RESEARCHREPORT

AN EMPIRICAL EDUCATION REPORT

Efficacy of the Alabama Math, Science, Technology Initiative (AMSTI) on Math, Science, and Reading Achievement

A REPORT OF A QUASI-EXPERIMENT IN ALABAMA

December 2019

Val Lazarev

Adam Schellinger

Jenna Zacamy

Denis Newman

Empirical Education Inc.

Acknowledgements

We are grateful to the staff at the Alabama State Department of Education (ALSDE) and the AMSTI team for providing the data needed to conduct this research. The research was sponsored by ALSDE which provided Empirical Education Inc. with independence in reporting the results.

The analysis presented in this report was conducted over two separate periods in 2018 and 2019. In 2018, we focused on the impact of AMSTI on math and science outcomes for students overall and for specific subgroups. Based on our initial findings and recommendations, ALSDE requested follow-on analysis in 2019, including investigating the impact of AMSTI on reading outcomes, further analysis of the impact on English Language Learners, and analysis on the impact on early career teachers. This report adds additional results to the original report: Lazarev, Schellinger, Zacamy, & Newman, 2018.

About Empirical Education Inc.

Empirical Education Inc. is a Silicon Valley-based research company that provides tools and services to help K-12 school systems make evidence-based decisions about the effectiveness of their programs, policies, and personnel. The company brings its expertise in research, data analysis, engineering, and project management to customers that include the U.S. Department of Education, educational publishers, foundations, leading research organizations, and state and local education agencies.

Table of Contents

Reference this report: Lazarev, V., Schellinger, A., Zacamy, J., & Newman, D. (2019). *Efficacy of the Alabama Math, Science, Technology Initiative (AMSTI) on Math, Science, and Reading Achievement. A Report of a Quasi-experiment in Alabama.* (Empirical Education report number: Empirical_ UM_AMSTI-7038-FR1-2019-O.1). Palo Alto, CA: Empirical Education Inc. Retrievable from bit.ly/AMSTIreport

Key Findings and Recommendations

OVERVIEW AND KEY FINDINGS

- Empirical Education Inc. conducted a quasi-experiment investigating the impact of the Alabama Math, Science, Technology Initiative (AMSTI) training. The study compared classes of fully-trained AMSTI teachers to classes taught by teachers with no AMSTI training within the same schools. The outcome measure was the ACT Aspire in math, science, and reading from spring 2017.
- On average, AMSTI training had a positive impact on student achievement in math, science, and reading. The results are equivalent to a gain of a 2-percentile point improvement on the ACT Aspire math assessment, 4-percentile points on the ACT Aspire science assessment, and 2-percentile points on the ACT Aspire reading assessment. Percentile gain is a standardized metric indicating the estimated improvement that the average comparison group class would see if the class had been taught by an AMSTI trained teacher.
- This overall result replicates a large-scale randomized control trial (RCT) that Empirical conducted over a decade ago.
- We found no difference in the effect of AMSTI training based on school characteristics, including elementary vs. middle school, school-wide percent of minority students or English Language Learners (ELLs), teacher characteristics (gender and advanced degree), or the percentage of classes in a school taught by an AMSTI teacher.
- AMSTI had a positive impact on math achievement for the following student subgroups: males, females, White students, Black students, students eligible for free or reduced-price lunch, and special education students.
- AMSTI had a positive impact on science achievement for the following student subgroups: males, females, and students eligible for free or reduced-price lunch. However, AMSTI had a negative effect on science achievement for ELLs.
- AMSTI had a positive impact on reading achievement for the following subgroups: males, females, White students, and students eligible for free or reduced-price lunch.

RECOMMENDATIONS FOR IMPROVEMENT AND FOLLOW-ON RESEARCH

- AMSTI had a strong benefit in science for females. The program is currently piloting units to help teachers address the standards in the digital literacy and computer science course of study. The AMSTI team should consider how the components of the program that may be more engaging for girls in science can be utilized with these new units. This could lead to more females continuing their education and career in STEM and other technical fields where they are traditionally underrepresented.
- Given the finding of the negative impact of AMSTI on science achievement for ELLs, AMSTI should consider how the language in the AMSTI science materials can be modified and improved to be more accessible, and what additional science training teachers can be provided with a focus on engaging ELLs.
- In math especially, economically disadvantaged students benefitted from AMSTI. Since these students are found predominantly in urban centers and rural areas, their success may provide an indicator of the kind of program that can continue to flourish in these locales that often fall behind academically.

- As the Alabama Course of Study continues to evolve and the state rolls out a new assessment, the impact of AMSTI should be measured each year.

Introduction

This report presents the findings of a quasi-experiment investigating the effectiveness of the Alabama Math, Science, Technology Initiative (AMSTI). This study was conducted by Empirical Education Inc. (Empirical) in response to the Alabama legislature's request of the State Superintendent of Education to commission "a study of the AMSTI program with a recognized research organization to ascertain the effectiveness of the program in schools utilizing AMSTI. The study shall focus on student achievement and teacher effectiveness in STEM and STEM related courses and the study shall include recommendations for improvements."

From 2005 to 2008, Empirical conducted an evaluation of AMSTI under a federal Department of Education contract. With productive cooperation of the Alabama State Department of Education (ALSDE) and the AMSTI staff, the experimental study of AMSTI involved 82 Alabama schools and over 700 teachers and showed an overall positive effect (Newman et al., 2012). In that study, we randomly assigned schools that had applied that year to join the program to either become AMSTI schools or serve as the control group and join AMSTI in the following year.

In the current study, we compared student achievement for classes of fully-trained AMSTI teachers to classes of teachers with no AMSTI training within the same school. The study used data from the state student information system, aggregated at the class level, including 2016/17 demographic data and spring 2017 ACT Aspire outcomes, as well as training records from the AMSTI database. This report describes the research questions addressed, the research methodology used, the data that was collected from the state, the results of the study, and potential areas for improvement and follow-on research.

Research Questions

This study addressed the following questions:

1. Is there a positive impact of AMSTI on math achievement for students in grades 4-8?
2. Is there a positive impact of AMSTI on science achievement for students in grades 5 and 7?
3. Is there a positive impact of AMSTI on reading achievement for students in grades 4-8?
4. Does the impact of AMSTI on student achievement vary across important student subgroups (e.g., depending on demographic characteristics of students and schools)?
5. Is there a positive impact of AMSTI on student achievement for early career teachers?
6. Where can improvement efforts be applied in the AMSTI process to produce greater program effectiveness?

While AMSTI is designed to primarily impact math and science achievement, findings from our study published in 2012 included a positive impact on reading achievement. The AMSTI team posited that this was because AMSTI had purposefully integrated reading and writing practices into its science training modules, based on recommendations from national and international mathematics and science standards. Therefore, we added the research question about reading outcomes to the current study to see if the findings would be replicated. To analyze this question, we looked at reading outcomes for students of fully-trained AMSTI science teachers.

Study Description

ABOUT AMSTI

AMSTI is the Alabama State Department of Education's initiative to improve math and science teaching statewide and has been training teachers since 2002. The two-year intervention is designed to better align classroom practices with national and statewide teaching standards, and ultimately to improve student achievement. The core beliefs on which AMSTI was built are that effective teachers clearly understand the standards they are to teach, have strong content knowledge of their subject, know how to engage students and scaffold their learning with age and grade appropriate activities, and revise instruction based on feedback from formative assessment. To help teachers develop in these areas, the initiative provides AMSTI teachers with three basic services: professional development, equipment and materials, and onsite support from content specialists.

STUDY DESIGN

This study compared achievement scores for two groups of classes within the same schools: (1) classes taught by fully-trained AMSTI teachers and (2) classes taught by teachers with no AMSTI training. For this study, "fully-trained" AMSTI teachers refer to teachers trained in all required AMSTI units for the grade and subject that they teach. Because approximately 80% of schools in the study dataset had teachers with and without AMSTI training, we compared classes taught by teachers with and without AMSTI within schools, rather than comparing classes across schools. By comparing classes from the same school, we were able to control for school characteristics thereby minimizing bias due to school adoption into AMSTI. Appendix A includes additional technical details of the methodology used.

The outcome measure used was the ACT Aspire in math, science, and reading from the 2016/17 school year. In consultation with ALSDE, the study team selected 2016/17 as the outcome year because data for the 2017/18 school year would not be available until late fall 2018. Additionally, the state had administered a new assessment (the Scantron) in spring 2018. One limitation of the decision to use spring 2017 test scores is that teachers trained in AMSTI science only had one year of training with the new AMSTI science materials prior to the outcome, as opposed to the two consecutive summers of training that AMSTI typically provides. The Alabama Course of Study (COS) in science was re-written in 2015, and the AMSTI science training and materials were subsequently re-written to align with the COS, with the first training using the new materials in summer 2016.

DATA COLLECTION AND SAMPLE IDENTIFICATION

This study used existing data collected from the information system at ALSDE and from the AMSTI team.

The ALSDE database team provided student achievement scores and demographics, along with teacher demographic information. Data included:

- School System, School, and Teacher ID
- Teacher demographic data (years teaching experience, highest degree, ethnicity, gender)
- Course information (course code, name, category, grade, subject area)
- Class-level student demographics (total number of students, % of students in the class based on: gender, race/ethnicity, socio-economic status, ELL status, Special Education status)

- Class-level student achievement data (ACT Aspire math, science, and reading scale scores and percent proficient for 2016/17 and 2015/16; overall and disaggregated by subgroups)

The AMSTI team provided active training certifications for the 2016/17 school year for both math and science, a list of all teachers who have ever received AMSTI certification, and a list of teachers who are inactive on their past AMSTI certifications. Figure 1 shows a map of Alabama with AMSTI regions identified. The blue dots represent all 831 schools with AMSTI trained teachers (according to data from the AMSTI team from the 2016/17 school year).

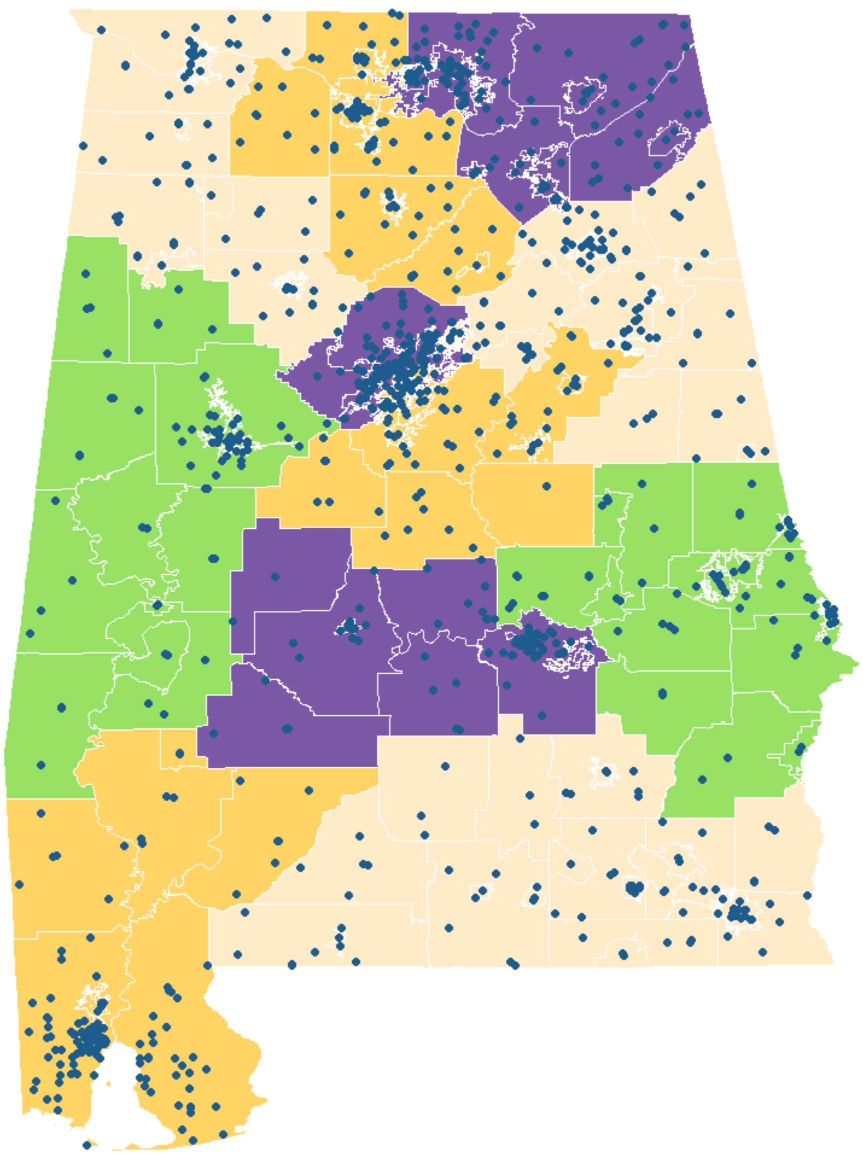

FIGURE 1. SCHOOLS WITH AMSTI TRAINED TEACHERS

We used the following selection criteria to identify classes to include in the primary analysis:

- Only general science and math classes (no STEM elective)
- Contain less than 100% special education students
- Class size greater than or equal to 8 students
- Pretest and outcome metrics are available
- Not taught by first year teachers who would not have had the opportunity to attend the AMSTI training.

We selected schools that had at least one class, satisfying the conditions specified above, taught by a fully-trained AMSTI teacher and at least one class taught by a teacher with no known training. All classes satisfying the conditions specified above in qualifying schools were included in the analysis.

Figure 2 and Figure 3 show a map of Alabama with AMSTI regions identified. The blue dots in the map in Figure 2 represent the 256 schools that had both teachers fully trained in AMSTI math and teachers with no known training, which were the schools used in the math analysis for this study. The blue dots in the maps in Figure 3 represent the 112 schools that had both teachers fully trained in AMSTI science and teachers with no known training, which were the schools used in the science analysis for this study. The smaller sample included in the science analysis, due to the AMSTI science training timeline and that science outcomes are only available in grades 5 and 7 in Alabama, led to less precision in the analysis of science outcomes. The analysis of the impact of AMSTI on student achievement in reading used a subset of science classes for which reading outcomes and pretest data were available.

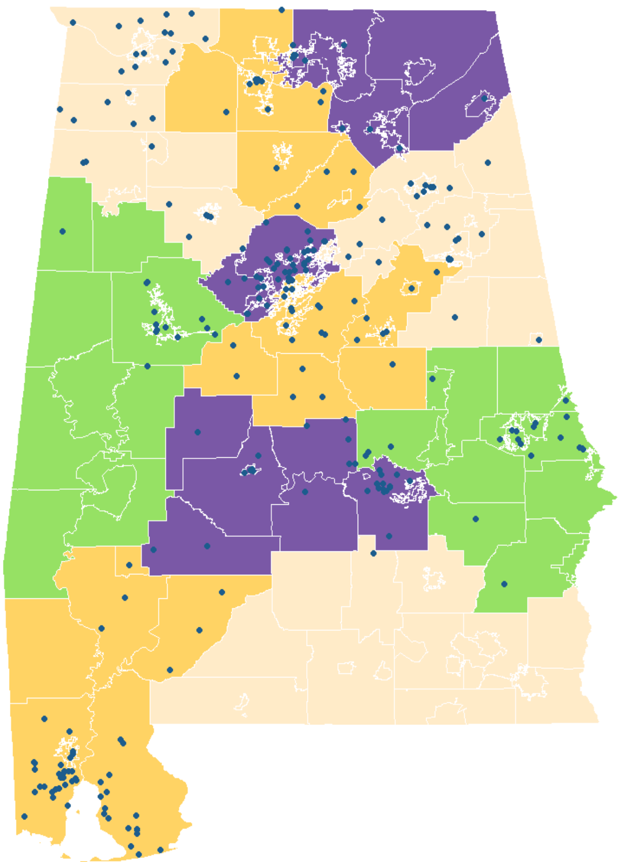

FIGURE 2. SCHOOLS USED IN MATH ANALYSIS

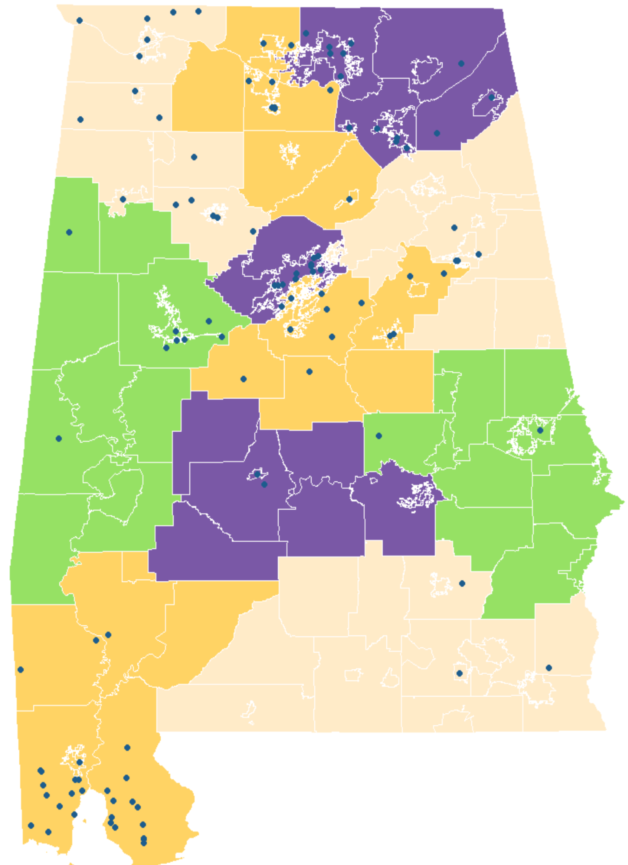

FIGURE 3. SCHOOLS USED IN SCIENCE ANALYSIS

Results

OVERALL

We found a positive impact of AMSTI training on ACT Aspire math, science, and reading. The analysis adjusted for differences in student demographics and pretest between classes taught by AMSTI-trained and non-AMSTI-trained teachers. The results are equivalent to a gain of 2-percentile points on the ACT Aspire math assessment, 4-percentile points on the ACT Aspire science assessment, and 2-percentile points on the ACT Reading assessment (Figure 4). Percentile gain is a standardized metric allowing us to compare improvements found with the different tests (math, science, and reading). Percentile gain can be understood as the change in percentile rank that the average comparison group class would see if the class had been taught by an AMSTI trained teacher. Appendix B includes the full results of the impact of AMSTI overall and by subgroups.

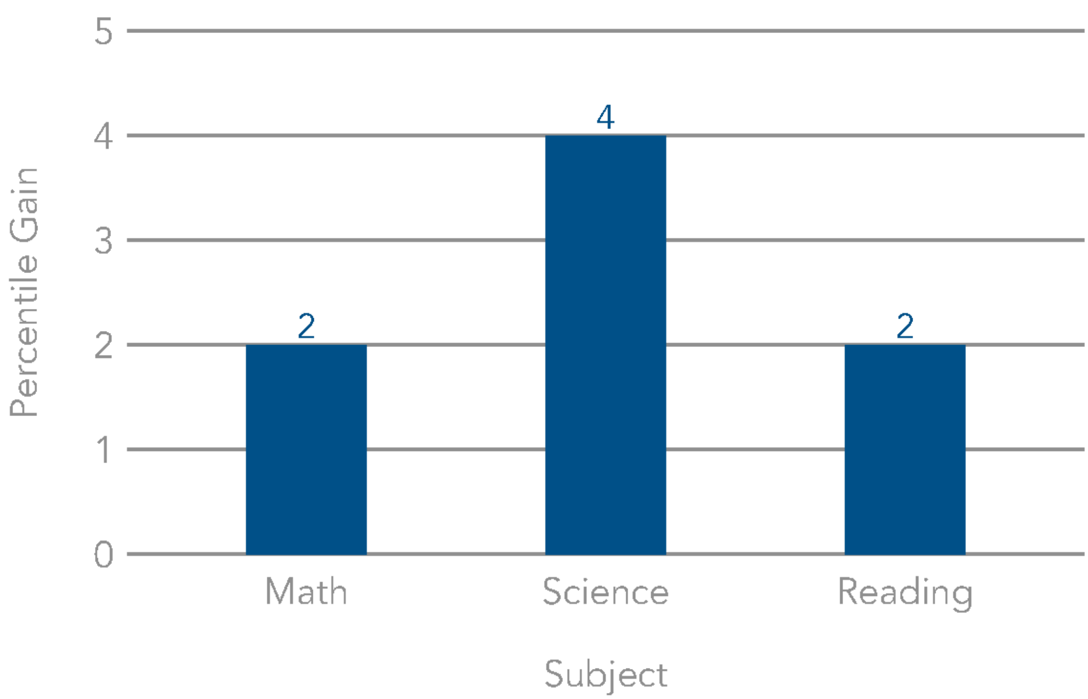

FIGURE 4. OVERALL IMPACT OF AMSTI IN MATH, SCIENCE, AND READING

EFFECT OF SCHOOL CHARACTERISTICS

We found no differential effect of AMSTI training based on school characteristics, including:

- Grade level: there were no differences in the impact of AMSTI between elementary and middle grades
- School-wide percent of minority students or ELLs: there were no differences in the impact of AMSTI for schools with higher proportion of minority or ELL students

- Teacher characteristics: there were no differences in the impact of AMSTI for teachers with an advanced degree or based on their gender
- Saturation of AMSTI teachers: there was no differences in the impact of AMSTI for schools with higher percentage of classes in the school taught by fully-trained AMSTI teachers

EFFECT OF STUDENT CHARACTERISTICS

There were generally positive results for subgroups defined in terms of some important student characteristics. Exploring the impact for different student subgroups and identifying for which groups the program does and does not work is essential in order to implement program improvement. In the study published in 2012, the Empirical research team found an overall impact of AMSTI very similar to the current study. But we also found a significant differential impact of AMSTI favoring non-minority students. In other words, we found that AMSTI was increasing the achievement gap between White and Black students in math, science, and reading. Figure 5 represents this result from the earlier (2012) study. The current study gave us the opportunity to investigate if that result would be replicated a decade later. This time we found a very different pattern—AMSTI benefitted both Black and White students. There are many differences between the prior and current studies to keep in mind. First, over a decade of program changes have taken place. Additionally, the two studies used different outcome measures, which may have different characteristics, and the ACT Aspire may be more amenable to what AMSTI is trying to teach. Most interestingly, the randomized control trial required that participating teachers were new to the program and would have only had one or two years of exposure, whereas the current study includes teachers with various years of experience with AMSTI. It is possible that teachers improve their AMSTI practice with different subgroups of students over time with more program experience.

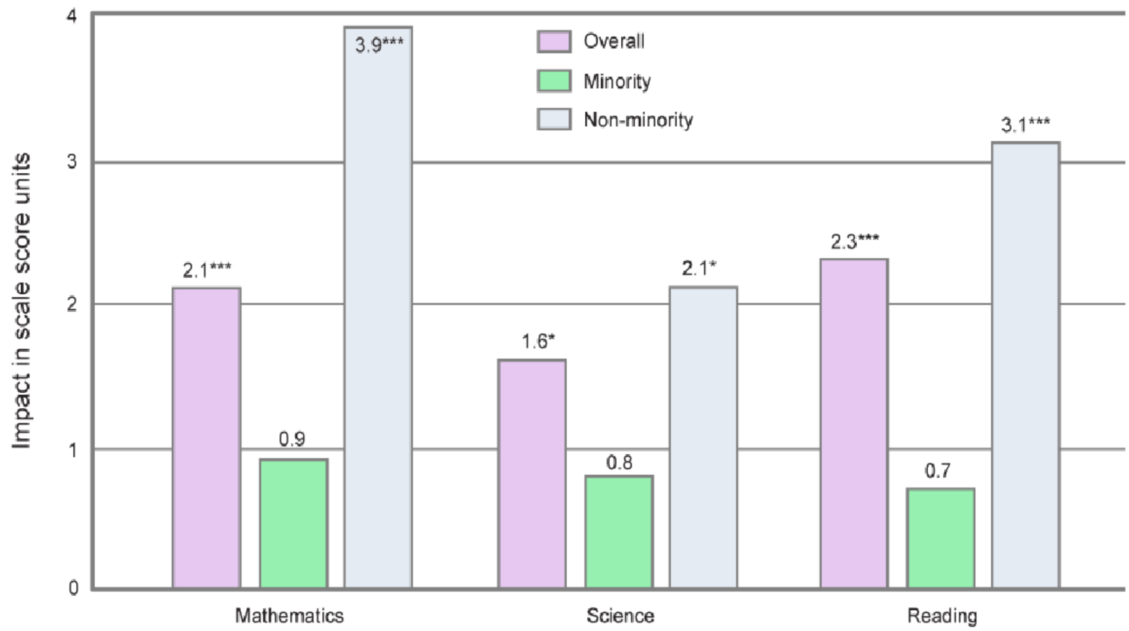

FIGURE 5. RESULTS FROM THE STUDY OF AMSTI PUBLISHED IN 2012: SUBGROUP ANALYSIS SHOWING MODERATION OF IMPACT BY MINORITY STATUS

*** $p < .01$

** $.01 \le p < .05$

* $.05 \le p < .10$

Black and White Students Math Achievement

In the current study, we found that AMSTI benefitted both White and Black students on their math achievement.[1] Figure 6 shows that the positive effects are the same size for White and Black students in math, and are equivalent to a 2-percentile point gain in math achievement over the comparison group (non-AMSTI classes).

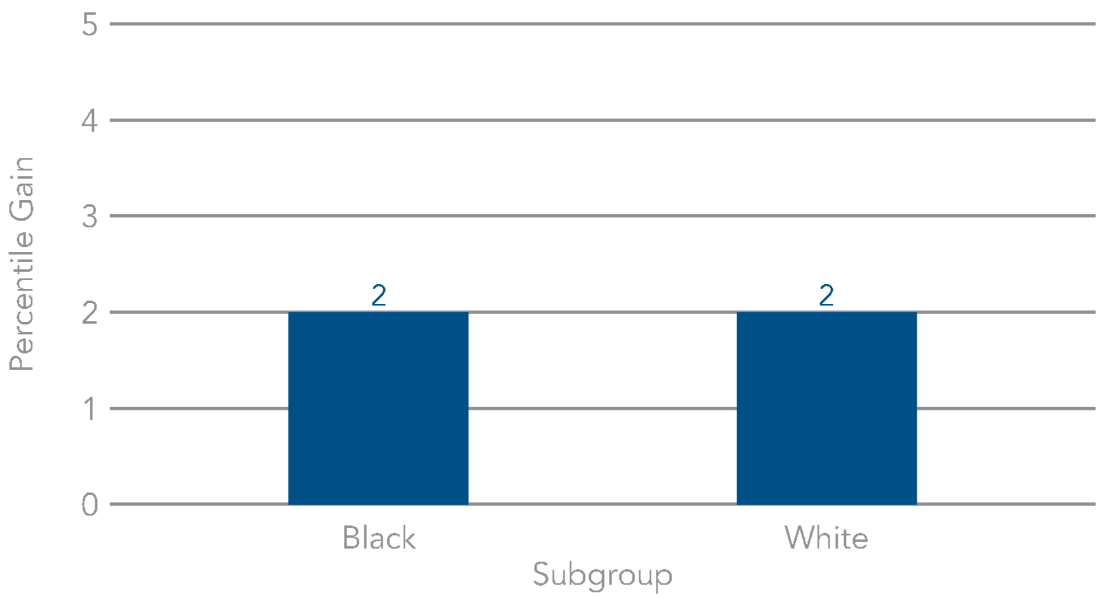

FIGURE 6. IMPACT OF AMSTI FOR BLACK AND WHITE STUDENTS ON THE ACT ASPIRE MATH ASSESSMENT

[1] We were not able to detect an impact of AMSTI for either White or Black students on their science achievement. We found a positive impact of AMSTI for White students on reading achievement, but could not detect an impact of AMSTI for Black students.

Males and Females

We also looked at the effect of AMSTI separately for males and females. We found that AMSTI impacts both genders favorably in math, science, and reading achievement (Figure 7). Interestingly, the effect in math and reading was larger for males, and in science, was larger for females. The effect of AMSTI on science achievement for females is equivalent to a 6-percentile point gain over the comparison group.

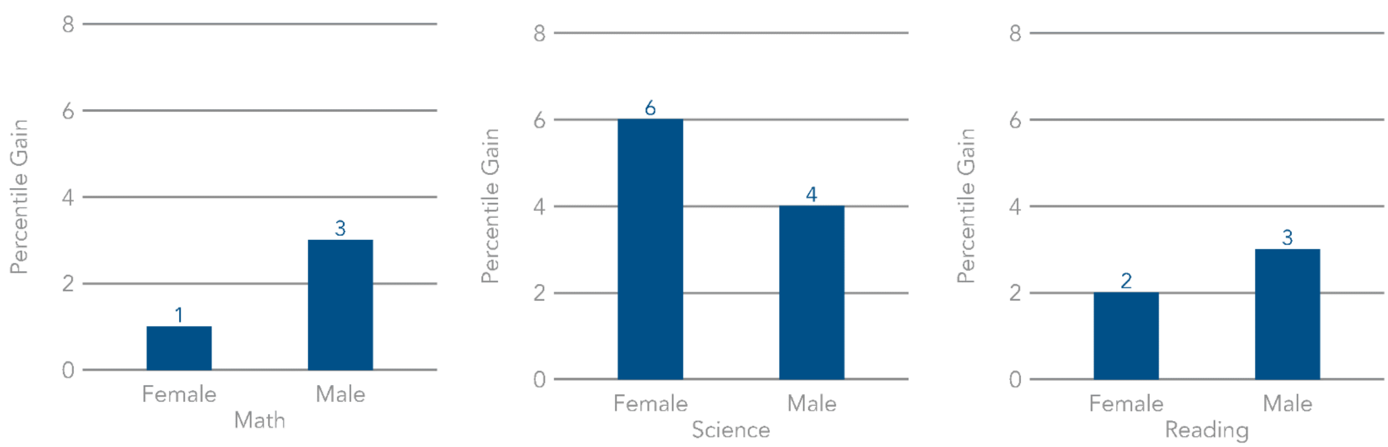

FIGURE 7. IMPACT OF AMSTI FOR FEMALE AND MALE STUDENTS ON THE ACT ASPIRE MATH, SCIENCE AND READING ASSESSMENTS

Economically Disadvantaged Students

We also found that AMSTI benefitted students with lower socio-economic status, as measured by eligibility for the free- or reduced-price lunch program. Figure 8 shows that the impact of AMSTI for students eligible for free- or reduced- price lunch is equivalent to a 3-percentile point gain in math achievement and a 4-percentile point gain in science and reading achievement.

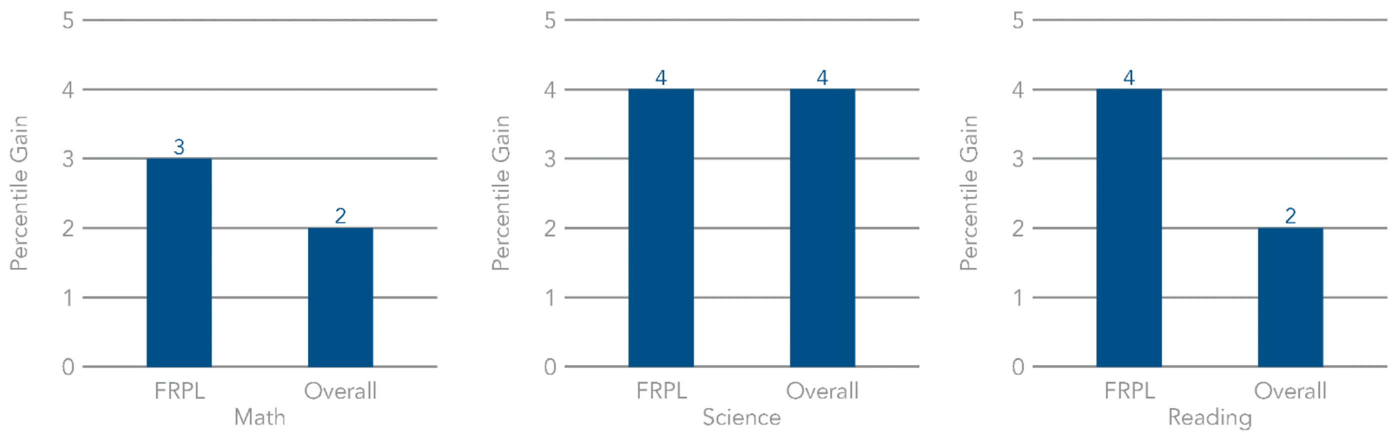

FIGURE 8. IMPACT OF AMSTI FOR STUDENTS ELIGIBLE FOR FREE/REDUCED PRICE LUNCH PROGRAM ON THE ACT ASPIRE MATH, SCIENCE, AND READING ASSESSMENT

Special Education Students

In Figure 9, we see that students in special education taught by fully-trained AMSTI teachers scored higher in math than special education students taught by non-AMSTI trained teachers. There was no effect of AMSTI for special education students on science or reading achievement.

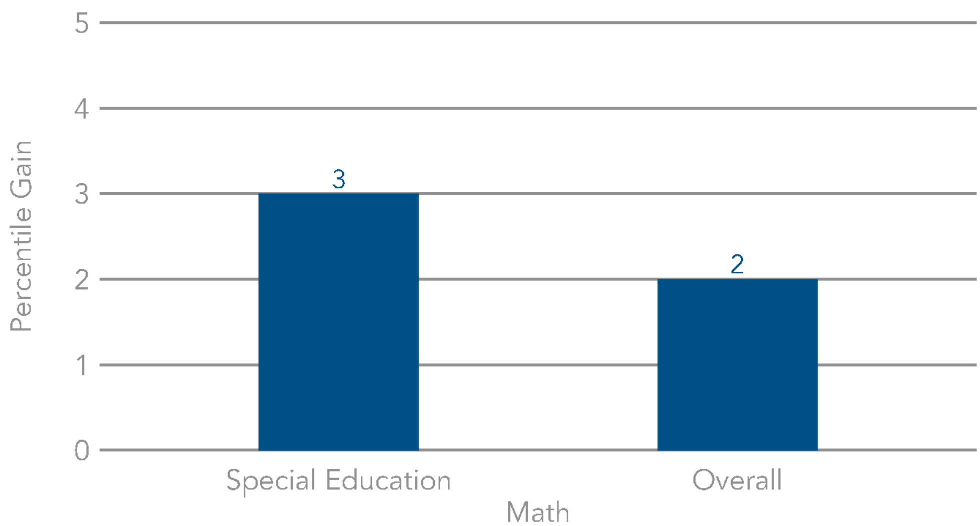

FIGURE 9. IMPACT OF AMSTI FOR SPECIAL EDUCATION STUDENTS ON ACT ASPIRE MATH ASSESSMENT

English Language Learners

We did find one subgroup that did not benefit from AMSTI. Figure 10 shows that in science, ELLs were at a serious disadvantage with AMSTI trained teachers. There was no detectable impact of AMSTI on math or reading achievement for ELLs.

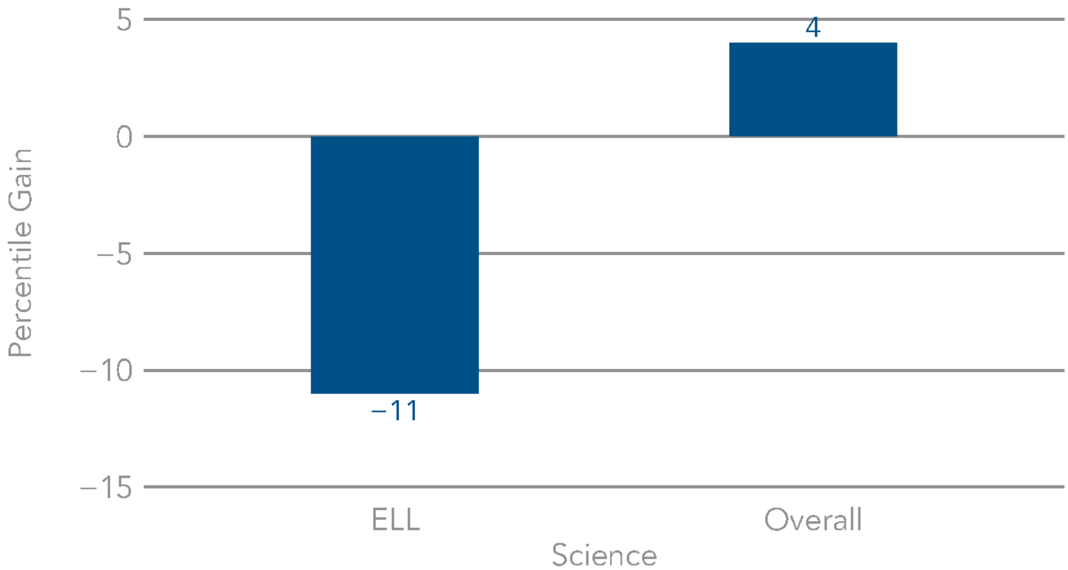

FIGURE 10. IMPACT OF AMSTI FOR ENGLISH LANGUAGE LEARNERS ON ACT ASPIRE SCIENCE ASSESSMENT

We continued to investigate this finding to identify specific schools where ELLs benefited from AMSTI. From the sample of 59 schools used in the science ELL subgroup analysis, we identified seven schools where there is a significant school-level positive effect on ELL students in AMSTI classes. We refer to these schools as "Model AMSTI science schools for ELLs" and shared this information with the AMSTI team (we do not identify the seven schools in this report).[2]

IMPACT ON EARLY CAREER TEACHERS

Based on our initial findings, ALSDE asked if the impact of AMSTI varies based on teachers' years of experience. For example, do early career teachers benefit more from participating in AMSTI training? The AMSTI team was also

[2] For each school, we estimated the school-level effect on ELLs (regardless of their participation in AMSTI) by comparing ELL performance within the school to ELL performance in the other schools in the sample. We also estimated the impact of AMSTI science on ELLs in each school. Schools were identified as "Model AMSTI science schools for ELLs" if the school-level impact of AMSTI science on ELLs was positive and at least marginally statistically significant (at $p < .20$). Additionally, if the school-level effect for all ELLs was negative and statistically significant, the positive impact of AMSTI for ELLs should outweigh the negative effect for ELLs in the school. Additional technical details of this analysis were provided directly to the AMSTI team.

interested in the impact of AMSTI for first year teachers specifically because they want to understand how participation in AMSTI pre-service certification training impacts student achievement. Therefore, we conducted a separate analysis to investigate the impact of AMSTI for early career teachers. The major limitation in this work was the unreliability of data on teachers' years of experience. These data are provided manually to ALSDE by individual districts and exhibited a disproportionate number of teachers with 0 years of experience. As an attempt at correction, we requested and obtained a list of all teachers who showed up in the ALSDE database for the first time in the study school year 2016/17. We initially sought to utilize first year teachers who had received AMSTI pre-service training during their degree, but the sample of teachers was insufficient. After inspection of AMSTI pre-service and compressed training records, along with the revised list of first year teachers, we identified a sample of early career teachers: that is, teachers in the revised list of first-time teachers with 0, 1, or 2 years of teaching experience, as measured by the ALSDE data, who were either partially or fully trained in AMSTI math or science and taught the same grade and subject in which they were trained. In the analysis of early career AMSTI math teachers, there were 88 classes of partially-trained AMSTI early career teachers, 22 classes of fully-trained AMSTI early career teachers, and 835 comparison classes. In the analysis of early career AMSTI science teachers, there were 142 classes of partially-trained AMSTI early career teachers, 33 classes of fully-trained AMSTI early career teachers, and 546 comparison classes.

While not directly comparable to the other results, as it utilizes two treatment indicators for fully- and partially-trained teachers, this analysis found similar positive impacts on both math and science, albeit with less reliability due to the small sample size. Students of early career teachers with partial AMSTI training in math had an impact equivalent to a 5-percentile-point gain on the ACT Aspire math assessment. Students of early career teachers who were fully-trained in AMSTI science had an impact equivalent to a 5-percentile-point impact on the ACT science assessment. Impacts for students of early career fully-trained AMSTI math teachers and early career partially-trained AMSTI science teachers were positive but did not reach levels of statistical significance. These results are promising; however, given the sample size and assumptions about the data, more research is needed to draw further conclusions.

Areas for Program Improvement and Follow-on Research

OPPORTUNITIES FOR FURTHER IMPROVEMENT

This study demonstrates the promising effect of AMSTI as a statewide initiative, in terms of its overall impact on math, reading, and science achievement, as well as its benefit for student subgroups that are traditionally underserved. There are, however, areas where the program can improve. These areas can be supported by additional analysis of the existing data.

Improving Outcomes for Females in Computer Science

AMSTI had a strong benefit in science for females. The program is currently piloting units to help teachers address the standards in the digital literacy and computer science course of study. The AMSTI team should consider how the components of the program that may be more engaging for girls in science can be utilized with these new units. This could lead to more females becoming engaged, and continuing their education and career in STEM and other technical fields where they are traditionally underrepresented.

Improving Outcomes for English Learners

The AMSTI and ALSDE team had a few hypotheses for why there was a negative impact of AMSTI on science achievement for ELLs. First, since the program is focused on inquiry-based learning, experimentation, and actually "doing science", it is possible that there is not enough focus on academic language that may be difficult for ELLs. Other program developers and researchers have grappled with the challenge of teaching increasingly complex concepts and language intensive practices that are called for by new science standards. These challenges are intensified for students who are not yet proficient in English (Estrella, Au, Jaeggi, & Collins, 2018; Garza, Huerta, Lara-Alecio, Irby, & Tong, 2017). A second hypothesis from the AMSTI and ALSDE team is that because the program encourages student collaboration, ELLs may feel more isolated in these collaborative environments if they struggle with language. Given the finding of the negative impact of AMSTI on science achievement for ELLs, the AMSTI team should consider how the language in the AMSTI science materials can be modified and improved to be more accessible, and what additional training for science teachers can be provided with a focus on engaging ELLs.

Improving the Achievement in Urban and Rural Locales

In math especially, economically disadvantaged students (those enrolled in the free/reduced price lunch program) benefited from AMSTI. Since these students are found predominantly in urban centers and rural areas, their success may provide an indicator of the kind of program that can continue to flourish in these locales that often fall behind academically.

FOLLOW-ON RESEARCH

There are several additional areas of investigation that can help focus improvement efforts.

Finding the Optimal Amount of Training

AMSTI has followed best practices in terms of offering 14-days of grade and content specific professional development across two consecutive summers (8 days in year one and 6 days in year two). Follow-on research could be conducted to identify how much training is optimal to improve classroom practices and, in turn, impact student achievement.

Identifying the Most Successful Schools

Another follow-on study can employ a mixed methods approach. With the data collected from this study, we can identify specific schools that were most successful with subgroups of interest. For example, we identified seven "Model AMSTI science schools for ELLs" and shared that information with ALSDE. We recommend that the ALSDE or AMSTI teams collect more qualitative data to understand the characteristics of these schools and how teachers are better supporting ELLs. They may explore characteristics, such as the type of ELL program being used (e.g., pull out programs for ELLs in science vs. programs where ELLs are integrated into general education science classes), presence of an ELL coordinator, or additional literacy supports in place for ELLs. We also found that these seven schools are not statistically different from the rest of the group of schools in the sample on known school characteristics, such as class size and student demographics; this further justifies looking for other factors responsible for better performance.

Annual Analysis

Finally, as the Alabama Course of Study continues to evolve and the state rolls out a new assessment, the impact of AMSTI should be measured each year. As more science teachers participate in the re-designed AMSTI science training, the larger sample would allow for a more precise analysis. Annual data collection would also allow for a more robust analysis of the impact of AMSTI on early career teachers.

References

Garza, T., Huerta, M., Lara-Alecio, R., Irby, B. J., & Tong, F. (April 2017). Pedagogical Differences during a Science and Language Intervention for English Language Learners. *The Journal of Educational Research, 111*(4), 487–496. Retrieved December 27, 2018 from https://www.tandfonline.com/doi/full/10.1080/00220671.2017.1302913

Estrella, G., Au, J., Jaeggi, S. M., & Collins, P. (2018). Is Inquiry Science Instruction Effective for English Language Learners? A Meta-Analytic Review. *AERA Open, 4*(2), 1-33. Retrieved December 27, 2018 from https://files.eric.ed.gov/fulltext/EJ1194148.pdf

Lazarev, V., Schellinger, A., Zacamy, J., & Newman, D. (2018). *Efficacy of the Alabama Math, Science, Technology Initiative (AMSTI). A Report of a Quasi-experiment in Alabama*. (Empirical Education report number: Empirical_ALSDE-7037-FR1-2018-O.1). Palo Alto, CA: Empirical Education Inc.

Newman, D., Finney, P.B., Bell, S., Turner, H., Jaciw, A.P., Zacamy, J.L., & Feagans Gould, L. (2012). *Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI). (NCEE 2012–4008)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

What Works Clearinghouse. (2018). *Standards handbook (Version 4.0)*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Appendix A: Technical Details of the Methodology Used

This report examines the effect of AMSTI teacher training on student performance on state assessments of math, science, and reading.

The study used data from the 2016/17 school year, as measured by the ACT Aspire math, science, and reading tests. Test results were available at the class level, with both percent proficient and mean scale score for the class and for subgroups used in this analysis. The test and pretest scores were normalized to establish comparability across grades.

For the main analysis, we compared (1) classes taught by teachers considered fully trained in all required AMSTI units for the grade and subject that they teach to (2) classes taught by teachers with no known AMSTI training. Classes of teachers who were trained in only some of the AMSTI units or trained to teach other grades/subjects were removed from the sample. The sample of classes used in this analysis were those in schools with both fully-trained AMSTI teachers and non-AMSTI teachers. By using classes taken from the same school, we were able to control for school characteristics thereby minimizing bias due to school selection into AMSTI.

Table A1, A2 and A3 show the baseline equivalence of the AMSTI and comparison classes. All differences between AMSTI and comparison classes are less than .25 pooled standard deviation, which is acceptable for a quasi-experimental comparison group study per What Works Clearinghouse guidelines (What Works Clearinghouse, 2018).

TABLE A1. BASELINE EQUIVALENCE OF THE AMSTI AND COMPARISON SAMPLES USED IN MATH ANAYSIS

Characteristic	AMSTI	Comparison	Pooled standard deviation	Difference as proportion of standard deviation
Class size (total number of students)	47.6	46.4	31.4	0.04
Pretest (from 2015/16 ACT Aspire)	0.09	-0.08	1.1	0.15
% Male	50.8	50.9	9.3	0.02
% Female	49.2	49.1	9.3	0.02
% White	55.4	50.0	32.9	0.16
% Black	33.0	38.4	34.1	0.16
% Hispanic	7.60	7.50	10.1	0.01
% Native American	0.62	0.57	2.4	0.02
% Economically disadvantaged	54.6	55.9	23.5	0.06
% English Language Learners	1.87	1.83	3.8	0.01
% Special education	12.7	12.1	11.5	0.06

TABLE A2. BASELINE EQUIVALENCE OF AMSTI AND COMPARISON SAMPLES USED IN SCIENCE ANAYSIS

Characteristic	AMSTI	Comparison	Pooled standard deviation	Difference as proportion of standard deviation
Class size (total number of students)	54.2	55.1	37.04	0.02
Pretest (from 2015/16 ACT Aspire)	0.19	-0.02	0.94	0.22
% Male	50.5	50.6	8.6	0.02
% Female	49.5	49.4	8.6	0.02
% White	59.5	53.6	30.18	0.19
% Black	26.5	33.4	31.46	0.22
% Hispanic	8.78	7.79	11.17	0.09
% Native American	1.23	1.43	4.2	0.05
% Economically disadvantaged	52.5	53.7	22.08	0.05
% English Language Learners	1.88	1.70	3.73	0.05
% Special education	11.6	11.9	9.48	0.03

TABLE A3. BASELINE EQUIVALENCE OF THE AMSTI AND COMPARISON SAMPLES USED IN READING ANAYSIS

Characteristic	AMSTI	Comparison	Pooled standard deviation	Difference as proportion of standard deviation
Class size (total number of students)	42.9	36.2	29.3	0.23
Pretest (from 2015/16 ACT Aspire)	0.35	0.14	1	0.21
% Male	50.8	51.4	13.2	0.05
% Female	49.2	48.6	13.2	0.05
% White	61.6	58.2	27.7	0.13
% Black	22.8	26.6	26.6	0.14
% Economically disadvantaged	49.7	52.6	23.4	0.13
% English Language Learners	2.6	3.5	6.6	0.13
% Special education	13.4	12.5	11.9	0.08

ANALYSIS

The estimate of AMSTI's impact was assessed by comparing scale scores for AMSTI classes and comparison classes, using a linear regression adjustment for differences in student demographics, school characteristics, and pretest scores. While the percent proficient was also available, we chose not to use it since it was less sensitive to subgroups that were on average well below or above the proficiency cut point.

An additional sensitivity analyses was conducted using a Bayesian Markov Chain Monte Carlo estimation with stratified propensity score matching, which encompasses a larger sample and confirmed the main findings. This analysis is not published as part of this report. It confirmed the ELL analysis but was not done for the other subgroups.

Appendix B. Full Results of Impact of AMSTI

Table B1, B2 and B3 present the full results of the overall and subgroup analysis, including the effect size, standard error, *p* value, percentile gain, and sample size. Below the tables we provide a description of the various metrics.

TABLE B1. IMPACT OF AMSTI TRAINING ON STUDENT ACHIEVEMENT IN MATH

Demographic	Subgroup	Effect size	Standard error	*p* value	% gain	No. of School Systems	No. of Schools	No. of AMSTI Classes	No. of Non-AMSTI Classes
Overall	All subgroups combined	0.05	0.02	.01	2	76	256	607	890
Gender	Males	0.07	0.03	<.01	3	76	256	607	888
	Females	0.04	0.03	.16	1	76	256	606	890
Race/ethnicity	White	0.05	0.03	.11	2	74	237	553	767
	Black	0.05	0.04	.17	2	74	248	531	801
Socio-economic status	Free/reduced-price lunch eligible	0.07	0.03	<.01	3	76	256	605	889
English Language Learner status	English Language Learners	No detectable impact	0.11	.77	N/A	50	148	185	254
Special education status	Special education students	0.08	0.05	.09	3	76	254	501	694

Note. We did not calculate percentile gain for subgroups with no detectible impact.

Source. Empirical Education staff calculations

TABLE B2. IMPACT OF AMSTI TRAINING ON STUDENT ACHIEVEMENT IN SCIENCE

Demographic	Subgroup	Effect size	Standard error	*p* value	% gain	No. of School Systems	No. of Schools	No. of AMSTI Classes	No. of Non-AMSTI Classes
Overall	All subgroups combined	0.11	0.04	.01	4	53	112	200	238
Gender	Males	0.09	0.05	.07	4	53	112	200	237
	Females	0.15	0.05	<.01	6	53	112	199	238
Race/ethnicity	White	No detectable impact	0.06	.29	N/A	49	105	187	210
	Black	No detectable impact	0.08	.94	N/A	50	104	175	209
Socio-economic status	Eligible for free/reduced-price lunch	0.12	0.05	.04	4	53	112	199	238
English Language Learner status	English Language Learners	-0.31	0.18	.12	-11	30	59	61	78
Special education status	Special education students	No detectable impact	0.10	.66	N/A	53	111	175	199

Note. We did not calculate percentile gain for subgroups with no detectible impact.

Source. Empirical Education staff calculations

TABLE B3. IMPACT OF AMSTI TRAINING ON STUDENT ACHIEVEMENT IN READING

Demographic	Subgroup	Effect size	Standard error	p value	% gain	No. of school systems	No. of schools	No. of AMSTI classes	No. of non-AMSTI classes
Overall	All subgroups combined	0.06	0.03	.03	2	49	95	267	690
Gender	Males	0.07	0.03	.01	3	49	95	265	676
	Females	0.05	0.03	.11	2	49	95	267	672
Race/ethnicity	White	0.05	0.04	.14	2	46	90	254	643
	Black	No detectable impact	0.06	.29	N/A	48	93	238	597
Socio-economic status	Free/reduced-price lunch eligible	0.09	0.04	.01	4	49	95	266	678
English Language Learner status	English Language Learners	No detectable impact	0.09	.57	N/A	31	58	83	201
Special education status	Special education students	No detectable impact	0.06	.37	N/A	49	94	230	528

Note. We did not calculate percentile gain for subgroups with no detectible impact.

Source. Empirical Education staff calculations

When we run the computations on the data, we produce several results: among them are the effect size and *p* values.

EFFECT SIZES

We translate the difference between program and control groups into a standardized effect size by dividing the average group difference by a measure of the variability in the outcome. This measure of variability is also called the standard deviation and can be thought of as the average distance of all the individual scores from the average score (more precisely, it is the square root of the average of squared distances). Dividing the difference by the standard deviation gives us a measure of the impact in units of standard deviation, rather than units of the scale used by the particular test. This standardized effect size allows us to compare the results we find with results from other studies that use different measurement scales. Effect sizes can also be translated into percentile gains (a more readily interpretable metric), which are shown in the bar graphs in the main body of the report.

P VALUES

The *p* value is very important, because it gives us a gauge of how confident we can be that the result we are seeing is not due simply to chance. Specifically, it tells us what the probability is that we would obtain a result with a magnitude as large as—or larger than—the magnitude of the one observed when in fact there is no effect. Roughly speaking, it tells us the risk of concluding that the program has had an effect when in fact it hasn't. This mistake is also known as a false-positive conclusion. Thus, a *p* value of .1 gives us a 10% probability of drawing a false-positive conclusion if in fact there is no impact of the program.

We can also think of the *p* value as the level of confidence, or the level of belief we have that the outcome we observe is not simply due to chance. While ultimately depending on the risk tolerance of the user of the research, we suggest the following guidelines for interpreting *p* values.

1. We have strong confidence when $p \leq .05$. (This is the level of confidence conventionally referred to as statistical significance.)
2. We have moderate confidence when $.05 < p \leq .15$.
3. While still providing some evidence of promise, we have limited confidence when $.15 < p \leq .20$.
4. We have no confidence when $p > .20$.

In reporting results with *p* values higher than conventional statistical significance, our goal is to inform the local decision makers with useful information and provide other researchers with data points that can be synthesized into more general evidence.