The Baltimore City Schools Middle School STEM Summer Program with VEX Robotics

Martha Abele Mac Iver and Douglas J. Mac Iver

July 2015

B·E·R·C

Baltimore Education Research Consortium

The authors gratefully acknowledge the assistance and helpful comments from the following individuals (listed alphabetically): Jennifer Bell-Ellwanger, Amie Bettencourt, Crystal Brice, Faith Connolly, Tracey Cooper, Rachel Durham, Linda Eberhart, Jeffrey Grigg, Monica Hetrick, Laura Jones, Theresa Jones, Anne Lilly, Heather Nolan, Victoria Parrish, Michael Pih, Bruce Randel, Lawrence Reid, Ryan Reid Salta, Kim Robinson, Eric Stadius, Glenn Starnes, Marc Stein, Casey Thomas, and Chris West.

Table of Contents

List of Tables

List of Figures

The Baltimore City Schools Middle School STEM Summer Program with VEX Robotics

Executive Summary

In 2011 Baltimore City Schools submitted a successful proposal for an Investing in Innovations (i3) grant to offer a three year (2012-2014) summer program designed to expose rising sixth through eighth grade students to VEX robotics.[1] The i3-funded Middle School Science, Technology, Engineering and Mathematics (STEM) Summer Learning Program was part of a larger Baltimore City STEM summer learning program entitled "Create the Solution" in 2012 and "22nd Century Pioneers" in 2013 and 2014. The five-week summer program offered in 2012, 2013, and 2014 consisted of a half-day of instruction in mathematics and science and a half-day of enrichment activities. The robotics workshop taught students the fundamentals of building robots and provided time for teams to build their own robots and participate in competitions. The larger program offered different enrichment activities such as sports or arts.

This report addresses research questions regarding the program's 1) implementation fidelity, 2) performance goals, 3) impact on student attendance and mathematics achievement outcomes, 4) impact on student aspirations for college, studying STEM subjects in college, and pursuing STEM careers, and 5) impact on measures of teacher effectiveness. We summarize findings below for each of these questions.

Implementation Fidelity. Instruction in mathematics and robotics was implemented with fidelity all three program years. Implementation fidelity was lower for the professional development in robotics and mathematics components of the program because teacher attendance rates did not meet the thresholds set by City Schools.

Enrollment Goals. Most program enrollment goals were not met. Enrollment in the i3-funded program was 193 students in 2012 (goal 400), 384 in 2013 (goal 500), and 386 in 2014 (goal 600). The program sought to enroll 80% low-performing students in mathematics each year, but fell significantly short of this goal despite the district's efforts to reach out to these students. In addition, the program goal of enrolling at least 50% female participants was not met. The program also sought to have at least 80% of students attend at least 70% of the time (17 of the 24 program days), but only 55% of students attended at that rate. The program did meet its goals for recruiting minority (at least 95%) and high poverty students (at least 80%) each year.

Program Impacts on Attendance. We found a significant program effect on attendance in the year following the 2012 program. Program students had average attendance rates of 1.4 percentage points higher than the comparison group the year following the program (97.0% vs. 95.6%). An even larger significant program effect for low-achieving students' attendance was found in the year following the 2012 program (96.4% vs. 93.8%). The 2013 program students had slightly but not significantly higher attendance rates than their matched comparison students in the year following the program.

[1] VEX Robotics is an organization that provides equipment and organizes competitions for robotics teams. See www.vexrobotics.com.

We also examined whether there was still a program effect on attendance a year later (2013-14) for the Summer 2012 participants. Program participants had average attendance rates of 1.5 percentage points higher than comparison students (95.2% vs. 93.7%). Among the low-achieving students the attendance difference was 2.4 percentage points (93.6% for program students vs. 91.2% for comparison students). These effects were not statistically significant.

Program Impacts on Mathematics Achievement. There were no program effects on mathematics achievement for either the 2012 or 2013 programs.

Program Impacts on Student Aspirations. There was no evidence from student survey data that the robotics program had a positive effect on student aspirations to attend college, study STEM subjects in college, or pursue a STEM career for either the 2013 or 2014 programs.

Program Impacts on Teacher Effectiveness. Analyses based on mean instructional effectiveness scores from Spring 2013 and Fall 2013 on the nine components of the district's teacher evaluation tool examined whether teachers who received the summer professional development in 2013 made gains in instructional effectiveness. The difference between program teachers' effectiveness scores before and after the professional development was not statistically significant. Data were not available to examine differences between program teachers and a comparable group of teachers who did not receive the summer professional development.

Recommendations

Findings from this evaluation study suggest that it is important to consider the following issues in summer program planning:

Potential to Increase Student Motivation and Engagement -- The positive program effect on student attendance the following school year is an encouraging sign that robotics summer programs have the potential to keep students engaged who might otherwise begin disengaging from school. Although significant effects on student attendance were not found for the second year of the program, it is important for district leaders to discuss how both in-school instruction and out-of-school time can help to increase secondary student motivation and engagement.

Student Recruitment – The i3 summer program fell short each year in meeting its enrollment goals. We think it is important for district-level conversations to occur about possible action steps (including different types of programming) that could increase the likelihood that students needing additional instructional time to improve their achievement will actually receive that instruction.

Meeting Attendance Goals – Low attendance during summer programs is a commonly identified problem nationally. If summer programs are to achieve their intended goals of improving student achievement through additional instructional time, it is crucial to improve summer attendance rates. It is important for district leaders to discuss how attendance at summer programming can be incentivized and increased.

The Baltimore City Schools Middle School STEM Summer Program with VEX Robotics

Martha Abele Mac Iver and Douglas J. Mac Iver

Background

Preparing students for science careers requires that schools succeed in keeping students engaged and on-track academically throughout the middle grades and high school -- particularly in mathematics. And this involves the need for schools to address the central issue of student motivation, which often boils down to two main questions in students' minds about what happens in school: "Can I do the task?" and "Do I want to do the task?" (Eccles, 2008). Classroom practices in middle and high school often do not help motivate students to "want to do the task," failing to address students' needs to experience the task's intrinsic value, attainment value, and utility value for the future, as well as needs for both autonomy and social-connectedness (Eccles & Midgely, 1989). Recent discussions of noncognitive factors affecting academic performance have emphasized the importance of helping students develop an academic mindset that will influence academic behaviors such as attendance and exerting effort in class and homework assignments. The key components of an academic mindset are: "1) I belong in this academic community; 2) My ability and competence grow with my effort; 3) I can succeed at this; and 4) This work has value for me" (Farrington et al., 2012).

The process of helping students to internalize these beliefs can occur not only in the core academic classroom, but also in elective activities that build a sense of competence and value in academic pursuits. Middle grades students need "high engagement electives that provide avenues for short-term success … Experiences like … robotics and chess in which students with good engineering or logic abilities but limited formal mathematics skills can demonstrate strengths are essential" (Balfanz, 2009, p. 1).

A number of studies have investigated the impacts of robotics, but the research evidence remains rather thin. The emerging literature on robotics programming for secondary students alludes to increased student engagement, but there do not appear to be studies that measure the impact of such programming on student engagement directly. While many articles describe robotics interventions or provide anecdotal evidence of student engagement, the focus of most published systematic research thus far has been on either student learning or attitudes towards science and STEM careers (e.g., Benitti, 2012; Beer, Chiel & Drushel, 1999; Coxon, 2012; Grubbs, 2013; Mauch, 2001; Welch and Huffman, 2011). These studies report some positive effects on learning of science concepts and mathematics.

Evaluation studies of out-of-school time programs (both summer school and after school) have found mixed results on program impacts (Dynarski et al., 2004; Gottfredson et al., 2010; Lauer, et al., 2006; Lauver, 2002). One research team specifically advocates the need for summer program evaluations to expand beyond academic outcomes, arguing that "special attention be paid to measures of … attitudes toward school, self-image, and attendance and discipline problems during the following school year" (Cooper et al., 2000, p. 102). This study seeks to contribute to this research agenda by assessing the effects of an out-of-school time robotics intervention aimed at increasing student interest in STEM and engagement with school.

Background on the Intervention

In 2011 Baltimore City Schools submitted a successful proposal for an Investing in Innovations (i3) grant to offer a three-year (2012-2014) summer program to expose rising sixth through eighth grade students to VEX robotics.[2] The i3-funded Middle School Science, Technology, Engineering and Mathematics (STEM) Summer Learning Program was part of a larger Baltimore City STEM summer learning program entitled "Create the Solution" in 2012 and "22nd Century Pioneers" in 2013 and 2014. Similar to STEM summer learning programs offered by Baltimore City Schools in 2010 and 2011 (prior to i3), the five-week summer program offered in 2012, 2013, and 2014 consisted of a half-day of instruction in mathematics and science and a half-day of enrichment activities. The robotics workshop taught students the fundamentals of building robots and provided time for teams to experiment, build their own robots, and later participate in competitions. The larger program offered different enrichment activities such as sports or arts.

Recruitment. Students scoring Basic on the prior year's Maryland School Assessment (MSA) mathematics test were specifically targeted for the i3 program (current year scores were not available during the recruitment period). Invitations were mailed to all qualifying students. In addition, information packets were sent to all schools serving students in the targeted grades. Information about the summer program was also available on the school district's website. Families were able to register their students on-line as well as using paper registration forms submitted to schools or the district office. Students were also allowed to enroll in the program once it had started (without pre-registering). The district used robot demonstrations in later program years in its recruitment efforts.

Program locations. The original proposal was for the summer program to be housed on three college campus sites, allowing students to experience life on a college campus and learn about admissions and financial aid. Although it was not possible for this to occur, Coppin State University was one of the program sites in 2013 and 2014, and all other sites were at City Schools. The locations and number of sites varied from year to year. The number of sites was 10 for 2012, 11 for 2013, and 9 for 2014. (See Appendix for a list of program sites each year.)

Program teachers. Baltimore City Schools teachers were recruited each year for the program through the regular district process of recruiting summer school teachers. Only teachers with at least a "Satisfactory" rating for the prior year were eligible to be hired. Selected teachers were assigned to teach either mathematics or science in the half-day academic component of the program. A subset of willing volunteers from among these teachers was selected to also teach the robotics component of the program. Teachers attended professional development sessions during the week prior to the start of the summer school program. Provision of professional development addressed one of the i3 program goals of helping improve teacher instruction and effectiveness during the school year. After the introductory session, teachers attended sessions specific to either mathematics or science[3] instruction. In addition, robotics teachers attended professional development sessions focused on robotics and each constructed their own robot

[2] VEX Robotics is an organization that provides equipment and organizes competitions for robotics teams. See www.vexrobotics.com.
[3] The summer program included science instruction, but this portion of the program was not funded by i3 and not part of the evaluation study.

during the week prior to professional development in math/science. Group and individual planning time (partially at the school sites) was included as part of the professional development week. In Years 1 and 2 the mathematics professional development was provided by City Schools staff and focused on the district Math Works curriculum. In Year 3, external consultants provided mathematics instruction professional development focused on implementing the Common Core curriculum from Engage NY (www.engageny.org).

Program evaluation framework. All i3-funded projects require an independent, third party evaluation. City Schools designated the Baltimore Education Research Consortium (BERC) as its independent evaluator. After awarding the i3 grants, the U.S. Department of Education designated Abt Associates to conduct a National Evaluation of i3 (NEi3), which includes assessing the quality of each of the independent evaluations of i3 grants. In late summer 2012 Abt Associates assigned a consultant to each independent evaluator to guide the evaluation planning and execution. The consultant worked with BERC throughout the entire process of the three-year evaluation. This involved submission of a revised final logic model and implementation fidelity measures, as well as a comprehensive design plan prior to beginning analyses, and a discussion of each set of analyses throughout the evaluation.

Logic Model

The logic model for this intervention (see Appendix A), developed collaboratively by the Baltimore City Schools development team and the BERC evaluation team, provides an overarching summary of how components of the Baltimore City Middle School STEM Summer Learning Program were designed to influence outcomes for high-needs middle school students. As illustrated in the logic model, the four key components of implementation fidelity for this intervention were professional development in mathematics, professional development in robotics, instruction in mathematics, and instruction in robotics. Professional development provided to teachers prior to the program start and throughout the program was expected to lead to high quality instruction in mathematics and robotics for summer school students. In particular, the professional development in Years 1 and 2 focused on equipping teachers to help students in the development of fact fluency and automaticity, while in Year 3 it focused more on Common Core standards emphasizing mathematical thinking and problem solving. In addition, the professional development emphasized the use of formative assessment data to inform instruction. Professional development in robotics was expected to lead to teacher delivery of a high quality robotics program with highly engaging competitions. City Schools program developers expected the robotics component to engage student interest and sustain summer program attendance. Summer program attendance and exposure to additional high quality mathematics instruction during the summer were expected to improve student mathematics achievement. In addition, participation in the robotics enrichment was expected to increase student engagement in general (measured by attendance the following year) as well as student interest in STEM careers (including college attendance and college coursework in mathematics and science required for those careers).

Methodology

This section will summarize the research questions, and describe the data and procedures used to respond to each question.

Research Questions

1) To what extent were key program components (professional development in mathematics, professional development in robotics, instruction in mathematics, and instruction in robotics) implemented with fidelity each year of the program?

2) To what extent were proposed performance goals (in student recruitment, program attendance, and teacher recruitment) met each year?

3) (Years 1 and 2) To what extent did the program have a positive impact on the following year's attendance rate of middle school students (compared to students who did not receive any of the district's STEM-related summer programs)?

4) (Years 1 and 2) To what extent did the program have a positive impact on the following year's math achievement of middle school students (compared to students who did not receive any of the district's STEM-related summer programs)?

5) (Years 2 and 3) Compared to students who received similar summer mathematics and science instruction but not robotics instruction during one of the district's concurrent Summer Learning programs with Arts or Sports/Fitness Enrichment, to what extent did participation in VEX Robotics STEM summer learning program have a positive impact on middle school students' aspirations for
   - Attending college?
   - Studying math and science in college?
   - Pursuing a STEM-related career?

6) (Year 2) To what extent did participation in the program's professional development have a positive impact on math teachers' teaching effectiveness ratings during the academic year following their program participation?

Data Sources

Implementation Fidelity

A process of collaborative discussion between BERC and City Schools staff led to the specification of indicators for each of the program's key components, as well as thresholds for assessing whether or not the program had been implemented as intended by the developers. These changed somewhat from year to year because of changes in the program over time. Data collection to measure whether all sites were implementing robotics and mathematics instruction as intended included the following components:

- Discussions (through email, by phone, and face-to-face) with program staff
- Observations at all summer program sites, which included conversations with site directors, observation of all math and robotics classrooms and all the robotics classrooms at all sites, and observations of robotics competitions. Each of the sites was visited for about two to three hours on one day.
- Document gathering (such as site schedules, etc.).

City Schools supplied data to BERC on teacher attendance at professional development sessions and teacher program attendance, which were used to measure whether the thresholds set by City Schools for adequate implementation fidelity of the professional development components were met.

For more details on implementation fidelity analyses, see Appendix A.

Performance Goals

Data on program students and their attendance were received from City Schools each program year. Data were matched to district administrative data to summarize program student characteristics. Data on program teachers' certifications and subject teaching experience were received from City Schools to conduct analyses addressing performance goal research questions. For more details, see Appendix B.

Attendance and Achievement Impact Analyses

District administrative data were used to construct matched comparison groups for program students in 2012 and 2013 through a combination of Mahalanobis and propensity score matching. Program impact analyses on attendance and mathematics achievement were conducted using hierarchical linear modeling. In Year 1, measures of mathematics achievement included both the Fall 2012 district benchmark tests in mathematics and the Spring 2013 MSA scores. In Year 2, only the Spring 2014 MSA scores were available. For more details on analyses and the transition in state assessments over the course of the study, see Appendix C.

Student Aspirations for Studying STEM in College and for STEM Careers

A short survey was administered to students at the beginning and end of the program in 2013 and 2014 to measure change in aspirations to attend college, study STEM subjects in college, and pursue a STEM career. It was impossible for logistical reasons to administer a survey to the comparison group of students identified by propensity score matching. Robotics students were therefore compared to program students who received the same math and science instruction but enrichment activities in sports and arts rather than robotics.

The survey contained several items to measure each of these aspirations, and the final aspirations measure used in analyses was the mean of the individual items (see Appendix C for more information on the survey items). Parental consent was required for survey data to be entered and analyzed. Student survey data were linked to student demographic and academic characteristics in the district administrative data. Analyses were conducted using hierarchical linear modeling to examine the impact of the robotics program (compared to other summer program enrichment activities) on student aspirations. For more details on the analyses, see Appendix C.

Teacher Effectiveness

City Schools provided de-identified data for the 2013 program math teachers, including "instructional performance" scores for each of the nine instructional components on the district teacher evaluation tool. Data generally included two observation scores (fall and spring) for each teacher in 2013 and 2014. We used a one group pretest-posttest design.   The comparison condition for the study was the teacher's effectiveness score in the year prior to the professional development intervention. There was no available comparison group of teachers not receiving the intervention for this study (because data were not available to the research team to construct a matched comparison group). Thus, the study cannot rule out other causes (e.g., general maturation experienced by all teachers) for any growth in instructional performance detected. For more details on the analyses, see Appendix C.

Findings

This section presents findings organized by each of the research questions specified above. Technical details about the analyses are included in the Appendix C.

Implementation Fidelity

1) To what extent were key program components (professional development in mathematics, professional development in robotics, instruction in mathematics, and instruction in robotics) implemented with fidelity each year of the program?

In Appendix A we present a matrix to summarize the implementation fidelity measures more fully, with data sources, values, and thresholds for each of the components and how these changed over the course of the study. In accordance with the implementation fidelity study guidelines for the National Evaluation of i3, each of the key components of the intervention is evaluated separately instead of specifying an overall measure of implementation fidelity for the intervention. Table 1 summarizes the study findings.

Table 1
Fidelity of Program Implementation, by Component and Year

	Year 1 (2012)	Year 2 (2013)	Year 3 (2014)
Mathematics Instruction	Implemented with fidelity	Implemented with fidelity	Implemented with fidelity
Robotics Instruction	Implemented with fidelity	Implemented with fidelity	Implemented with fidelity
Mathematics Professional Development	Not Implemented with fidelity	Not Implemented with fidelity	Missing Data[4]
Robotics Professional Development	Implemented with fidelity	Not Implemented with fidelity	Not Implemented with fidelity

Program Performance Goals

2) To what extent were proposed performance goals (in student recruitment, program attendance, and teacher recruitment) met each year?[5]

[4] Data on teacher attendance at professional development were available for 2 of the 3 scheduled days.
[5] A copy of the program performance goals from the i3 proposal is included in Appendix B.

The goal for student recruitment to the i3-funded program was 400 students in 2012, 500 in 2013, and 600 in 2014. As Table 2 indicates, these goals were not met. Many of the full summer program participants chose sports or arts rather than the robotics program.

The goals for recruiting minority (at least 95%) and high poverty students (at least 80%) were met each year. The program sought to enroll 80% low-performing students in mathematics each year, but fell significantly short of this goal despite the district's efforts to reach out to these students. In addition, the program had set a goal of enrolling at least 50% female participants. This goal was not attained (see Table 2 below).

Table 2
Middle School Summer STEM Program Performance Goal Status[6]

	Program Goals	Year 1 (2012)	Year 2 (2013)	Year 3 (2014)
Number of students recruited	Year 1 – 400 Year 2 – 500 Year 3 – 600	193[7]	384	386
% Female	50%	26.2%	40.9%	32.9%
% Free/Reduced Lunch	80%	**85.6%**	**87.8%**	**88.3%**
% Minority	95%	**95.4%**	**94.8%**	**94.6%**
% Below Proficient on prior year's Math MSA	80%	37.4%	43.8%	47.3%
% of students attending at least 70%	80%	54.9%	58.1%	54.9%
% teachers who taught upper elementary grades the year prior	15% to 20%	From Gr 3-5: 12.2% From Gr K-5: 18.3%[8]	From Gr 3-5: 12.8% From Gr K-5: 22.0%	**From Gr 3-5: 20.3%** From Gr K-5: 24.4%

[6] Boldface indicates that program goals were met.
[7] See Appendix B for more details about how program students were identified. In 2012 there were approximately 50 more Robotics students at two additional sites which did not provide data required to identify students in Robotics program.

[8] Percentage of teachers from all elementary grades added for context.

One of the goals of the program in the original proposal was to "recruit 15-20% teachers from upper elementary grades, to encourage STEM learning in elementary grades." This goal was met in Year 3, and nearly met in the two previous years.

The program also sought to have at least 80% of students attend at least 70% of the time (17 of the 24 program days). This goal was also not achieved (see Table 2).

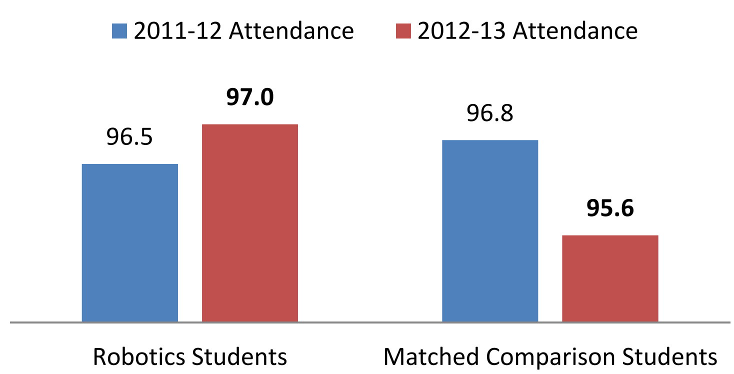
Program Impacts on Attendance

3) (Years 1 and 2) To what extent did the program have a positive impact on the following year's attendance rate of middle school students (compared to students who did not receive any of the district's STEM-related summer programs)?

For the Year 1 (2012) program, a total of 166 robotics students were promoted and had no missing data for analyses. Using propensity score matching, we identified a total of 486 comparison students who had not attended summer school in 2012. Analyses verified that these two groups of students had virtually identical average attendance in 2011-12, prior to the program, and were closely matched on other demographic and school characteristics.

A significant program effect was found for attendance in the year following the program. Program students had average adjusted[9] attendance rates of 1.4 percentage points higher than the comparison group the year following the program (97.0% vs. 95.6%) (Figure 1). Another way of stating the impact is that treatment students attended about 2.5 days more of the 180-day school year on average.

Figure 1. Adjusted Average Attendance Rates During the School Years Before and After 2012 Summer Program, for Robotics Students and Matched Comparison Group[10]

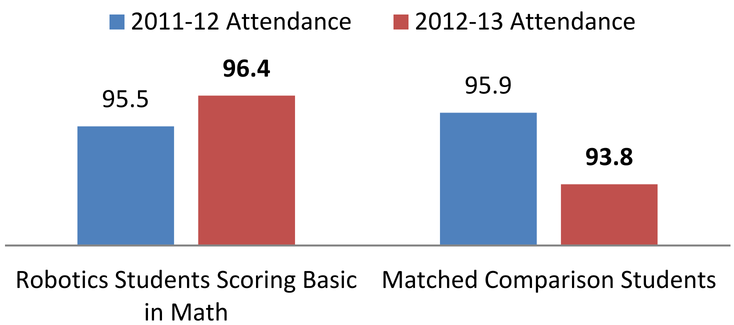


Parallel analyses found an even larger significant program effect for low-achieving (scoring Basic on the mathematics MSA) students' attendance in the year following the program. The

[9] For technical details on analyses, see Appendix C.
[10] Significant difference between robotics and comparison students in 2012-13

Basic program students had average attendance rates of 2.6 percentage points higher than the comparison group the year following the program (96.4% vs. 93.8%) (see Figure 2). Another way of stating the impact is that treatment students in the subsample attended, on average, about a week more of school than did the control students in the subsample (i.e., attended 4.7 days more during the course of the 180-day school year.)

Figure 2. Adjusted Average Attendance Rates During the School Years Before and After 2012 Summer Program, for Low-Performing Robotics Students and Matched Comparison Group[11]



The same analyses were conducted for the Year 2 (2013) program students. A total of 358 robotics students were promoted and had no missing data for analyses. These were matched to a total of 1057 comparison students who had not attended summer school in 2013. Analyses verified that these two groups of students had virtually identical average attendance in 2012-13, prior to the program.

The 2013 program students had adjusted average attendance rates of 0.6 percentage points higher than the comparison group the year following the program (95.0% vs. 94.4%), which was not a statistically significant difference. Parallel analyses for low-achieving students in mathematics (157 program students closely matched to 453 comparison students) found that those program students had average attendance rates of 1.5 percentage points higher than comparison students (94.9% vs. 93.4%), but this did not reach the threshold for statistical significance.

Finally, we examined whether there was still a program effect on attendance a year later (2013-14) for the 2012 robotics program participants. About a quarter of the 2012 participants had progressed to ninth grade, a year when attendance typically declines notably. A total of 157 program students and 462 comparison students had 2013-14 attendance data, and remained closely matched on 2011-12 attendance. Identical analyses were conducted using 2013-14 attendance as the dependent variable. Program students had average attendance rates of 1.5 percentage points higher than comparison students (95.2% vs. 93.7%). Among the low-achieving students (56 program students closely matched to 159 comparison students), the attendance difference was 2.4 percentage points (93.6% for program students vs. 91.2% for comparison students). Although program students had higher attendance, neither effect was statistically significant.

[11] Significant difference between robotics and comparison students in 2012-13

Program Impacts on Mathematics Achievement[12]

4) (Years 1 and 2) To what extent did the program have a positive impact on the following year's math achievement of middle school students (compared to students who did not receive any of the district's STEM-related summer programs)?

In Year 1 it was possible to examine the impact of the summer program on a proximal measure of mathematics achievement, the Fall 2012 district mathematics benchmark test, as well as on the Spring 2013 MSA test. In Year 2 only the Spring 2014 MSA test scores were available (for those students who did not take the PARCC assessment in 2014). Analyses were conducted using the same methodology as for the attendance analyses described above. Program students did not have significantly higher mathematics achievement scores than comparison students, for either the full sample or the subgroup of low-achieving students, on any of the mathematics achievement measures.[13]

Program Impacts on Student Aspirations

5) (Years 2 and 3) Compared to students who received similar summer mathematics and science instruction but not robotics instruction during one of the district's concurrent Summer Learning programs with Arts or Sports/Fitness Enrichment, to what extent did participation in the VEX Robotics STEM summer learning program have a positive impact on middle school students' aspirations for
    - Attending college?
    - Studying math and science in college?
    - Pursuing a STEM-related career?

No significant effects of the robotics program (compared to the other enrichment programming) were found on any of the aspirations measures (college-going, taking science and mathematics courses in college, or pursuing a STEM career) in either year. The sample size each year was too small to detect any small program effects. For tables summarizing details of analyses, see Appendix C.

[12] For technical details on analyses, see Appendix C.

[13] It is important to note that the district changed its math curriculum during the period of the evaluation. During the 2011-12 school year, the math curriculum was aligned to MSA standards. For 2012-13, a hybrid curriculum of MSA and Common Core Standards was used by the district. In 2013-14, the mathematics curriculum was completely aligned to Common Core Standards. Despite these changes in curriculum, students were tested all three years using the MSA.

Program Effects on Teacher Effectiveness

6) (Year 2) To what extent did participation in the program professional development have a positive impact on math teachers' teaching effectiveness ratings during the academic year following their program participation?

Instructional effectiveness scores on the nine components of the district's teacher evaluation tool were available for 2012-13 and 2013-14 for 52 of the 58 program teachers from 2013. Means of the nine component scores (each ranging from 1 to 4) were computed for an overall observation score. Because two observations were conducted each year, the best estimation of the impact of the summer professional development intervention is a comparison of the second mean score from the year prior to the intervention (from an observation generally conducted in the spring) with the first mean score from the year following the intervention (from an observation generally conducted in the fall) -- the observations closest in proximity to the professional development intervention. A total of 44 teachers had full data available for this analysis. The difference between the Fall 2013 mean score and the Spring 2013 mean score was not statistically significant.[14]

Data were not available to ascertain whether scores also rose at the same rate for a comparable group of teachers who did not receive the intervention. It is therefore not possible to confidently attribute the improved scores to the intervention itself.

[14] See Appendix C for technical details of analyses.

Discussion

Although the impact on student achievement in mathematics that program developers had hoped to achieve was not detected, the 2012 program's impact on student attendance rates the following year is an important finding. Compared to students who had the very same attendance rate prior to the intervention, students who attended the 2012 robotics summer program had significantly higher attendance rates at school in the year after the program. Attendance rates were also higher for 2013 program students (particularly the low-performing students) than matched comparisons, though the effect did not meet the threshold of statistical significance in the second year.

It is important to acknowledge the limitations of this study, compared to studies using random assignment of students. It is possible that even though matched comparison groups students were equivalent to program students on prior school attendance rates, as well as on all the other matching variables, they differed in some unmeasured way that would explain their higher school attendance in the year following the program. For example, it is possible that there was some underlying motivation difference or family support difference between program attenders and non-attenders, related to recent experience that did not affect prior year's attendance, which could account for their attendance difference in the subsequent year.

At the same time, the study had a rigorous design in a context where random assignment of students to a summer program was not possible. And the study did find some evidence that one important goal of the STEM Robotics Summer Learning Program – to produce an impact on students' engagement that would last beyond the summer by providing relevant, authentic, minds-on, hands-on instruction in technology, mathematics, and science– was realized. The program provided students with the opportunity to develop an understanding of the relevance (and exciting but practical uses) of STEM to the field of robotics. This opportunity to construct, program, and operate a robot to carry out specific tasks and test the robot in competitions seems to have produced the envisioned increases in students' commitment to attend middle school regularly. Although the study did not include qualitative interview data with treatment students that could have illumined the specific mechanisms underlying this attendance effect, the informal observations of program classrooms conducted by the research team found high levels of student engagement in the process of building robots, and a sense of accomplishment in student demeanors as they operated robots and participated in both informal and formal competitions. These observations were consistent with potential growth or maintenance of the components of an academic mindset related to academic behaviors such as attendance: feelings of competence and success, belief in the value of effort to increase competence, and perceptions of the value of the task. Future research may be able to measure the specific intervening attitudes and determine their relationship to the behavioral outcome variable of school attendance.

It is important to note that the 2012 program had a significantly positive effect for males as well as for females. Our informal classroom observations found boys deeply engaged in building robots (who were generally not as engaged in the regular academic classroom environment). The students served by the program were nearly all African-American, and more than two-thirds of the program students in 2012 were African American males. Given the widely discussed educational challenges of African-American males in American society (e.g. Lewis et al., 2010;

Holzman, 2010), this study's finding of a program effect on their school engagement is an important contribution. While research has documented the lower academic achievement of males in American society that has contributed to their gender gap in college attendance and linkages to a gender gap in "noncognitive" academic behaviors have been made (e.g., DiPrete & Buchmann 2013; Jacob, 2002), there has been little attention to the gender gap in school attendance rates (except in discussions of the higher rates of suspension among males, and particularly minority males). Keeping boys interested in academic pursuits remains a challenge, particularly in high poverty inner city contexts. When activities such as robotics engage them actively in building something complex that they can then manipulate and enjoy, they can see tangible results of their efforts, take pride in the competence they have demonstrated, and gain a vision for how what happens in school can be relevant for them both in the present and in the future.

"Motivating the academically unmotivated" (Hidi and Harackiewicz, 2000, p. 151) is one of the critical issues of the 21st century. Finding ways to stir up student interest in pursuing learning activities to maintain even the crudest indicator of engagement, simple school attendance, remains a challenge for most high-poverty secondary schools. There is a need for more focused theoretical attention to the role of "interest" in student motivation and engagement (Ainsley, 2010). In particular, how can schools and other organizations focused on youth development both trigger and maintain situational interest so that students begin to internalize interests leading to increased motivation and engagement? Researchers argue that "active participation, engagement, and effort are promoted by tasks that are hands-on, heads-on, project-based, relevant, progressive, and integrated across subject matter, or in other words, intrinsically motivating, inherently interesting, and fun" (Skinner and Pitzer, 2012, pp. 28-29). The results of this study suggest that continued investment in high-interest elective activities such as robotics could have a significant impact on helping students remain engaged in school, who otherwise may have begun a process of disengagement that would lower their chances for successful eventual transition in to college and career.

Even if there was no detectable program impact on achievement in mathematics, based on the state assessment results, the impact on attendance, as a basic behavioral measure of engagement, is an encouraging finding. Prior research has noted that attendance (together with behavior and course grades) is much more important than test scores as a predictor of high school graduation (e.g., Balfanz, Herzog, & Mac Iver, 2007; Allensworth & Easton, 2007). While test scores may become more important as states continue to institute graduation testing requirements, and certainly predict the need for remediation as students make the transition to college, it is important not to ignore the impact of improving attendance on students' college and career readiness. Attending class in college and showing up to work every day are critical determinants of college and career success. Increasing attendance in middle and high school is the first step to getting students ready for college and career.

Recommendations

Findings from this evaluation study suggest that it is important to consider the following issues in summer program planning:

Potential to Increase Student Motivation and Engagement -- The positive program effect on student attendance the following school year is an encouraging sign that robotics summer programs have the potential to keep students engaged who might otherwise begin disengaging from school. Although significant effects on student attendance were not found for the second year of the program, it is important for district leaders to discuss how both in-school instruction and out-of-school time can help to increase secondary student motivation and engagement.

Student Recruitment -- As noted above, the i3 summer program fell short each year in meeting its goals for number of students served in the program. We think it is important for district-level conversations to occur about possible action steps that could increase the likelihood that students needing additional instructional time to improve their achievement will actually receive that instruction.

Meeting Attendance Goals – Low attendance during summer programs is a commonly identified problem. If summer programs are to achieve their intended goals of improving student achievement through additional instruction time, it is crucial to improve summer attendance rates. It is important for district leaders to discuss how attendance at summer programming can be incentivized and increased.

References

Ainley, M. (2012). Students' interest and engagement in classroom activities. In S. Christenson, A. Reschly, & C. Wylie (Eds). *Handbook of research on student engagement* (pp. 283-302). New York: Springer.

Allensworth, E., & Easton, J. (2007). *What matters for staying on-track and graduating in Chicago public high schools.* Chicago: Consortium on Chicago School Research.

Balfanz, R. (2009). *Putting middle grades students on the graduation path*. Baltimore, MD: Everyone Graduates Center.

Balfanz, R., Herzog, L. & Mac Iver, D.J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist, 42*(4), 223-235.

Beer, R., Chiel, H. & Drushel, R. (1999). Using autonomous robotics to teach science and engineering. *Communications of the ACM*, 42, 85-92.

Benitti, F. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education, 58,* 978-988.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, *65*(Serial No. 260), 1–118.

Coxon, S. (2012). The malleability of spatial ability under FIRST LEGO league-based robotics simulation. *Journal for the Education of the Gifted*, 35, 291-316.

Deci, E. L. (1998). The relation of interest to motivation and human needs: The self-determination theory viewpoint. In L. Hoffman, A. Krapp, K.A. Renninger, & J. Baumert (Eds.), *Interest and learning* (pp. 146-162). Kiel, Germany: Institute for Science Education.

DiPrete, T., & Buchmann, C. (2013) *The rise of women: The growing gender gap in education and what it means for American schools.* New York, NY: Russell Sage Foundation.

Dong, N. and Maynard, R. A. (2013). PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Sample Size Requirements for Experimental and Quasi-experimental Designs. *Journal of Research on Educational Effectiveness*, 6, 24-67.

Duckart, J. (1998). An evaluation of the Baltimore Community Lead Education and Community Corps (CLEARCorps) program. *Evaluation Review*, 22, 373-402.

Dynarski, M., James-Burdumy, S., Moore, M., Rosenberg, L., Deke, J., & Mansfield, W. (2004). *When schools stay open late: The national evaluation of the 21st Century Community*

*Learning Centers Program: New findings (*U.*S. Department of Education, National Center for Education Evaluation and Regional Assistance)*.Washington, DC: U.S. Government Printing Office.

Eccles, J. S. (2008). Can middle school reform increase high school graduation rates? *California Dropout Research Policy Brief, 12*.

Eccles, J. S., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for young adolescents. *Research on motivation in education, 3*, 139-186.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D.W., & Beechum, N. O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago, IL: University of Chicago Consortium on Chicago School Research.

Gottfredson, D. Cross, A.B., Wilson, D., Rorie, M. & Connell, N. (2010). Effects of participation in after-school programs for middle school students: A randomized trial *Journal of Research on Educational Effectiveness*, 3, 282–313.

Grubbs, M. (2013, March). Robotics intrigue middle school students and build STEM skills. *Technology and Engineering Teacher*, 12-16.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, 70 (2), 151-179.

Holzman, M. (2010). *Yes we can: The Schott 50 state report on public education and black males*. Schott Foundation for Public Education.

Jacob, B.A. (2002). Where the boys aren't: Noncognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review, 21*, 589-598.

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

Lauver, S. C. (2002).Assessing the benefits of an after -school program for urban youth: An impact and process evaluation. *Dissertations available from ProQuest.* Paper AAI3043903. http://repository.upenn.edu/dissertations/AAI3043903

Lewis, S., Simon C., Uzzell R., Horwitz, A., & Casserly, M. (2010). *A call for change: The social and educational factors contributing to the outcomes of black males in urban schools*. Washington, D.C.: Council of Great City Schools.

Mauch, E. (2001). Using technological innovation to improve the problem solving skills of middle school students: Educators' experiences with the LEGO Mindstorms robotic invention system. *The Clearing House*, 74, 211-213.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, *95*(450), 573-585.

Skinner, E. & Pitzer, J. (2012). Developmental dynamics of students' engagement, coping, and everyday resilience.. In S. Christenson, A. Reschly, & C. Wylie (Eds). *Handbook of research on student engagement* (pp. 21-44). New York: Springer.

Appendices

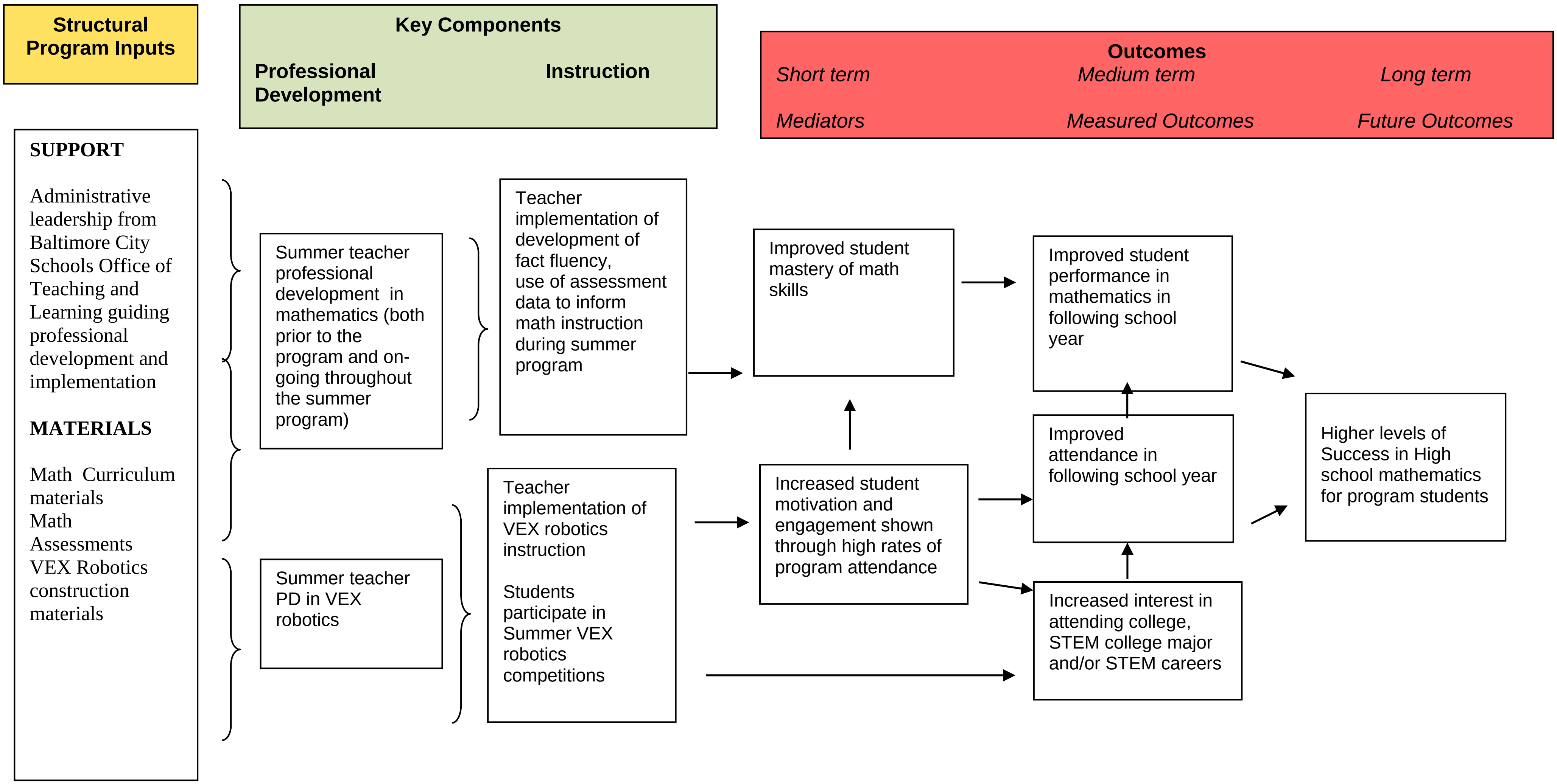
Appendix A
Implementation Fidelity

To meet the standards for the National Evaluation of i3 (NEi3), the evaluation of the STEM Middle School Summer Program required the following implementation study components: 1) *a logic model for all tested interventions that identifies key components of the intervention, the mediators through which the intervention is expected to have its intended outcomes, and the student outcome domains the intervention is designed to improve; 2) measures of implementation fidelity that cover all key components of the intervention, and descriptions of how they will be constructed; 3) specified thresholds for acceptable implementation; and 4) reported summaries of implementation fidelity for the entire study sample.* The focus on the implementation study was on whether key components of the intervention were delivered as intended. It did not extend into measures of "quality" of implementation (e.g., quality of classroom instruction, etc.).

The logic model for the intervention and all the required implementation fidelity measures and thresholds were developed in a collaborative process between the evaluator (BERC) and City Schools over the course of the 2012-13 school year. Due to the scheduling of the NEi3 process, these collaborative discussions occurred after the first summer of implementation, and so measurement of implementation fidelity in Year 1 (Summer 2012) differed somewhat from the plans developed for Years 2 and 3 (2013 and 2014).

The collaborative discussions focused on the original proposal's logic model and intervention design, as well as ways it had been modified over the course of the project period. Figure 1 presents the final logic model that was submitted for the NEi3 and which guided the development of the implementation fidelity measures.

Figure A1.   Baltimore Middle School STEM Summer Learning Program Logic Model



Implementation Fidelity Measures

As illustrated in the logic model, the four key components of implementation fidelity for this summer program intervention are professional development in mathematics, professional development in robotics, instruction in mathematics, and instruction in robotics.

Professional Development in Mathematics

To measure implementation fidelity for this program component, two indicators were required: measurement of 1) teacher attendance at pre-program professional development, and 2) attendance at the ongoing, daily professional development time specified in the proposal.

Program developers from City Schools decided that teacher attendance for the full time of the pre-program professional development was required for adequate implementation fidelity. To measure implementation fidelity for this indicator across the entire program, City Schools specified a threshold of 90% of teachers attending all pre-program professional development.

Because the original proposal also specified daily professional development time during the program (daily collaboration with colleagues in addition to the weekly "formal" professional development session), we specified the second indicator of teacher participation in the on-going professional development to be measured by teacher program attendance. City Schools program developers defined adequate implementation of this PD component as a teacher attending at least 22 of the 24 days of the program. To measure implementation fidelity for this indicator across the entire program, City Schools specified a threshold of 90% of teachers attending at least 22 of the 24 days of the program.

To measure overall implementation fidelity for this program component, developers required that the threshold be met on both indicators for the component to be implemented with fidelity. Thus, program implementation with fidelity requires at least 90% of teachers attend all pre-program PD and at least 90% of teachers attend 22 days or more of the program.

Data on teacher professional development attendance were delivered to BERC from City Schools. In Year 1 (2012), data were available only for the first indicator, teacher attendance at the pre-program professional development. In Year 1, 68% of math teachers attended all three days of the pre-program sessions for which attendance were taken. This did not meet the threshold specified by City Schools for the program component to be implemented with fidelity.

In Year 2 (2013), 70% of math teachers attended all days of the pre-program professional development, and 78% of teachers attended at least 22 of 24 program days. Neither of the indicators met the threshold specified by City Schools for the program component to be implemented with fidelity.

In Year 3 (2014), data on teacher attendance at the third day the pre-program professional development was missing. A total of 91% of math teachers attended the first two days of pre-program professional development and 97% of teachers attended at least 22 of 24 program days. Although the second indicator met the threshold specified by City Schools for the program component to be implemented with fidelity, the fact that data were missing on pre-program attendance leads to the conclusion that it is not known whether this component was implemented with fidelity in Year 3.

Professional Development in Robotics

To measure implementation fidelity for this program component, two indicators were specified: measurement of 1) teacher attendance at pre-program professional development, and 2) teacher completion of a robot during the pre-program sessions. Because these were not measured in Year 1 (which occurred prior to the specification of implementation fidelity measures), the Year 1 indicator of this component was district delivery of the week-long professional development in robotics.

Program developers from City Schools decided that teacher attendance for the full time of the pre-program professional development was required for adequate implementation fidelity. To measure implementation fidelity for this indicator across the entire program, City Schools specified a threshold of 90% of teachers attending all pre-program professional development. Completion of the robot during the pre-program session was also specified as necessary for adequate implementation fidelity. To measure implementation fidelity for this indicator across the entire program, City Schools also specified a threshold of 90% of teachers completing a robot during pre-program professional development.

In Year 1 (2012), data were not available on teacher attendance at professional development. Professional development was delivered, and this met the threshold for implementation fidelity on this program component for Year 1.

In Year 2 (2013), 76% of robotics teachers attended all days of the pre-program professional development, and 76% of teachers completed a robot. This did not meet the threshold specified by City Schools for the program component to be implemented with fidelity.

In Year 3 (2014), 69% of robotics teachers attended all days of the pre-program professional development. Data were missing on the percentage of teachers who completed a robot. Because both indicators were required to meet the threshold specified by City Schools for the program component to be implemented with fidelity, and fewer than 90% of teachers attended all professional development days, this program component was not implemented with fidelity in Year 3.

Mathematics Instruction

To measure implementation fidelity for this program component, three indicators were specified. In Years 1 and 2 these were measurement of 1) time devoted to math instruction overall; 2) time devoted to fact practice; and 3) whether weekly assessments were conducted. Because of changes in the mathematics curriculum in Year 3 to align more closely with the Common Core Standards, these were changed in Year 3 to: 1) time devoted to math instruction overall; 2) implementation of the provided Common Core-aligned mathematics curriculum; and 3) whether pre- and post-assessments were conducted.

Classroom observations were conducted to gather data on implementation fidelity. Instead of a formal protocol, we used a running record methodology. Classroom visits were at least 15- 20 minutes long, and sometimes classrooms were visited more than once during the 80 minute mathematics period.

For each of these indicators, the site was coded as implementing with fidelity if all math classrooms were coded adequate from data gathered from program documents and site visits. At the program level, each indicator was coded as implemented with fidelity if at least 90% of the sites were coded adequate. At the program level, the “Mathematics Instruction” component was judged to be implemented with fidelity if at least 90% of sites were judged adequate on all three indicators.

In each of the three years, 100% of sites met the threshold for implementation fidelity on all three of these indicators.

Robotics Instruction

To measure implementation fidelity for this program component, two indicators were specified: measurement of 1) time devoted to robotics instruction overall; and 2) whether sites took students to the program-wide robotics competitions.

Program developers from City Schools decided that for robotics instruction to be implemented with fidelity, each site needed:

- daily robotics instruction programming for at least 2 hours/day for all instructional days
- to take its robotics students to both of the program-wide robotics competitions.

For each of these indicators, the site was coded as implementing with fidelity if all math classrooms were coded adequate. At the program level, each indicator was be coded as implemented with fidelity if at least 90% of the sites were coded adequate. At the program level, the "Robotics Instruction" component was judged to be implemented with fidelity if at least 90% of sites were judged adequate on both indicators.

In each of the three years, 100% of sites met the threshold for implementation fidelity on both these indicators.

Tables A1 through A3 (Appendix A) summarize these implementation fidelity findings.

Table A.1
Year 1 Middle School Summer STEM Program Implementation Fidelity Summary

	Indicator 1	Indicator 2	Indicator 3	Implemented with Fidelity?
Professional Development – Math	90% of teachers attending all PD			NO
	68% attending all PD -- NO			
Professional Development - Robotics	District delivery of full week PD			YES
	YES			
Instruction – Math	80 minutes math instruction daily	At least 5 minutes/day fact practice	4 weekly assessments	YES
	YES	YES	YES	
Instruction – Robotics	2 hours instruction daily	All sites brought students to competitions		YES
	YES	YES		

Table A.2

Year 2 Middle School Summer STEM Program Implementation Fidelity Summary

	Indicator 1	Indicator 2	Indicator 3	Implemented with Fidelity?
Professional Development – Math	90% of teachers attending all pre-program PD	90% of teachers attending at least 22 program days		NO
	70% attending all PD -- NO	78% attending at least 22 days[15] – NO		
Professional Development - Robotics	90% of teachers attending all pre-program PD	90% of teachers building robot in pre-program PD		NO
	76% attending all PD -- NO	76% completion – NO		
Instruction – Math	80 minutes math instruction daily	At least 5 minutes/day fact practice	4 weekly assessments	YES
	YES	YES	YES	
Instruction – Robotics	2 hours instruction daily	All sites brought students to competitions		YES
	YES	YES		

[15] 91% of math teachers attended at least 21 of the 24 program days.

Table A.3
Year 3 Middle School Summer STEM Program Implementation Fidelity Summary

	Indicator 1	Indicator 2	Indicator 3	Implemented with Fidelity?
Professional Development – Math	90% of teachers attending all pre-program PD	90% of teachers attending at least 22 program days		UNKNOWN (missing data)
	Attendance data available for only 2 of 3 PD days	97% attending at least 22 program days -- YES		
Professional Development - Robotics	90% of teachers attending all pre-program PD	90% of teachers building robot in pre-program PD		NO
	69% attending all PD -- NO	Data not available		
Instruction – Math	80 minutes math instruction daily	Implementation of provided Common Core Curriculum	Pre and Post assessments	YES
	YES	YES	YES	
Instruction – Robotics	2 hours instruction daily	All sites brought students to competitions		YES
	YES	YES		

Appendix B
Performance Goals

Program Goals (from Submitted Proposal)

Recruitment Goals: 1500 students will be recruited to participate in the program (Year 1 – 400; Year 2 – 500; Year 3 – 600). 80% of these students will be eligible for Free and Reduced Meals (FARM, a measure of poverty); 95% will represent minority racial and ethnic groups; 80% will have scored below Proficient on the previous year's math MSA; and 50% will be female.

Attendance Goal: At least 80% of the students who are enrolled in the summer program will attend at least 70% of the time. Student attendance will be tracked using the Student Management System (SMS).

Student Achievement Goals: 50% of students who scored below Proficient on the Math MSA the year prior to the Summer Learning Program will achieve a score of Proficient or Advanced the year following the program.

100% of students will maintain and/or increase their June Mathematics benchmark results when the same assessment is given in August, to show both retention and growth in mathematics grade-level aptitude.

All students who participate in the summer program will demonstrate mastery of the selected summer math skills, which are aligned to the Maryland State Curriculum, by scoring at least an 80% on concept assessments given during the summer program. The selected math skills will be pulled from the end-of-year math benchmark data – skills on which the majority of students show deficiency.

Participating students will increase their (1) desire to attend college, and (2) desire to engage in a STEM college major and/or career, as measured by a pre- and post-program survey.

Teacher Effectiveness Goals: In addition to goals aimed at student improvement, a secondary goal is to recruit teachers from underrepresented STEM areas and to improve their teaching effectiveness. The project provides two weeks of extensive and targeted professional development to participating teacher leaders to bolster their abilities in the STEM field, as well as professional development for one hour each day of the program.

Goals for teachers include the following:

Recruit 15-20% teachers from upper elementary grades, to encourage STEM learning in elementary grades.

All participating teachers will increase instructional performance scores on the City Schools Instructional Framework observation the year following their participation in the Summer Learning Program, as compared to the previous two years.

Summer Program Sites

2012 Program Sites

East Side

Bluford Drew Jemison STEM Academy (East),[16]
Northeast Middle School
Glenmount Elementary/Middle School
The Stadium School
Reach! Partnership
John Ruhrah Elementary/Middle School

West Side

Fallstaff Elementary/Middle School
Rognel Heights Elementary/Middle School
Violetville Elementary Middle (housing the program originally intended for Beechfield Elementary/Middle School as well)

2013 Program Sites:

East Side

Roland Park Elementary/Middle School
Hamilton Elementary/Middle School
The Stadium School
Reach! Partnership School

West Side

Coppin State University
Edmondson-Westside High School
Pimlico Elementary/Middle School
William Pinderhughes Elementary
Westport Elementary/Middle School
Rognel Heights Elementary/Middle School
Beechfield Elementary/Middle School

---

[16] The program was originally intended to be housed in the BDJ West school building, but was moved because of air conditioning issues.

2014 Program Sites:

East Side

Northeast Elementary/Middle School
Highlandtown (237) Elementary/Middle School
The Stadium School (on Baltimore's east side)

West Side

Coppin State University campus
Edmondson-Westside High School
Grove Park Elementary/Middle School
William Pinderhughes Elementary
Rognel Heights Elementary/Middle School
Beechfield Elementary/Middle School

Table B.1
Middle School Summer STEM Program
Math Teacher Certification and Experience

While teacher certification status and experience in teaching mathematics at the 6th grade level or higher were not explicitly mentioned within the original proposal, City Schools provided data on these measures and they appear relevant to discuss in this report. Table B2 reports the percentages of teachers with mathematics certification (either for grades 4-9 or for grades 7-12) and experience teaching mathematics at grade 6 or higher by program year.

	Year 1 (2012) N=43	Year 2 (2013) N=46	Year 3 (2014) N=34
% of SS Math Teachers certified for secondary math (either 4-9 or 7-12)	30.2%	39.1%	38.2%
% of SS Math Teachers who taught math at grade 6 level or higher in previous year	37.2%	56.5%	58.8%
% of SS Math Teachers who taught math at grade 6 level or higher in following year	51.2%	NA	58.8%

Note – None of the math teachers in Year 1 were designated in the data file received as teaching only 5th grade during the summer program, so all math teachers are included in the analysis group. In Year 2, the math teachers who were listed as only teaching 5th grade were excluded because the i3 study focused only on rising 6th to 8th graders. In Year 3 the program served only rising 6th, 7th, and 8th graders.

Table B.2
Characteristics of 2012 Middle School STEM Summer School Attenders and Non-Attenders, By Grade Level

	5th (to 6th)			6th (to 7th)			7th (to 8th)			All Middle Grades Students (Rising 6th to 8th)		
	Robotics Students	All program Students	Non-program students	Robotics Students	All program Students	Non-program students	Robotics Students	All program Students	Non-program students	Robotics Students [17]	All program Students	Non-program students
N	67	252	5663	73	286	5372	53	236	5440	195	776	16,475
% Female	26.9	43.3	49.4	28.8	37.4	49.9	20.8	30.9`	49.3	26.2	37.4	49.5
% FARMS	86.6	89.3	88.3	87.7	87.8	87.8	81.1	88.6	86.7	85.6	88.5	87.6
% Minority	94.0	96.0	87.7	97.3	95.5	88.4	94.3	92.4	89.0	95.4	94.7	88.3
% SPED	16.4	19.0	19.9	23.3	22.0	19.6	24.5	20.8	19.1	22.2	20.7	19.6
% Chronically absent in 2011-12	4.5	7.1	11.5	12.3	8.4	15.0	0.0	8.1	17.0	6.2	7.9	14.4

[17]This column and the next one both include 2 students in grade 8 in 2011-12.

Table B.3
Characteristics of 2013 Middle School STEM Summer School Attenders and Non-Attenders, By Grade Level

	5th (to 6th)			6th (to 7th)			7th (to 8th)			All Middle Grades Students (Rising 6th to 8th)		
	Robotics students	All program students	Non-program students	Robotics students	All program students	Non-program students	Robotics students	All program students	Non-program students	Robotics students	All program students	Non-program students
N	146	396	5694	136	338	5580	102	241	5589	384	591	16,863
% Female	41.1	48.7	48.9	41.2	48.2	48.9	40.2	52.3	48.8	40.9	49.4	48.9
% FARMS	91.1	89.6	88.8	86.0	87.6	88.0	87.8	88.4	88.2	87.8	88.6	88.2
% Minority	96.6	95.5	86.5	92.6	94.1	88.0	95.1	96.3	88.7	94.8	95.2	87.5
% SPED	19.9	21.5	17.8	16.9	18.6	20.6	24.5	21.2	19.6	20.1	20.4	19.4
Average Attendance Rate in 2012-13	96.2	96.2	94.0	95.4	95.3	93.6	96.1	95.6	92.8	95.9	95.7	93.5

Table B.4
Characteristics of 2014 Middle School STEM Summer School Attenders and Non-Attenders, By Grade Level

	5th (to 6th)			6th (to 7th)			7th (to 8th)			All Middle Grades Students (Rising 6th to 8th)		
	Robotics students	All program students	Non-program students	Robotics students	All program students	Non-program students	Robotics students	All program students	Non-program students	Robotics students	All program students	Non-program students
N	134	312	5661	140	304	5664	112	247	5578	386	863	16,903
% Female	33.6	44.9	48.5	32.1	40.5	49.6	33.0	45.3	48.8	32.9	43.5	48.9
% FARMS	89.6	92.3	88.7	87.1	91.8	89.4	88.4	89.9	88.1	88.3	91.4	88.7
% Minority	92.5	95.2	85.6	97.1	97.0	86.7	93.7	95.1	87.6	94.6	95.8	86.6
% SPED	33.6	34.3	21.2	32.9	30.3	21.3	32.1	30.0	23.0	32.9	31.6	21.8
Average Attendance Rate in 2013-14	95.2	94.8	93.6	94.9	95.1	93.2	94.1	94.4	91.9	94.8	94.8	92.9

Table B.5
2012 Middle School STEM Summer School Attendance Rates, by Grade Level
Robotics Students and All Program Students

		5th (to 6th)			6th (to 7th)			7th (to 8th)		All Middle Grades Students (Rising 6th to 8th)	
		Robotics Students	All Program Students		Robotics Students	All Program Students		Robotics Students	All Program Students	Robotics Students[18]	All Program Students
N		67	226		73	246		53	209	195	683
Average attendance rate		65%	58%		62%	58%		75%	62%	66%	59%
% attending at least 25%		91%	88%		88%	84%		91%	84%	89%	85%
% attending at least 50%		79%	66%		74%	65%		85%	68%	79%	67%
% attending at least 70%		51%	38%		45%	41%		75%	49%	55%	43%
% attending at least 80%		33%	23%		33%	27%		51%	31%	37%	27%

Note: Figures based on nine sites with reliable attendance data.

[18] Includes 2 students in grade 8 in 2011-12.

Table B.6
2013 Middle School STEM Summer School Attendance Rates, by Grade Level
Robotics Students and All Program Students

	5th (to 6th)		6th (to 7th)		7th (to 8th)		All Middle Grades Students (Rising 6th to 8th)	
	Robotics Students	All Program Students	Robotics Students	All Program Students	Robotics Students	All Program Students	Robotics Students	All Program Students
N	146	396	136	338	102	241	384	975
Average attendance rate	68%	68%	65%	65%	66%	63%	67%	66%
% attending at least 25%	90%	91%	92%	91%	91%	86%	91%	90%
% attending at least 50%	78%	79%	77%	76%	73%	71%	76%	76%
% attending at least 70%	64%	62%	52%	52%	58%	54%	58%	57%
% attending at least 80%	44%	39%	32%	35%	37%	35%	38%	37%

Table B.7
2014 Middle School STEM Summer School Attendance Rates, by Grade Level
Robotics Students and All Program Students

	5th (to 6th)		6th (to 7th)		7th (to 8th)		All Middle Grades Students (Rising 6th to 8th)	
	Robotics Students	All Program Students	Robotics Students	All Program Students	Robotics Students	All Program Students	Robotics Students	All Program Students
N	134	312	140	304	112	247	386	863
Average attendance rate	66%	67%	64%	64%	64%	63%	65%	65%
% attending at least 25%	87%	86%	84%	81%	88%	85%	86%	84%
% attending at least 50%	76%	77%	70%	70%	71%	72%	72%	73%
% attending at least 70%	57%	60%	56%	57%	52%	51%	55%	56%
% attending at least 80%	41%	44%	36%	41%	38%	35%	38%	40%

Appendix C
Methodology

Identification of Program Students for Analyses

2012 Program

Data on summer school participation in the 2012 program were received from both site-level files and the district student management system. Data on whether a student was enrolled in robotics or one of the other enrichment programs were missing for two of the ten sites. Nine of the ten program sites submitted program attendance data, which were used as the primary source for identifying program participants and their attendance. All students with 0 days attendance were excluded. For students with multiple records in the attendance files (sometimes from different schools), we selected the record with the most days attended recorded (or flipped a coin if both records had the same number of days attended). For the tenth school, we used the summer school SMS records to identify program participants, but all these students were missing attendance data. Students without identification numbers (32 total) were excluded, and an additional 16 students who could not be found in 2011-12 district files were also excluded. Only students who were in grades 5 through 7 in 2011-12 district files (and not coded as withdrawing from the district in the final record) were included as middle school student participants for analysis (2 students in grade 8 in 2011-12 were excluded). Attendance data were systematically missing for some weeks for three of the nine schools with attendance records. In these cases, students were assigned their attendance rate in the period where data were available.

2013 Program

Pre-enrollment records and program record data for 2013 were received from the district office, indicating which students were enrolled in robotics and their daily program attendance. These data were merged with other student record data to conduct analyses. Program records indicated that 397 students had at least one day of program attendance in robotics. Thirteen of these were eliminated from the analysis group because they either had no valid district identification number, had no district records in 2012-13, or were in grade levels other than those included in the i3 proposal. The total program analysis group therefore consisted of 384 robotics students at eleven sites.

2014 Program

Program record data for 2014 were received from the district office, indicating which students were enrolled in robotics and their daily program attendance. These data were merged with other student record data to conduct analyses. A total of 564 students were pre-enrolled or on the enrollment list at some point. Program records indicate that 397 of these students had at least one day of program attendance in robotics. Eleven of these were eliminated from the analysis group because they either had no valid district identification number, had no district records of attendance in 2013-14, or were in grade levels other than those included in the i3 proposal. The total program analysis group therefore consisted of 386 robotics students at nine sites.

Propensity Score and Mahalanobis Metric Matching Method

Given the large differences between summer program students and others, it was important to identify a closely matched comparison group to the treatment students with baseline equivalence on the pre-test measures of attendance and mathematics achievement. To this end, we combined propensity score and Mahalanobis metric matching using a caliper matching technique studied and recommended by Rubin and Thomas (2000). Only potential control students who did not attend summer school and who did have data on the prognostic covariates (prior year's attendance and prior year's state mathematics z-score) as well as the outcome variables were included in the matching analyses. In this two-step method, all control subjects meeting these requirements who were within ± .2 of the estimated propensity score of each treated subject were identified as potential matches. Then, Mahalanobis metric matching on the two prognostic covariates (prior attendance and prior mathematics score) was used to make a final selection of up to three matches for each treated student (3 to 1 matching).

Within each prior grade level (fifth, sixth, and seventh), we selected a comparison group subsample from among our larger sample of potential control subjects (so that the comparison group subsample had similar covariate values to the treatment sample on 15 covariates, including on the 2 key "prognostic" covariates). All of the matching was performed using nearest-remaining-neighbor matching, beginning with the most difficult to match treated subject (the one with the highest propensity score) and proceeding to the subject with the lowest propensity score. The propensity scores were estimated using logistic regression with linear terms for each covariate: the two prognostic covariates, eight student characteristics (dummy variables indicating whether the student is: a member of a minority group that is underrepresented in STEM careers in the U.S., male, a recipient of free or reduced lunches, in special education services, overage for grade, one who transferred between schools at least once in the prior year, one who attended summer school the prior summer, and one who was suspended at least once in the prior year), and five covariates measuring prior year characteristics of the student's post-summer program school (enrollment, % of students receiving free or reduced-price lunch, two dummy variables representing the three different gradespan types – K8, middle grades only, and middle high school, a dummy variable indicating whether the school is a district school or a charter school, and the school's average state mathematics assessment z-score in grades 6-8 for the prior year). We matched on prior year characteristics of the student's post-summer program school to address potential school effects on the outcome achievement score that was measured more than seven months after the end of the intervention. Because school choice (of post-summer program school) occurred before the summer intervention, the distributions on these variables were not affected by the intervention.

For Year 1 analyses, we conducted one matching procedure within each grade level for the attendance and Maryland School Assessment math achievement outcomes, and a separate matching procedure within each grade level for the Fall 2012 district mathematics benchmark outcome (because the large amount of missing data for the benchmark outcomes significantly reduced the sample size available for those analyses). Because there was a large amount of missing MSA data in 2014 (due to PARCC testing in some classrooms), we conducted two separate matching procedures (within each grade level) for the Year 2 attendance analysis sample and the Year 2 math achievement analyses.

Tables C1 and C2 illustrate (for the Year 1 attendance and MSA analyses) the difference between program students and non-program students by grade level, and the close matching achieved between treatment and comparison groups by the matching method. The other matching procedures (for mathematics benchmark analyses in Year 1, math MSA achievement in Year 2, and attendance in Year 2) had similar results (tables not included here).

List of Variables used in Propensity Score Matching for Comparison Group and in Analyses of Student Attendance/Achievement Effects

Pre-test mathematics achievement – z-score on previous school year's Maryland School Assessment-Mathematics
Pre-test attendance – Prior school year's attendance rate.

Student characteristics
- Gender (1=Male, 0=Female)
- Eligibility for free/reduced price lunch (1=yes, 0=no)
- Minority group underrepresented in STEM careers (1=Black\Hispanic\Native American\other non-white, non-Asian, 0=Caucasian/Asian)
- Special education status (1=yes, 0=no)
- Baseline overage for grade status (1=yes, 0=no)
- Prior year grade level (a set of dummy variables)
- Within-year school transfer during prior year (1=yes, 0=no)
- Suspended during the prior year (1=yes, 0=no)
- Attended summer school the prior school year
- Known to have attended the treatment (STEM Summer Learning with VEX Robotics) for two summers (1=yes, 0=no; this covariate is available only for Year 2 analyses)

School characteristics (prior year data on student's post-program school)
- Total student enrollment
- % of students eligible for free/reduced price lunch
- Type of school (1=District, 0=Charter)
- Gradespan (with K-8 schools as the reference category)
  - Middle school (1=yes, 0=no)
  - Middle-High school (1= yes, 0 = no)
- Average Maryland School Assessment z-score in grades 6-8

Table C.1
Differences in Group Means between Robotics Summer Program Students and Non-Program Students
Before Propensity Score Matching, By Prior Grade Level

	Grade 5			Grade 6			Grade 7		
	Program Students (n=57)	Non-program Students (n=4527)	Std. Mean Diff.	Program Students (n=63)	Non-Program Students (n=4388)	Std. Mean Diff.	Program Students (n=46)	Non-Program Students (n=4406)	Std. Mean Diff.
Male	0.72 (.45)	0.48 (.50)	0.53	0.70 (.46)	0.48 (.50)	0.50	0.76 (.43)	0.49 (.50)	0.00
FRL	0.86 (.35)	0.89 (.32)	-0.07	0.87 (.34)	0.88 (.32)	-0.01	0.80 (.40)	0.86 (.34)	-0.17
Minority	0.91 (.29)	0.87 (.34)	0.17	0.97 (.18)	0.88 (.33)	0.53	0.96 (.21)	0.88 (.33)	0.41
Spec. Ed	0.12 (.33)	0.15 (.36)	-0.10	0.22 (.42)	0.15 (.35)	0.16	0.20 (.40)	0.14 (.35)	0.17
Overage	0.11 (.45)	0.21 (.41)	-0.37	0.27 (.45)	0.21 (.41)	0.15	0.20 (.40)	0.22 (.41)	-0.08
Changed schools	0.02 (.13)	0.07 (.25)	-0.38	0.06 (.25)	0.08 (.27)	-0.08	0.04 (.21)	0.07 (.25)	-0.12
Suspended	0.07 (.26)	0.08 (.27)	0.03	0.10 (.30)	0.12 (.32)	-0.09	0.09 (.29)	0.12 (.33)	-0.15
Summer School prior year	0.46 (.50)	0.23 (.42)	0.46	0.27 (.45)	0.14 (.35)	0.33	0.52 (.51)	0.12 (.33)	0.84
Prior Math z-score	0.02 (1.13)	-0.03 (.99)	0.02	0.35 (.82)	0.38 (.92)	0.01	-0.12 (1.01)	0.06 (1.0)	-0.11
Prior attendance	96.9 (4.39)	95.0 (5.38)	0.42	95.96 (5.6)	94.84 (5.74)	0.18	97.07 (2.56)	94.35 (6.72)	1.11
Prior Year Characteristics of Students Post-Intervention School									
Enrollment	232 (141)	220 (135)	0.07	271 (146)	221 (136)	0.38	300 (161)	226 (138)	0.52
%FRL	81.2 (13.8)	84.9 (12.5)	-0.30	82.3 (14.5)	85.0 (12.5)	-0.21	82.6 (13.1)	85.3 (11.9)	-0.29
Charter	0.19 (.40)	0.19 (.39)	0.00	0.13 (.34)	0.18 (.39)	-0.18	0.07 (.25)	0.17 (.38)	-0.45
Middle School	0.28 (.45)	0.25 (.43)	0.06	0.30 (.46)	0.25 (.43)	0.09	0.22 (.42)	0.27 (.44)	-0.15
Middle High	0.16 (.37)	0.13 (.34)	0.06	0.22 (.42)	0.13 (.33)	0.25	0.22 (.42)	0.13 (.34)	0.18
Avg. Math z-score	0.04 (.51)	0.02 (.49)	0.05	-0.05 (.46)	0.03 (.48)	-0.14	0.05 (.46)	0.02 (.49)	0.14

(Standard deviations in parentheses)

Table C.2
Results for Differences in Group Means between Robotics Summer Program Students and Comparison Students after Propensity Score Matching, By Prior Grade Level

	Grade 5			Grade 6			Grade 7		
	Program Students (n=57)	Comparison Students (n=169)	Std. Mean Diff.	Program Students (n=63)	Comparison Students (n=189)	Std. Mean Diff.	Program Students (n=46)	Comparison Students (n=128)	Std. Mean Diff.
Male	0.72 (.45)	0.73 (.44)	-0.03	0.70 (.46)	0.70 (.46)	0.00	0.76 (.43)	0.76 (.43)	0.00
FRL	0.86 (.35)	0.89 (.32)	-0.08	0.87 (.34)	0.89 (.31)	-0.06	0.80 (.40)	0.80 (.40)	0.01
Minority	0.91 (.29)	0.92 (.27)	-0.04	0.97 (.18)	0.97 (.16)	-0.03	0.96 (.21)	0.96 (.19)	-0.04
Spec. Ed	0.12 (.33)	0.09 (.29)	0.09	0.22 (.42)	0.18 (.39)	0.10	0.20 (.40)	0.20 (.40)	-0.02
Overage	0.11 (.45)	0.12 (.33)	-0.06	0.27 (.45)	0.25 (.43)	0.05	0.20 (.40)	0.21 (.40)	-0.04
Changed schools	0.02 (.13)	0.02 (.13)	0.00	0.06 (.25)	0.05 (.21)	0.07	0.04 (.21)	0.06 (.24)	-0.09
Suspended	0.07 (.26)	0.07 (.26)	0.00	0.10 (.30)	0.06 (.24)	0.11	0.09 (.29)	0.12 (.32)	-0.10
Summer School prior year	0.46 (.50)	0.47 (.50)	-0.03	0.27 (.45)	0.25 (.43)	0.05	0.52 (.51)	0.46 (.50)	0.13
Prior Math z-score	0.02 (1.13)	0.04 (1.10)	-0.02	0.35 (.82)	0.37 (.80)	-0.02	-0.12 (1.01)	-0.12 (.93)	0.00
Prior attendance	96.9 (4.39)	96.8 (4.1)	0.02	95.96 (5.6)	96.39 (4.77)	-0.08	97.07 (2.56)	97.50 (2.39)	-0.17
Prior Year Characteristics of Students Post-Intervention School									
Enrollment	232 (141)	245 (164)	-0.09	271(146)	274 (161)	- 0.02	300 (161)	310 (157.1)	-0.06
%FRL	81.2 (13.8)	81.6 (17.8)	-0.03	82.3 (14.5)	82.0 (16.1)	0.02	82.6 (13.1)	83.6 (14.9)	-0.06
Charter	0.19 (.40)	0.19 (.39)	0.03	0.13 (.34)	0.17 (.38)	-0.14	0.07 (.25)	0.14 (.35)	-0.30
Middle School	0.28 (.45)	0.23 (.42)	0.10	0.30 (.46)	0.34 (.48)	-0.09	0.22 (.42)	0.31 (.46)	-0.22
Middle High	0.16 (.37)	0.12 (.32)	0.11	0.22 (.42)	0.19 (.39)	0.07	0.22 (.42)	0.24 (.43)	-0.06
Avg. Math z-score	0.04 (.51)	0.07 (.53)	-0.06	-0.05 (.46)	-0.03 (.53)	-0.02	0.05 (.46)	-0.06 (.54)	0.21

Description of Statistical Analyses for Attendance and Mathematics Achievement Outcomes

Measures

The primary outcome variables for this study were 1) yearly attendance rate (percent of days attended), calculated for all students from district administrative records on attendance in the year following the summer program treatment; and 2) mathematics achievement. Specifically, the achievement outcomes for the 2012 program were the student's z-score of mathematics scale score for the Fall 2012 district mathematics benchmark and the Spring 2013 administration of the Maryland School Assessment in mathematics. For the 2013 program students, only the Spring 2014 state Maryland School Assessment in mathematics was available. Only students with regular scores were included in analyses, as the alternative version of the test is scaled completely differently. Z-scores were calculated using the full district means and standard deviations for each grade level in the particular year.

The treatment variable indicates for each student whether he or she was in the robotics enrichment during summer school group (coded 1) or the matched comparison group with no summer school (coded 0). All covariates were grand mean centered in the impact models. Students were nested in the summer treatment sites, not the post-treatment schools, so in addition to their individual characteristics (from district administrative records), we estimated as level 1 variables the value of school level characteristics of the school attended the year following the intervention (measured the year prior to intervention, so as to ensure that covariates could not have been influenced by any impact of the intervention on student attributes). This allowed us to control for school level characteristics that could be associated with student attendance and achievement the year following treatment. School characteristics associated with the outcome variable, such as prior average achievement score and prior attendance rates, were included in the impact models to control for their effects on the outcome in the year following the summer program treatment. All covariates were pre-specified and included in the final model, regardless of their statistical significance.

Student level covariates included:

- Prior school year's attendance rate
- Z-score on previous school year's state assessment-in mathematics
- Gender (1=Male, 0=Female)
- Eligibility for free or reduced price lunch (1=yes, 0=no)
- Minority group underrepresented in STEM careers (1=Black\Hispanic\Native American\other non-white, non-Asian, 0=Caucasian or Asian)
- Special education status (1=yes, 0=no)
- Baseline overage for grade status (1=yes, 0=no)
- Prior year grade level (a set of dummy variables)
- Within-year school transfer during prior year (1=yes, 0=no)
- Suspended during the prior year (1=yes, 0=no)
- Attended summer school the prior school year (1=yes, 0=no)

Post-treatment school characteristics (from the prior year), assigned at Level 1, included:

- Total student enrollment
- % of students eligible for free or reduced price lunch
- Type of school (1=District, 0=Charter)
- Gradespan (with K-8 schools as the reference category)
  - Middle school 6-8 (1=yes, 0=no)
  - Middle-High school 6-12 (1= yes, 0 = no)
- Average State assessment z-score in grades 6-8
- Average attendance rate in grades 6-8

Statistical Model

We used a two-level random intercept model with covariates that assumed homogeneity of the treatment effects across sites. The treatment students were nested in eight summer treatment sites, and control students were nested together in a ninth site (no treatment). (Summer site effects were not a focus of the study.)This follows the constant block effect model described by Dong and Maynard (in press).

Level 1: Students within Sites

Level 1 describes the relationship between students' outcomes, student-level characteristics, and their treatment status. The level 1 model is

$$Y_{iJ} = \beta_{0j} + \beta_{1J}T_i + \Sigma\beta_{2s}X_{sij} + e_{ij},$$

where

Y_{iJ} is an outcome for student $_I$ in site $_j$;

T_i is 1 if the student is the treatment group and 0 otherwise;

X_{ij} is a set of S student-level covariates (described above) for student $_I$ in site $_j$, measured in the year prior to treatment exposure and centered on the grand mean in the sample; and

e_{ij} is a random error term for student $_I$ from site $_j$, assumed to be independently and identically distributed across students within sites (i.e., the "within-site" residual).

Level 2: Sites

$$\beta_{0j} = \gamma_{00} + u_0$$

$$\beta_{1J} = \gamma_{10}$$

$$\beta_{2s} = \gamma_{2S} \textit{ (and so on for each covariate)}$$

where

γ_{00} is the grand mean of the outcome variable (attendance)

γ_{10} is the main effect of treatment

The set of γ_{2S} regression coefficients represent the relationships between students' outcomes and the covariates, with each coefficient assumed to be constant across sites, U_{0j} J=1,…,J are fixed effects associated with each site effect, and are constrained to have a mean of zero.

All available covariates described earlier were included in the final model, regardless of their statistical significance. The purpose for including the prognostic covariates (pre-intervention achievement and attendance variables) was to control for students' prior achievement and prior attendance and to increase the precision of our impact estimates. The other covariates were intended to control for important student and school characteristics and to increase the precision of our impact estimates.

To test for baseline equivalence between the treatment and control students on attendance, we estimated a hierarchical linear model in the form specified above in which prior year's attendance was predicted by treatment status (controlling for grade level dummy variables).

Program Impacts on Attendance

In analyses conducted for the 2012 program, the intercept (adjusted mean attendance rate in the matched control group) was 95.6% of the days enrolled. The adjusted mean attendance rate of the treatment students was 1.4 percentage points higher, 97.0% of the days enrolled. This impact was both statistically significant (t (631) = 3.52, p =.001), and large enough to be educationally meaningful, Hedges's g = .26 Another way of stating the impact is that treatment students attended about 2.5 days more of the 180-day school year on average.

The Robotics Summer Program was designed by the district specifically as a way of reaching out and engaging students' who were not yet proficient in mathematics and the district's recruitment efforts were especially targeted toward enrolling such students --though students who were not low-performing in mathematics were also accepted into the program on a "space available" basis. Given the program's focus, in addition to the full sample analyses reported above, we also pre-planned to follow up these analyses with a subgroup analysis that estimated the effects of the 2012 robotics summer program on treatment students whose state achievement test proficiency level in mathematics in Spring 2012 was in the lowest category ("Basic").

Roughly a third (35%, 60 of 171) of the treatment students from the full sample had scored "basic" in mathematics on the state assessment in the spring prior to the summer program. These 60 treatment students (and their 167 matches from the comparison group) are the focus of our subgroup analyses. A formal baseline equivalence test that compared the prior attendance rates of treatment and comparison students in this "non-proficient" subsample using a two-level random intercept HLM model which took account of students' grade level and nesting in summer program sites is summarized in Table 3. The estimated difference between treatment and control means in the model was -0.42, a small nonsignificant prior attendance advantage for the control group. Given baseline equivalence in prior attendance, we then tested for program impacts on subsequent attendance of treatment and control students in this subsample using our two-level random intercept model.

The intercept (adjusted mean 2012-2013 attendance rate in the subsample's matched control group) was about 93.8% of the days enrolled. The adjusted mean attendance rate of the treatment students in the subsample was 2.6 percentage points higher, 96.4% of the days enrolled. This impact was both statistically significant (t (206) = 2.865, p. = .005), and large enough to be educationally meaningful, Hedges's g = .37. Another way of stating the impact is that treatment students in the subsample attended, on average, about a week more of school than did the control students in the subsample (i.e., attended 4.7 days more during the course of the 180-day school year.)

Parallel analyses were conducted for the 2013 program. After adjusting attendance rates statistically for all the demographic, status, achievement, and school characteristics variables included in the analyses, 2013 program students had average attendance rates of 0.6 percentage points higher than the comparison group the year following the program (95.0% vs. 94.4%), which was not a statistically significant difference. Parallel analyses for low-achieving students in mathematics (157 program students closely matched to 453 comparison students) found that those program students had average attendance rates of 1.5 percentage points higher than comparison students (94.9% vs. 93.4%). This was also not a statistically significant difference. Statistical analyses for the full group of students which included an interaction term did indicate, however, that there was a significantly greater attendance impact of the program for low-achieving students compared to higher-achieving students.

Finally, we examined whether there was still a program effect on attendance a year later (2013-14) for the 2012 robotics program participants. About a quarter of the 2012 participants had progressed to ninth grade, a year when attendance typically declines notably. A total of 157 program students and 462 comparison students had 2013-14 attendance data, and remained closely matched on 2011-12 attendance. Identical analyses were conducted using 2013-14 attendance as the dependent variable. Program students had average attendance rates of 1.5 percentage points higher than comparison students (95.2% vs. 93.7%), an effect that approached statistical significance (p=.066) but did not meet the formal criteria. Among the low-achieving students (56 program students closely matched to 159 comparison students), the attendance difference was 2.4 percentage points (93.6% for program students vs. 91.2% for comparison students), but this effect was not statistically significant (due to the small number of students and large variation in attendance rates).

Program Impacts on Mathematics Achievement

In Year 1 it was possible to examine the impact of the summer program on a proximal measure of mathematics achievement, the Fall 2012 district mathematics benchmark test. Because not all schools administered the test, scores were available for less than three-quarters of the program students. Separate matching procedures were conducted to identify a closely matched comparison group for program students who could be included in these analyses. The comparison group (n=352) was virtually identical to the group of program students (n=121) in prior mathematics achievement and attendance, and closely matched on more than a dozen other demographic and school characteristics.

Analyses were conducted for achievement using the same models as for the attendance analyses described above. Program students did not have significantly higher mathematics achievement scores than comparison students, for either the full sample or the subgroup of low-achieving students.

Parallel analyses were conducted using the Spring 2013 mathematics MSA scores with the same sample of students as in the Year 1 attendance analyses (166 program students, 486 comparison students). Program students did not have significantly higher mathematics achievement scores than comparison students, for either the full sample or the subgroup of low-achieving students.

For the Year 2 (2013) program students, only the Spring 2014 mathematics MSA scores were available for analyzing program impact on mathematics achievement. Because some students at each school took the PARCC assessment rather than the MSA, scores were not available for the full group of Year 2 students. Separate matching procedures were conducted to identify a closely matched comparison group for program students who could be included in these analyses. The comparison group (n=286) was virtually identical to the group of program students (n=828) in prior mathematics achievement and attendance, and closely matched on more than a dozen other demographic and school characteristics.

Analyses were conducted for achievement using the same models as for the analyses described above. Program students did not have significantly higher mathematics achievement scores than comparison students, for either the full sample or the subgroup of low-achieving students.

Aspiration Scales
Description of Statistical Analyses for Aspirations Outcomes

Measures

Outcome measures of student aspirations for studying STEM subjects in college and for STEM careers were created from multiple items designed by the evaluation team for the survey administered to summer program students at the beginning and end of the program. Each of the scales was constructed as the mean response to items from a multi-item scale. Each item has an identical response scale: strongly disagree (1), disagree (2), agree (3), strongly agree (4). We created three scales:

College-going Aspirations Scale (Cronbach's alpha = 0.71)
Going to college after high school is important
Going to college is important for achieving my future goals.
I plan to go to college following high school.

Aspirations to Study STEM in College Scale (Cronbach's alpha = 0.76)
I want to study math and science in college
I would enjoy taking math and science classes in college
Taking math and science in college is important for achieving my future goals.

STEM Career Aspirations Scale (Cronbach's alpha =0.80)
I would like to be a scientist.
I would like a job working with robot.
I would like a job where I invent things.
I would like to design machines that help people.
I would enjoy a job helping to protect the environment.
I would enjoy a job in the medical field.
I would enjoy a job doing scientific research.

The treatment variable indicates for each student whether he or she was in the robotics enrichment during summer school group (coded 1) or the comparison group of summer school participating in the arts and sports enrichments (coded 0).

All covariates were grand mean centered in the impact models. Students were nested in the summer treatment sites. All covariates were pre-specified and included in the final model, regardless of their statistical significance.

Student level covariates included:
Prior school year's attendance rate
Z-score on previous school year's state assessment-in mathematics
Gender (1=Male, 0=Female)
Eligibility for free or reduced price lunch (1=yes, 0=no)
Minority group underrepresented in STEM careers (1=Black\Hispanic\Native American\other non-white, non-Asian, 0=Caucasian or Asian)
Special education status (1=yes, 0=no)

Baseline overage for grade status (1=yes, 0=no)
Prior year grade level (a set of dummy variables)
Within-year school transfer during prior year (1=yes, 0=no)
Suspended during the prior year (1=yes, 0=no)
Attended summer school the prior school year (1=yes, 0=no)

Statistical Model

We used a two-level random intercept model with covariates that assumed homogeneity of the treatment effects across sites. The treatment students were nested in summer treatment sites, and control students were nested together in a separate site (no treatment). (Summer site effects were not a focus of the study.)This follows the constant block effect model described by Dong and Maynard (2013).

Level 1: Students within Sites

Level 1 describes the relationship between students' outcomes, student-level characteristics, and their treatment status. The level 1 model is

$$Y_{iJ} = \beta_{0j} + \beta_{1J}T_i + \Sigma\beta_{2s}X_{sij} + e_{ij},$$

where

Y_{iJ} is an outcome for student $_I$ in site $_j$;

T_i is 1 if the student is the treatment group and 0 otherwise;

X_{ij} is a set of S student-level covariates (described above) for student $_I$ in site $_j$, measured in the year prior to treatment exposure and centered on the grand mean in the sample; and

e_{ij} is a random error term for student $_I$ from site $_j$, assumed to be independently and identically distributed across students within sites (i.e., the "within-site" residual).

Level 2: Sites

$$\beta_{0j} = \gamma_{00} + u_0$$

$$\beta_{1J} = \gamma_{10}$$

$$\beta_{2s} = \gamma_{2S} \text{ (and so on for each covariate)}$$

where

γ_{00} is the grand mean of the outcome variable (attendance)

γ_{10} is the main effect of treatment

The set of γ_{2S} regression coefficients represent the relationships between students' outcomes and the covariates, with each coefficient assumed to be constant across sites,
U_{0j} J=1,…,J are fixed effects associated with each site effect, and are constrained to have a mean of zero.

All available covariates described earlier were included in the final model, regardless of their statistical significance.

To test for baseline equivalence between the treatment and control students in attitudes, we estimated a hierarchical linear model in the form specified above in which pre-test aspirations measures were predicted by treatment status (controlling for grade level dummy variables).

Table C.3

Pre and Post-Survey Unadjusted Aspiration Scale Means for 2013 i3 Middle School STEM Summer Program Students

	Robotics Students	Students in Other Enrichments	Overall
	n=40	n=98	n=138
College-Going			
Pre	3.73 (.456)	3.70 (.466)	3.71 (.461)
Post	3.58 (.617)	3.68 (.599)	3.65 (.603)
College STEM Course Taking			
Pre	3.14 (.773)	3.18 (.667)	3.17 (.697)
Post	3.13 (.754)	3.09 (.773)	3.10 (.765)
STEM Career Aspirations			
Pre	2.42 (.735)	2.51(.585)	2.48 (.631)
Post	2.48 (.732)	2.56 (.674)	2.53 (.690)

Table C.4
Pre and Post-Survey Unadjusted Aspiration Scale Means for 2014 i3 Middle School STEM Summer Program Students

	Robotics Students n=70	Students in Other Enrichments n=100	Overall n=170
College-Going			
Pre	3.70 (.457)	3.79 (.395)	3.75 (.422)
Post	3.67 (.531)	3.69 (.557)	3.68 (.545)
College STEM Course Taking			
Pre	3.12 (.696)	3.12 (.786)	3.12 (.748)
Post	3.13 (.712)	3.15 (.747)	3.14 (.731)
STEM Career Aspirations			
Pre	2.67 (.662)	2.63(.615)	2.65 (.633)
Post	2.62 (.743)	2.63 (.681)	2.62 (.705)

Scales ranged from 1.0 (lowest aspirations) to 4.0 (highest aspirations).

Impact of Professional Development on Teacher Effectiveness Scores

Sample

The full sample for the evaluation study was the population of mathematics teachers during Year 2 of the i3-funded intervention (a total of 58 teachers).[19] Teachers received 4 days of common professional development in mathematics instruction prior to the beginning of the summer program, and additional professional development during the 5-week program at the 12 different sites throughout the district.

Teachers voluntarily applied to teach in the summer program by the end of March prior to the summer school professional development in mid-June. Those with unsatisfactory ratings the prior school year were excluded from selection for the program. Eligible teachers were selected by district summer program staff and site coordinators.

Data collection

All data are from Baltimore City Public Schools teacher administrative records. Pre-intervention data were available for 2012-13 (prior to Year 2 Summer program), and post-intervention data were available for the following year, 2013-14.

Dependent Variable

Teacher "instructional performance" scores are from the official, district-wide evaluation of teacher performance (The Baltimore City Teacher Observation Rating Form), based on a combined rating on nine instructional components. Teachers are scored as either ineffective (1), developing (2), effective (3), or highly effective (4) on each of the following nine areas:

1) Communicate standards-based lesson objectives
2) Present content clearly
3) Use strategies and tasks to engage all students in rigorous work
4) Use evidence-dependent questioning
5) Check for understanding and provide specific, academic feedback
6) Facilitate student-to-student interaction and academic talk
7) Implement routines to maximize instructional time
8) Build a positive, learning-focused classroom culture
9) Reinforce positive behavior, redirect off-task behavior, and de-escalate challenging behavior

The reliability of the outcome measure was calculated from the post-intervention data and yielded a Cronbach's alpha of 0.90.

[19] Because the teacher evaluation system has just changed in the district, prior instructional performance scores were not available for teachers in Year 1 of the program. Post-scores for teachers in Year 3 of the program will not be available until after the evaluation funding has ended. It is therefore only possible to conduct this teacher impact study for Year 2 of the intervention.

The effectiveness score was constructed as the mean of the 9 indicators, which were each scored on a scale of 1 to 4. The total possible average score range for each observation was therefore between 1 and 4.

Analysis

We used a paired t-test[20] to estimate whether gains in instructional performance scores after receiving professional development through this i3-funded intervention are statistically different from zero. Other teacher covariates were not available for analysis.

Because two observations were conducted each year, the best estimation of the impact of the summer professional development intervention is a comparison of the second score from Year 1 (from an observation generally conducted in the spring) with the first score from Year 2 (from an observation generally conducted in the fall). A total of 44 teachers have full data available for this benchmark analysis.

As the first sensitivity analysis (based on an n of 52 teachers with full data available), we compared the mean of the observations for Year 2 with the mean of the observations for Year 1. This analysis is more likely to provide an overestimation of the treatment effect, because the mean for Year 1 is affected by the generally lower first observation score and is generally lower than the second observation score. It is the second Year 1 observation score that is the closest in time to the intervention and the best estimate of the baseline score.

No missing outcome data or pre-test data were imputed. We calculate an "effect size" according to the following formula: $(X_2 - X_1) / SD_1$, where

X_1 = Mean of teachers instructional performance scores Time 1 (pretest)
X_2= Mean of teachers instructional performance scores Time 2 (posttest)
SD_1= Standard Deviation of instructional performance scores at Time 1 (estimated from the data at Time 1).

For the benchmark analysis, we conducted a matched sample difference of means test for the 44 teachers who had both a second observation score in Year 1 (later in year) and first observation score in Year 2 (closer in proximity to the professional development intervention). The difference between the first Year 2 mean score (2.86) and the last Year 1 mean score (2.77) was 0.09. The standard deviation in instructional effectiveness scores in Year 1 was 0.57, for an effect size of 0.16. The t value (0.76) was not significant (p=.288).

For the first sensitivity analysis, we conducted a similar paired t-test on the difference between the "average effectiveness scores" for Year 1 and Year 2 for the 52 teachers who had at least one effectiveness score in each of the two years. The difference between the average Year 2 score (2.93) and the average Year 1 score (2.76) was .17. The standard deviation in instructional effectiveness scores in Year 1 was .48, for an effect size of .35. The t value (2.52) was significant at p=.015.

[20] See Duckart, 1998, for a similar analytical strategy.

We conducted an additional sensitivity analysis on average scores, using the sample of 44 teachers used in the benchmark test. The difference between the average Year 2 score (2.90) and the average Year 1 score (2.74) was .16. The standard deviation in instructional effectiveness scores in Year 1 was .49, for an effect size of .33. The t value (1.97) had a p value of .055.

Although the increase in teacher effectiveness scores was statistically significant in the first sensitivity test, it was not significant in the approach using time points that give the best estimate of impact. The sensitivity analysis conducted with the smaller sample in the second sensitivity test yielded a much smaller p value than the benchmark test, which suggests that the difference in significance results between the approaches is not due to the smaller sample size.

Even if we could conclude with confidence that the increase in effectiveness scores after the professional development intervention was statistically significant, data are not available to ascertain whether scores also rose at the same rate for a comparable group of teachers who did not receive the intervention. It is therefore not possible to confidently attribute the improved scores to the intervention itself.