes, I have some interest

○ Yes, I am very interested

How do you feel about your overall experience in Citizen Schools this year?

○ Very Unhappy

○ Unhappy

○ Neither happy nor unhappy

○ Happy

○ Very happy

Please answer whether you agree, disagree, or are unsure about the following statements:

	Agree	Disagree	Unsure
I plan to graduate from high school.	○	○	○
I plan to go to college.	○	○	○
I plan to graduate from college.	○	○	○

This is the end of the survey.

Thank you for taking the time to answer our questions!