NCEE 2012-4008

U. S. DEPARTMENT OF EDUCATION

Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI)

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI)

Final Report

February 2012

Authors:

Denis Newman, Empirical Education Inc., Co-Principal Investigator

Pamela B. Finney, Research Management Leader, Regional Educational Laboratory Southeast, Co-Principal Investigator

Steve Bell, Abt Associates Inc., Task 2 Methodology Leader

Herb Turner, ANALYTICA, Technical Consultant

Andrew P. Jaciw, Empirical Education Inc., Director of Experimental Design and Analysis

Jenna L. Zacamy, Empirical Education Inc., Senior Research Manager

Laura Feagans Gould, Academy for Educational Development, Senior Research Scientist

Project Officer:
Sandra Garcia, Institute of Education Sciences

NCEE 2012-4008

Richard Sawyer, Academy for Educational Development, was the original co-principal investigator.

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
Rebecca A. Maynard
Commissioner

February 2012

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract ED–06-CO–0028.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the report.

This report is available on the Institute of Education Sciences website at http://ncee.ed.gov and the Regional Educational Laboratory Program website at http://edlabs.ed.gov.

Alternate Formats Upon request, this report is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department’s Alternate Format Center at 202-260-9895 or 202-205-8113.

Disclosure of potential conflict of interest

None of the authors or other staff involved in the study from Abt Associates, Academy for Educational Development, ANALYTICA, Empirical Education Inc., or the Regional Educational Laboratory-Southeast at SERVE Center at the University of North Carolina at Greensboro has financial interests that could be affected by the content of this report.[1] No one on the nine-member Technical Working Group, convened twice annually by the research team to provide advice and guidance, has financial interests that could be affected by the study findings.

[1] Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

Acknowledgments

This study was the collaborative effort of a team that included professional staff at the Regional Educational Laboratory-Southeast at SERVE Center at the University of North Carolina at Greensboro; Empirical Education Inc.; Abt Associates Inc.; and the Academy for Educational Development (AED). From the Regional Educational Laboratory-Southeast (SERVE), the authors acknowledge Jean Scott, senior education policy research analyst, for assistance in maintaining a strong relationship with the Alabama State Department of Education (ALSDE); David Schaefer, technical editor and advisor; Melissa Williams, evaluation specialist, and De Fuller, administrative support.

From Empirical Education Inc., the authors acknowledge Laurel Sterling, Margit Zsolnay, and Ruthie Chang for their contributions to data collection, project management, and report writing, and Boya Ma, Qingfeng (Breeze) Zhao, Akiko Lipton, and Xiaohui Zheng, for planning and conducting the statistical analyses.

From Abt, the authors acknowledge Rob Olsen, for his valuable advice on statistical modeling, and Cay Bradley, for her collaboration on the analysis of the two-year impact.

From AED, the authors acknowledge Harold Wenglensky, Richard Sawyer, Amy Detgen, Robert Bozick, and Jason Snipes, for their leadership in earlier phases of the project.

Finally, the authors express their deep gratitude to the people at ALSDE and administrators and teachers in the participating school systems for their assistance and cooperation in conducting this research.

Contents

DISCLOSURE OF POTENTIAL CONFLICT OF INTEREST III

ACKNOWLEDGMENTS IV

SUMMARY XX

1: INTRODUCTION AND STUDY OVERVIEW 1

Strengthening skills of mathematics and science teachers nationwide 1

The Alabama Math, Science, and Technology Initiative (AMSTI): A state-level strategy to strengthen skills of mathematics and science teachers 2

Alabama Math, Science, and Technology Initiative (AMSTI) theory of action 3

Alabama Math, Science, and Technology Initiative (AMSTI) program components and elements 5

Prior research on the effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement 6

Overview of study design 7

Research questions 9

Structure of the report 12

2: STUDY DESIGN AND METHODOLOGY 13

Rationale for experimental design 13

Recruitment, selection, and random assignment of schools 14

Incentives to participate 20

Defining the term "Alabama Math, Science, and Technology Initiative (AMSTI) teacher" 20

Data sources, collection, and cleaning 21

Threats to internal validity 30

Data analysis methods 35

3: ANALYSES OF ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) IMPLEMENTATION 65

Alabama Math, Science, and Technology Initiative (AMSTI) implementation components 65

Alabama Math, Science, and Technology Initiative (AMSTI) implementation results 66

Conclusions 79

4: EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON MATHEMATICS PROBLEM SOLVING ACHIEVEMENT, SCIENCE ACHIEVEMENT, AND CLASSROOM INSTRUCTIONAL STRATEGIES AFTER ONE YEAR 81

Effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement after one year 81

Effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on classroom practices after one year 86

5: EFFECT ON MATHEMATICS PROBLEM SOLVING AND SCIENCE ACHIEVEMENT AFTER TWO YEARS 91

Effect on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement 92

Effect on Stanford Achievement Test Tenth Edition (SAT 10) science achievement 94

6: EFFECTS OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON READING ACHIEVEMENT, TEACHER CONTENT KNOWLEDGE, STUDENT ENGAGEMENT, AND VARIATIONS IN EFFECTS FOR STUDENT SUBGROUPS AFTER ONE YEAR 96

Effects on student achievement in reading 97

Effects on teacher content knowledge and student engagement 99

Subgroup differences in average effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement after one year 103

7: SUMMARY OF FINDINGS AND STUDY LIMITATIONS 106

Study design..106

Summary of confirmatory findings 106

Summary of exploratory findings..107

Study limitations 109

APPENDIX A. EXPLANATION OF PRIMARY AND SECONDARY CONFIRMATORY OUTCOME MEASURES.............. A-1

APPENDIX B. EXPLANATION OF EXPLORATORY RESEARCH QUESTIONS B-1

Effects of two years of exposure to the Alabama Math, Science, and Technology Initiative (AMSTI) B-1

Effect on student achievement in reading... B-1

Effect on teacher content knowledge B-2

Effect on student engagement B-3

Effects on subgroups of students... B-3

APPENDIX C. SELECTION AND RANDOM ASSIGNMENT OF SCHOOLS C-1

Subexperiments C-1

Limiting the pool of schools for selection C-1

Selecting and assigning schools C-1

The real-time and interactive nature of pair selection C-3

APPENDIX D. STATISTICAL POWER ANALYSIS...D-1

APPENDIX E. DATA COLLECTION PROCEDURES AND TIMELINE E-1

APPENDIX F. DESCRIPTION OF PROGRAM IMPLEMENTATION DATA COLLECTED BUT NOT USED IN REPORT......F-1

Classroom observations F-1
Professional development observations F-3
Principal surveys F-4

APPENDIX G. ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) TEACHER SURVEY #3 .... G-1

APPENDIX H. DATA CLEANING AND DATA FILE CONSTRUCTION H-1

Student-level data H-1
Program implementation data H-1

APPENDIX I. ATTRITION THROUGH STUDY STAGES FOR SAMPLES USED IN THE CONFIRMATORY ANALYSIS I-1

Selection of the Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome sample I-1
Selection of the Stanford Achievement Test Tenth Edition (SAT 10) I-3
science outcome sample I-3
Selection of the sample for the active learning score outcome I-6

APPENDIX J. DESCRIPTION OF DEGREE RANK J-1

APPENDIX K. EQUIVALENCE OF YEAR 1 BASELINE AND ANALYZED SAMPLES FOR CONFIRMATORY STUDENT-LEVEL AND CLASSROOM PRACTICE OUTCOMES K-1

APPENDIX L. INTERNAL CONSISTENCY AND VALIDITY OF ACTIVE LEARNING MEASURES L-1

APPENDIX M. NUMBER OF STUDENTS AND TEACHERS IN SCHOOLS IN ANALYTIC SAMPLES USED TO ANALYZE YEAR 1 CONFIRMATORY QUESTIONS M-1

APPENDIX N. ATTRITION THROUGH STUDY STAGES FOR SAMPLES USED IN YEAR 1 EXPLORATORY ANALYSIS N-1

Stanford Achievement Test Tenth Edition (SAT 10) reading outcome N-1
Level of teacher content knowledge in mathematics N-3
Level of student engagement in mathematics N-6
Level of student engagement in science attrition N-7

APPENDIX O. TESTS OF EQUIVALENCE FOR BASELINE AND ANALYTIC SAMPLES FOR YEAR 1 EXPLORATORY OUTCOMES O-1

APPENDIX P. STATISTICAL POWER ANALYSES FOR MODERATOR ANALYSES P-1

APPENDIX Q. DERIVATION AND MOTIVATION OF THE BELL-BRADLEY ESTIMATOR WHEN MEASURING ESTIMATED TWO-YEAR EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) Q-1

APPENDIX R. ATTRITION THROUGH STUDY STAGES FOR SAMPLES CONTRIBUTING TO ESTIMATION OF TWO-YEAR EFFECTS R-1

SELECTION OF THE STANFORD ACHIEVEMENT TEST TENTH EDITION (SAT 10) MATHEMATICS PROBLEM SOLVING SAMPLE R-1

SELECTION OF THE STANFORD ACHIEVEMENT TEST TENTH EDITION (SAT 10) SCIENCE SAMPLE R-5

APPENDIX S. EXAMINATION OF EQUIVALENCE IN BASELINE AND ANALYTIC SAMPLES USED IN THE ESTIMATION OF TWO-YEAR EFFECTS S-1

APPENDIX T. ESTIMATION MODEL FOR TWO-YEAR EFFECTS OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) T-1

APPENDIX U. TOPICS AND INSTRUCTIONAL METHODS USED AT THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) SUMMER INSTITUTE U-1

APPENDIX V. PARAMETER ESTIMATES ON PROBABILITY SCALE FOR ODDS-RATIO TESTS OF DIFFERENCES BETWEEN ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) AND CONTROL CONDITIONS IN YEAR 1 (ASSOCIATED WITH SUMMER PROFESSIONAL DEVELOPMENT AND IN-SCHOOL SUPPORT OUTCOMES) V-1

APPENDIX W. DESCRIPTIVE STATISTICS FOR VARIABLES THAT CHANGE TO A BINARY SCALE USED IN THE W-1

APPENDIX X. COMPARISON OF ASSUMED PARAMETER VALUES AND OBSERVED SAMPLE STATISTICS FOR STATISTICAL POWER ANALYSIS AFTER ONE YEAR X-1

APPENDIX Y. PARAMETER ESTIMATES FOR STANFORD ACHIEVEMENT TEST TENTH EDITION (SAT 10) MATHEMATICS PROBLEM SOLVING AFTER ONE YEAR Y-1

APPENDIX Z. PARAMETER ESTIMATES FOR STANFORD ACHIEVEMENT TEST TENTH EDITION (SAT 10) SCIENCE AFTER ONE YEAR Z-1

APPENDIX AA. PARAMETER ESTIMATES FOR ACTIVE LEARNING IN MATHEMATICS AFTER ONE YEAR AA-1

APPENDIX AB. PARAMETER ESTIMATES FOR ACTIVE LEARNING IN SCIENCE AFTER ONE YEAR AB-1

APPENDIX AC. SENSITIVITY ANALYSES OF EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STANFORD ACHIEVEMENT TEST TENTH EDITION (SAT 10) MATHEMATICS PROBLEM SOLVING ACHIEVEMENT AFTER ONE YEAR AC-1

APPENDIX AD. SENSITIVITY ANALYSES OF EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STANFORD ACHIEVEMENT TEST TENTH EDITION (SAT 10) SCIENCE ACHIEVEMENT AFTER ONE YEAR AD-1

APPENDIX AE. SENSITIVITY ANALYSES OF EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON ACTIVE LEARNING INSTRUCTIONAL STRATEGIES IN MATHEMATICS CLASSROOMS AFTER ONE YEAR AE-1

APPENDIX AF. SENSITIVITY ANALYSES OF EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON ACTIVE LEARNING INSTRUCTIONAL STRATEGIES IN SCIENCE CLASSROOMS AFTER ONE YEAR AF-1

APPENDIX AG. TESTS FOR VIOLATIONS OF FACTORS ASSOCIATED WITH ASSUMPTION OF EQUAL FIRST YEAR EFFECTS ON STUDENTS IN ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) AND CONTROL SCHOOLS AG-1

Sponsor's guidelines for the parameters of the intervention AG-2
District support AG-3
Composition of schools that did not implement Alabama Math, Science, and Technology Initiative (AMSTI) AG-3
Characteristics of teachers who did not participate in Alabama Math, Science, and Technology Initiative (AMSTI) AG-9
Efforts by schools to implement Alabama Math, Science, and Technology Initiative (AMSTI) AG-9
Characteristics of student cohorts AG-12

APPENDIX AH. POST HOC ADJUSTMENT TO STANDARD ERROR FOR ESTIMATE OF TWO-YEAR EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON MATHEMATICS ACHIEVEMENT AFTER TWO YEARS AH-1

APPENDIX AI. PARAMETER ESTIMATES FOR EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) AFTER TWO YEARS AI-1

APPENDIX AJ. PARAMETER ESTIMATES FOR EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STUDENT READING ACHIEVEMENT AFTER ONE YEAR AJ-1

APPENDIX AK. PARAMETER ESTIMATES FOR TEACHER CONTENT AND STUDENT ENGAGEMENT AFTER ONE YEAR AK-1

APPENDIX AL. ESTIMATES OF EFFECTS FOR TERMS INVOLVING THE INDICATOR OF TREATMENT STATUS IN THE ANALYSIS OF THE MODERATING EFFECT OF THE THREE-LEVEL PRETEST VARIABLE AL-1

APPENDIX AM. PARAMETER ESTIMATES FOR THE ANALYSIS OF THE MODERATING EFFECT OF RACIAL/ETHNIC MINORITY STATUS ON THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON READING AFTER ONE YEAR AM-1

APPENDIX AN. PARAMETER ESTIMATES FOR ANALYSIS OF AVERAGE EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON READING BY RACIAL/ETHNIC MINORITY STUDENTS AFTER ONE YEAR....AN-1

APPENDIX AO. PARAMETER ESTIMATES FOR EFFECT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON READING FOR WHITE STUDENTS AFTER ONE YEAR....AO-1

REFERENCES....REF-1

Figures

Figure 1 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year....xxv

Figure 2 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year....xxvi

Figure 1.1 Alabama Math, Science, and Technology Initiative (AMSTI) theory of action....4

Figure 2.1 Regions, schools, and school year of Subexperiment 1 and Subexperiment 2....16

Figure 2.2 Flow chart of school sample allocation from random assignment to inclusion in data analysis....19

Figure 3.1 Average number of days Alabama Math, Science, and Technology Initiative (AMSTI) trainers covered topics in training manual to a moderate extent or more in grade 5 mathematics....68

Figure 3.2 Average number of days the Alabama Math, Science, and Technology Initiative (AMSTI) trainers covered topics in training manual to a moderate extent or more in grade 5 science....69

Figure 3.3 Average number of days Alabama Math, Science, and Technology Initiative (AMSTI) trainers used various instructional methods more than 25 percent of the time in grade 5 mathematics and science individually....71

Figure 3.4 Average number of days Alabama Math, Science, and Technology Initiative (AMSTI) trainers used various instructional methods more than 25 percent of the time in grade 7 mathematics and science combined....71

Figure 3.5 Reported access to Alabama Math, Science, and Technology Initiative (AMSTI) materials by Alabama Math, Science, and Technology Initiative (AMSTI) teachers....72

Figure 3.6 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control mathematics teachers receiving summer professional development....73

Figure 3.7 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control science teachers receiving summer professional development....74

Figure 3.8 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control mathematics teachers reporting access to materials and manipulatives....75

Figure 3.9 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control science teachers reporting access to materials and manipulatives....76

Figure 3.10 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control mathematics teachers requesting support for instruction....77

Figure 3.11 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control mathematics teachers receiving support for instruction 77
Figure 3.12 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control science teachers requesting support for instruction 78
Figure 3.13 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control science teachers receiving support for instruction 78
Figure 4.1 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year 84
Figure 4.2 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year 86
Figure 4.3 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms after one year 88
Figure 4.4 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms after one year 90
Figure 5.1 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after two years 93
Figure 5.2 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after two years 95
Figure 6.1 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) reading achievement after one year 98
Figure D1 Record of technical settings and result of power analysis D-2

Tables

Table 1.1 Alabama Math, Science, and Technology Initiative (AMSTI) components and corresponding elements 5
Table 2.1 Regions, schools, and school year of Subexperiment 1 and Subexperiment 2 16
Table 2.2 Data collection activities 21
Table 2.3 Student achievement data analyzed in Year 1 24
Table 2.4 Year 1 teacher survey response rates 27
Table 2.5 School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome after one year 31
Table 2.6 School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) science outcome sample after one year 32
Table 2.7 School and teacher attrition associated with active learning in mathematics outcome after one year 32
Table 2.8 School and teacher attrition associated with active learning in science outcome after one year 33

Table 2.9 Distribution of teachers by number of valid responses to active learning score items in mathematics and science 39
Table 2.10 School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) reading outcome after one year 49
Table 2.11 School and teacher attrition associated with teacher content knowledge in mathematics outcome after one year 50
Table 2.12 School and teacher attrition associated with teacher content knowledge in science outcome after one year 50
Table 2.13 School and teacher attrition associated with student engagement in mathematics outcome after one year 51
Table 2.14 School and teacher attrition associated with student engagement in science outcome after one year 51
Table 2.15 Minimum detectable differential effect sizes for moderators for mathematics problem solving, science, and reading outcomes....53
Table 2.16 Factors that must remain stable over time (from Year 1 to Year 2) for the effect of the Alabama Math, Science, and Technology Initiative (AMSTI) to be the same in the Alabama Math, Science, and Technology Initiative and control group schools (in their respective first year of Alabama Math, Science, and Technology Initiative implementation) 59
Table 2.17 School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome after two years 62
Table 2.18. School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) science outcome after two years 63
Table 4.1 Sample statistics for analytic sample of grade 4–8 students used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year 82
Table 4.2 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year 83
Table 4.3 Sample statistics for analytic sample of grade 5 and 7 students used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year 85
Table 4.4 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year 85
Table 4.5 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms after one year 87

Table 4.6 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms after one year 87
Table 4.7 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms after one year 89
Table 4.8 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms after one year 89
Table 5.1 Summary statistics on analytic sample used to determine estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after two years 92
Table 5.2 Estimated effect of two years of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after two years 93
Table 5.3 Sample statistics for analytic sample used to determine estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after two years 94
Table 5.4 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after two years 94
Table 6.1 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) reading achievement after one year 97
Table 6.2 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) reading after one year 98
Table 6.3 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher content knowledge in mathematics after one year 99
Table 6.4 Teacher counts in each response category for teacher content knowledge in mathematics 99
Table 6.5 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher content knowledge in science after one year 100
Table 6.6 Teacher counts in each response category for teacher content knowledge in science 100
Table 6.7 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on student engagement in mathematics after one year 101
Table 6.8 Teacher counts in each response category for student engagement in mathematics 101
Table 6.9 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher student engagement in science after one year 102
Table 6.10 Teacher counts in each response category for student engagement in science 102
Table 6.11 Average effects and variation in effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement for subgroups of students after one year 104

Table B1 Achievement gaps on the Stanford Achievement Test Tenth Edition (SAT 10) in Alabama, 2005/06 .... B-4
Table C1 Primary matching characteristics of sites, by region C-2
Table E1 Completion dates for data collection procedures E-1
Table I1 Attrition from analytical sample associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome I-1
Table I2 Attrition from analytical sample associated with Stanford Achievement Test Tenth Edition (SAT 10) science outcome I-4
Table I3 Attrition from analytical sample associated with active learning in mathematics outcome I-6
Table I4 Attrition from analytical sample associated with active learning in science outcome I-7
Table K1 Year 1 mean baseline sample characteristics associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving and science outcomes after one year K-1
Table K2 Year 1 mean baseline sample characteristics associated with active learning outcomes after one year K-5
Table K3 Year 1 mean analytic sample characteristics associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving and science outcomes after one year K-7
Table K4 Year 1 mean analytic sample characteristics associated with active learning outcomes after one year K-11
Table L1 Correlation coefficients for instructional strategies for active learning for mathematics L-1
Table L1 Correlation coefficients for instructional strategies for active learning for science L-1
Table M1 Number of students and teachers in schools in analytic samples used to analyze Year 1 confirmatory questions M-1
Table N1 Attrition from analytical sample associated with Stanford Achievement Test Tenth Edition (SAT 10) reading outcome N-1
Table N2 Attrition from analytical sample associated with level of teacher content knowledge in mathematics outcome N-3
Table N3 Attrition from analytical sample associated with level of teacher content knowledge in science outcome N-5
Table N4. Attrition from analytical sample associated with level of student engagement in mathematics outcome N-6
Table N5. Attrition from analytical sample associated with level of student engagement in science outcome N-7
Table O1 Year 1 mean baseline sample characteristics associated with reading achievement outcome after one year O-1
Table O2 Year 1 mean baseline sample characteristics associated with level of teacher content knowledge outcomes after one year O-3
Table O3 Year 1 mean baseline sample characteristics associated with level of student engagement outcomes after one year O-5
Table O4 Year 1 mean analytic sample characteristics associated with reading achievement outcome after one year O-7
Table O5 Year 1 mean analytic sample characteristics associated with level of teacher content knowledge outcomes after one year O-9

Table O6 Year 1 mean analytic sample characteristics associated with level of student engagement outcomes after one year O-11

Table P1 Specification of covariates used in moderator analyses for mathematics problem solving outcome P-3

Table P2 Sample-based assumptions used in estimating power for moderator analyses of mathematics problem solving outcome P-5

Table P3 Sample-based assumptions used in estimating power for moderator analyses of science outcome P-6

Table R1 Attrition from Year 1 and Year 2 analytical samples contributing to estimation of two-year effect (associated with Stanford Achievement Test Tenth Edition [SAT 10)] mathematics problem solving outcome) R-1

Table R2 Attrition from Year 1 and Year 2 analytical samples contributing to estimation of two-year effect (associated with Stanford Achievement Test Tenth Edition [SAT 10)] science outcome) R-6

Table S1 Mean characteristics of baseline sample contributing to the Year 2 component of the Bell-Bradley estimate (associated with Stanford Achievement Test Tenth Edition [SAT 10] mathematics problem solving and science outcomes) S-1

Table S2 Mean characteristics of analytic sample contributing to the Year 2 component of the Bell-Bradley estimate (associated with Stanford Achievement Test Tenth Edition [SAT 10] mathematics problem solving and science outcomes) S-4

Table T1 Multiple observations per student used in analyzing two-year effects T-1

Table U1 Grade 5 mathematics topics covered to a moderate extent or more by Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1 ....U-1

Table U2 Grade 5 science topics covered to a moderate extent or more by Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1 ....U-1

Table U3 Grade 7 mathematics and science topics covered by Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1....U-2

Table U4 Use of instructional methods more than 25 percent of the time by grade 5 mathematics Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1....U-3

Table U5 Use of instructional methods more than 25 percent of the time by grade 5 science Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1....U-3

Table U6 Use of instructional methods more than 25 percent of the time by grade 7 mathematics and science Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1 ....U-3

Table V1 Parameter estimates on probability scale for odds-ratio tests of differences between Alabama Math, Science, and Technology Initiative (AMSTI) and control conditions in Year 1 associated with summer professional development and in-school support outcomes V-1

Table V2 Parameter estimates for odds-ratio tests of differences between Alabama Math, Science, and Technology Initiative (AMSTI) and control conditions in Year 1.... V-2

Table W1 Descriptive statistics for variables that change to binary scale in Year 1 ....W-1

Table X1 Comparison of assumed parameter values and observed sample statistics for statistical power analysis associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome after one year X-1
Table X2 Comparison of assumed parameter values and observed sample statistics for statistical power analysis associated with Stanford Achievement Test Tenth Edition (SAT 10) science outcome after one year X-2
Table X3 Comparison of assumed parameter values and observed sample statistics for statistical power analysis associated with active learning in mathematics outcome after one year X-2
Table X4 Comparison of assumed parameter values and observed sample statistics for statistical power analysis associated with active learning in science outcome after one year X-3
Table Y1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after one year Y-1
Table Y2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after one year Y-2
Table Y3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after one year Y-3
Table Z1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after one year Z-1
Table Z2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after one year Z-2
Table Z3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after one year Z-2
Table AA1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) effect on teaching for active learning in mathematics after one year AA-1
Table AA2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teaching for active learning in mathematics after one year AA-1
Table AA3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning in mathematics after one year ... AA-2
Table AB1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teaching for active learning in science after one year AB-1
Table AB2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teaching for active learning in science after one year AB-1

Table AB3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning in science after one year AB-2
Table AC1 Sensitivity analyses of the one-year effect of Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement AC-1
Table AD1 Sensitivity analyses for one-year effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement AD-1
Table AE1 Sensitivity analyses for one-year effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms AE-1
Table AF1 Sensitivity analyses for one-year effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms AF-1
Table AG1 Mean characteristics of nonimplementing and implementing control group schools AG-5
Table AG2 Implementation of the Alabama Math, Science, and Technology Initiative (AMSTI) by Alabama Math, Science, and Technology Initiative and control group schools during first year of Alabama Math, Science, and Technology Initiative intervention AG-10
Table AG3 Mean background characteristics of students in first year of Alabama Math, Science, and Technology Initiative (AMSTI) implementation in Year 1 and Year 2 .... AG-13
Table AI1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after two years AI-1
Table AI2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after two years AI-2
Table AI3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after two years AI-3
Table AI4 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after two years AI-4
Table AI5 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after two years AI-5
Table AI6 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after two years AI-5
Table AJ1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student reading achievement after one year AJ-1
Table AJ2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student reading achievement after one year AJ-2

TABLE AJ3 ESTIMATES OF MATCHED-PAIR FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STUDENT READING ACHIEVEMENT AFTER ONE YEAR AJ-3
TABLE AK1 ESTIMATES OF FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE T OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON TEACHER CONTENT KNOWLEDGE IN MATHEMATICS AFTER ONE YEAR AK-1
TABLE AK2 ESTIMATES OF RANDOM EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON TEACHER CONTENT KNOWLEDGE IN MATHEMATICS AFTER ONE YEAR AK-1
TABLE AK3 ESTIMATES OF FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON TEACHER CONTENT KNOWLEDGE IN SCIENCE AFTER ONE YEAR AK-2
TABLE AK4 ESTIMATES OF MATCHED-PAIR FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON TEACHER CONTENT KNOWLEDGE IN SCIENCE AFTER ONE YEAR AK-3
TABLE AK5 ESTIMATES OF FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STUDENT ENGAGEMENT IN MATHEMATICS AFTER ONE YEAR AK-4
TABLE AK6 ESTIMATES OF RANDOM EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STUDENT ENGAGEMENT IN MATHEMATICS AFTER ONE YEAR AK-4
TABLE AK7 ESTIMATES OF FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STUDENT ENGAGEMENT IN SCIENCE AFTER ONE YEAR AK-5
TABLE AK8 ESTIMATES OF RANDOM EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STUDENT ENGAGEMENT IN SCIENCE AFTER ONE YEAR AK-5
TABLE AK9 ESTIMATES OF MATCHED-PAIR FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON STUDENT ENGAGEMENT IN SCIENCE AFTER ONE YEAR.... AK-6
TABLE AK10 ANALYTIC SAMPLE USED TO ASSESS VARIATION IN EFFECTS OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON ACHIEVEMENT FOR SUBGROUPS OF STUDENTS AFTER ONE YEAR AK-7
TABLE AL1 ESTIMATES OF EFFECTS FOR TERMS INVOLVING THE INDICATOR OF TREATMENT STATUS IN THE ANALYSIS OF THE MODERATING EFFECT OF THREE-LEVEL PRETEST VARIABLE AL-2
TABLE AM1 ESTIMATES OF FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE MODERATING EFFECT OF MINORITY STATUS ON THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON READING AFTER ONE YEAR AM-1
TABLE AM2 ESTIMATES OF RANDOM EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE MODERATING EFFECT OF MINORITY STATUS ON THE IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON READING AFTER ONE YEAR AM-2
TABLE AM3 ESTIMATES OF MATCHED-PAIR FIXED EFFECTS FROM THE BENCHMARK MULTILEVEL ANALYSIS OF THE MODERATING EFFECT OF MINORITY STATUS ON IMPACT OF THE ALABAMA MATH, SCIENCE, AND TECHNOLOGY INITIATIVE (AMSTI) ON READING AFTER ONE YEAR AM-2

Table AN1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by racial/ethnic minority students after one year AN-1
Table AN2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading achievement by racial/ethnic minority students after one year AN-2
Table AN3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading achievement by racial/ethnic minority status after one year AN-2
Table AO1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by White students after one year AO-1
Table AO2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by White students after one year AO-2
Table AO3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by White students after one year AO-2

Box

Box AG1 Factors that need to remain stable over time for effect of the Alabama Math, Science, and Technology Initiative (AMSTI) to be the same in the first year of implementation in both groups of schools (treatment and control) that received it AG-1

Summary

Partly motivated by the 1996 National Assessment of Educational Progress scores, which were below the national average for Alabama's grade 4–8 students in mathematics and grade 8 students in science, the Alabama State Department of Education (ALSDE) developed a statewide initiative to improve mathematics and science teaching and student achievement in kindergarten through grade 12 (K–12). The Alabama Math, Science, and Technology Initiative (AMSTI) is a two-year intervention intended to better align classroom practices with national and statewide teaching standards—and ultimately to improve student achievement—by providing professional development, access to materials and technology, and in-school support for teachers.

AMSTI, a schoolwide intervention, was introduced in a set of 20 schools in 2002. Each year since then, the state has rolled out the program to additional schools within its 11 regions.[2] By 2009, about 40 percent of the state's 1,518 schools were designated as AMSTI schools. Funding for the program from the state legislature was $46 million in 2009.[3]

Given the policy relevance and level of investment in AMSTI, the Regional Educational Laboratory Southeast mounted a longitudinal, cluster randomized controlled trial to determine the effectiveness of AMSTI in grades 4–8, as implemented in five regions in the state. Previous evaluations of the program's effects on students in K–12 did not use randomized controlled trials. The most recent evaluation (Miron and Maxwell 2007) reported that students in grade 5 in AMSTI schools outperformed students in non-AMSTI schools in mathematics, science, and reading and students in grade 4 in AMSTI schools outperformed their counterparts in non-AMSTI schools in reading only. These evaluations used a study design that compared school-level test scores of AMSTI schools with non-AMSTI schools in the same district but did not establish preintervention comparability. This study's randomized controlled trial design improves on previous evaluations because it eliminates selection bias and establishes the preintervention comparability of the two groups.

The AMSTI theory of action posits that in order to improve student achievement, teacher instructional strategies should include higher levels of hands-on, inquiry-based instruction. The three components of the program that foster this type of instruction are comprehensive professional development delivered through a 10-day summer institute and follow-up training during the school year; access to program materials, manipulatives, and technology needed to deliver hands-on, inquiry-based instruction; and in-school support by AMSTI lead teachers and site specialists who offer mentoring and coaching for instruction. The full program is delivered over the course of two years. In each region, AMSTI site specialists partner with a local university or college. ALSDE oversees the professional development and implementation of the program.

[2] Alabama has 11 regional inservice centers (RICs), which were established by the Alabama Legislature in 1984 to provide "rigorous inservice training in critical needs areas for the state's public school personnel." The 11 AMSTI regions follow the same boundaries as the RICs.
[3] Accessed from the web on May 8, 2010 (http://www.alsde.edu/general/quick_facts.pdf).

The AMSTI theory of action provided the theoretical basis for selecting the research questions addressed in this report. The primary confirmatory analyses address the effect of AMSTI on student achievement in mathematics problem solving and science after one year. These outcomes, though related and expected to be positively correlated, are from different content domains. The primary research question looks at whether the intervention had an effect on mathematics problem solving or science knowledge.[4]

The secondary research question addresses the effect of AMSTI on classroom practices, which are the mediating link between the intervention components and student achievement. The effect of AMSTI on classroom practices is measured by a composite variable of teacher self-reported time (in minutes) using hands-on instruction, inquiry-based instruction, and instruction promoting student use of higher-order thinking skills. This composite "active learning" score was computed separately for mathematics and science instruction. As the initiative may be successful at increasing active learning instruction for one subject area but not the other, the study examines whether the intervention had an effect on either domain—active learning instruction in mathematics or active learning instruction in science.

The study addresses the following confirmatory research questions:

Primary confirmatory research question: effects on student achievement after one year

- What is the effect of AMSTI on:
    a. student achievement in mathematics problem solving *after one year*?
    b. student achievement in science *after one year*?

Secondary confirmatory research question: effects on classroom practice after one year

- What is the effect of AMSTI on:
    a. the use of active learning instructional strategies by mathematics teachers *after one year*?
    b. the use of active learning instructional strategies by science teachers *after one year*?

The study also addresses the following exploratory research questions:

Exploratory research question: effects on student achievement after two years

- What is the effect of AMSTI on:
    a. student achievement in mathematics problem solving *after two years*?
    b. student achievement in science *after two years*?

[4] The decision to examine these as separate outcomes was further warranted by the program design elements (see table 1.1 in chapter 1). During the professional development, trainers use content- and grade-specific instructional methods; there are separate mathematics and science specialists; and separate curriculum modules.

Exploratory research question: effects on student achievement in reading after one year

- What is the effect of AMSTI on student achievement in reading *after one year*?

Exploratory research questions: effects on teacher content knowledge and student engagement after one year

- What is the effect of AMSTI on:
  a. mathematics teachers' reported level of content knowledge *after one year*?
  b. science teachers' reported level of content knowledge *after one year*?
- What is the effect of AMSTI on:
  a. mathematics teachers' reported level of student engagement *after one year*?
  b. science teachers' reported level of student engagement *after one year*?

Exploratory research questions: variations in effects on student achievement for specific subgroups of students after one year

- Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by student pretest scores? What is the effect of AMSTI on these outcomes after one year for students with pretest scores that fall in the low, middle, and high ranges?
- Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by low-income status, proxied by enrollment in the free or reduced-price lunch program (as part of the National School Lunch Program)? What is the effect of AMSTI on these outcomes after one year for students enrolled in the free or reduced-price lunch program and students who are not enrolled?
- Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by racial/ethnic minority status? What is the effect of AMSTI on these outcomes after one year for racial/ethnic minorities and for White students?
- Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by gender? What is the effect of AMSTI on these outcomes after one year for boys and for girls?

Although AMSTI is a two-year program, the confirmatory analyses address the effect of the program after the first year. The effect of AMSTI after the full intervention was implemented (that is, after two years) cannot be estimated without additional assumptions because, as detailed in chapter 2, the control group entered the program after its first year and was no longer a pure control group in the second year. Researchers selected an appropriate method to estimate the two-year effects; however, the necessity of additional assumptions makes the analyses exploratory rather than confirmatory. This limitation on the study's design means that only the one-year effect on mathematics problem solving and science can be considered confirmatory.

Beyond the two-year impacts, the exploratory questions pertain only to the first year of AMSTI. Unlike questions concerning two-year effects, they can be answered without additional assumptions necessitated by the entry of the control group into the AMSTI program in the second year of the study. These analyses address the effect of AMSTI on student achievement in reading, teacher content knowledge, student engagement, and variations in effects on student achievement for particular subgroups of students. These questions are important to understanding the full effects of AMSTI and in potentially identifying ways to improve the program. The rationale for selecting these questions arises from several sources: the AMSTI theory of action; interest from program developers; prior research on AMSTI and within the fields of science, technology, engineering, and mathematics; and measured state achievement gaps.

The study took advantage of ALSDE's rollout of AMSTI to specific regions during the study years. To participate in the study, schools must have housed at least one grade between grades 4 and 8, and at least 80 percent of a school's mathematics and science teachers must have agreed to participate. From the eligible schools that applied to the program, researchers made a purposeful effort to select a sample that was representative of the population of schools in the regions involved. Pairs of similar schools were selected from the pool of applicants based on similarity in mathematics achievement, the percentage of minority students, and the percentage of students from low-income households. Within each pair, schools were randomly assigned either to the AMSTI condition, in which teachers received AMSTI training and program materials, or to the control condition, in which teachers used their existing mathematics and science programs.

Because Alabama did not plan to introduce the program in the number of schools required by the experiment in one year, the experiment combined two "subexperiments", one starting in 2006 and the other starting in 2007. The full sample combined the two samples from the two "subexperiments" and included 82 schools, with about 780 teachers and 30,000 students in grades 4–8 across the two subexperiments.[5] In Subexperiment 1, the first set of 40 schools (within three regional AMSTI sites) was randomized to conditions in the winter of 2006. In Subexperiment 2, the second set of 42 schools (within two regional AMSTI sites) was randomized to conditions in the winter of 2007. To estimate the effects of AMSTI after one year (confirmatory analysis), data from both subexperiments were pooled and analyzed together after their respective first year. The integrity of the samples used in the confirmatory analysis was maintained, because the difference in attrition between the intervention and control groups was less than 5 percentage points and overall attrition was 2.5 percent or less for all outcomes. To estimate the effects of AMSTI after two years, data from both subexperiments were pooled and analyzed together after the respective second year.

Data were collected at multiple levels. Sources included classroom rosters, student achievement and demographic data, professional development training logs and observations, professional development teacher surveys, interviews with teachers and principals, classroom observations, and web-based surveys of teachers and principals.[6] In both subexperiments,

---

[5] This number represents the approximate number of teachers and students in the 82 study schools during years 1 and 2 of Subexperiment 1 and Subexperiment 2. For precise numbers of teachers and students used in each analysis, see chapter 2.

[6] Training logs, in-person interviews with teachers and principals, and classroom observations were conducted only with Subexperiment 2, because researchers did not receive approval from the Office of

teachers in AMSTI schools were trained in the program the summer following randomization and before their first year of implementation (2006/07 for Subexperiment 1, 2007/08 for Subexperiment 2).

Inferential tests on web-based teacher survey data were conducted to examine the differences between AMSTI and control schools in the presence of the three main intervention components (summer professional development, access to materials and manipulatives, and in-school support). AMSTI teachers were more likely to have participated in summer professional development than were control teachers (87 percent versus 24 percent for mathematics teachers, 84 percent versus 24 percent for science teachers). AMSTI teachers also reported having greater access to materials than did control teachers (78 percent versus 41 percent for mathematics teachers, 61 percent versus 33 percent for science teachers). AMSTI teachers were more likely to receive in-school support than were their control counterparts (59 percent versus 40 percent for mathematics teachers, 65 percent versus 25 percent for science teachers). All these differences were statistically significant at $p < .05$ (for specifics see chapter 3).[7]

The effect of AMSTI on student achievement in mathematics after one year, as measured by end-of-the-year scores on the Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving assessment of students in grades 4–8, was 2.06 scale score units (figure 1). The difference of 0.05 standard deviation in favor of AMSTI schools is equivalent to a gain of 2 percentile points on the SAT 10 mathematics problem solving assessment for the average control group student had the student received AMSTI. The 0.05 standard deviation is statistically significant but smaller than the effect the research team believed would be detectable by the experiment as designed. Whether this size effect is educationally important is an open question. It may be useful to convert this effect into a more policy-relevant metric—additional student progress measured in days of instruction. In these terms, the average estimated effect of AMSTI was equivalent to 28 days of additional student progress over students receiving conventional mathematics instruction.[8] The effect of AMSTI on student achievement in science,

---

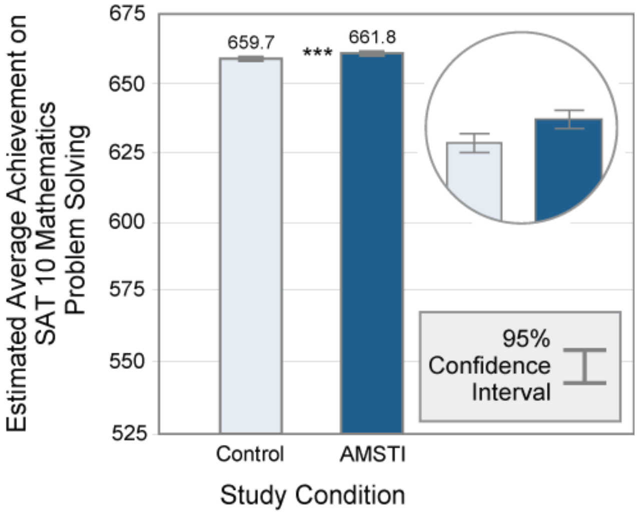
Management and Budget in time to collect these implementation data during the 2006/07 school year for Subexperiment 1. Student-level data and web-based survey data from teachers and principals were collected for Year 1 of Subexperiment 1 through a research grant (IES: #R305E040031) from the Institute of Education Sciences to Empirical Education Inc., with permission from the IES program officer.

[7] The implementation analyses presented in this report aim simply to describe program implementation for each program component. The study design did not include assessment and analysis of the AMSTI implementation quality since objective benchmarks for AMSTI implementation do not exist.

[8] To obtain this value, we express the estimated average score gain in the treatment group as a proportion of the score gain in the control group (T=treatment, C=control):

$\frac{\hat{\overline{Y}}_{T(post)} - \hat{\overline{Y}}_{T(pre)}}{\hat{\overline{Y}}_{C(post)} - \hat{\overline{Y}}_{C(pre)}} = \frac{\hat{\overline{Y}}_{C(post)} - \hat{\overline{Y}}_{C(pre)} + \hat{T}}{\hat{\overline{Y}}_{C(post)} - \hat{\overline{Y}}_{C(pre)}} = 1 + \frac{\hat{T}}{\hat{\overline{Y}}_{C(post)} - \hat{\overline{Y}}_{C(pre)}}$. We then multiply this value by 180 (assuming a 180-day school year in Alabama) which yields the estimated projected number of days of schooling by the control group, had they been in the treatment condition. Subtracting 180 from this as measured by end-of-the-year scores on the SAT 10 science assessment, required only in grades 5 and 7, was not statistically significant after one year (figure 2).

Figure 1 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year



** Significant at $p < .05$; *** Significant at $p < .01$
Note: $n = 82$ schools; 18,713 students
Source: Student achievement data from tests administered as part of the state's accountability system.

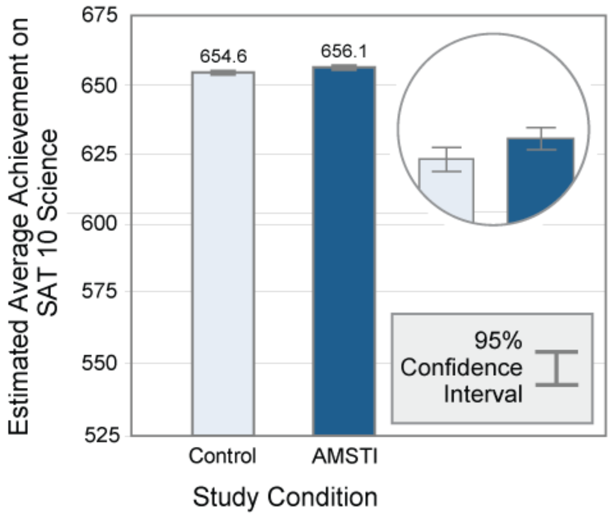
---

quantity yields an estimate of the treatment effect in terms of additional learning growth as translated into additional days of

schooling: $$IMPACT = \left(1 + \frac{\hat{T}}{\hat{\overline{Y}}_{C(post)} - \hat{\overline{Y}}_{C(pre)}}\right) \times 180 - 180 = \left(\frac{\hat{T}}{\hat{\overline{Y}}_{C(post)} - \hat{\overline{Y}}_{C(pre)}}\right) \times 180 = 28 days .$$

This calculation assumes that the treatment effect accumulates linearly with time over the course of a grade. A formal test of the linearity of the accrual of the treatment effect was not conducted. If the treatment effect does not accrue linearly then this extrapolation of the number of days may not be accurate.

Figure 2 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year



Note: n = 79 schools; 7,528 students
Source: Student achievement data from tests administered as part of the state's accountability system.

AMSTI also had a positive and statistically significant effect on classroom practices in mathematics and science after one year. Based on multiple surveys in which teachers reported the number of minutes of active learning strategies used during the previous 10-day period, AMSTI mathematics teachers averaged 49.83 more minutes, and AMSTI science teachers averaged 40.07 more minutes than control teachers. These estimated effects are equivalent to 0.47 standard deviation in mathematics and 0.32 standard deviation in science. Although teachers in both the AMSTI and control groups reported using active learning instructional strategies, teachers in AMSTI schools reported spending more time engaged in this type of instruction.

The exploratory investigation of the two-year effect of AMSTI on student achievement on the SAT 10 mathematics problem solving test found a positive and statistically significant result of 3.74 scale score units. This effect represents a difference of 0.10 standard deviation in favor of AMSTI schools, equivalent to a gain of 4 percentile points for the average control group student had the student received AMSTI for two years. This estimate of the average effect of AMSTI after two years can be translated into an estimated 50 days of additional student progress over students receiving conventional mathematics instruction.

The exploratory investigation of the two-year effect of AMSTI on student achievement in science also found a statistically significant result, with a magnitude of 4.00 scale score units. This effect represents a difference of 0.13 standard deviation in favor of AMSTI schools, equivalent to a gain of 5 percentile points for the average control group student had the student received AMSTI for two years.[9]

The effect of AMSTI on student achievement in reading after one year, as measured by end-of-the-year scores on the SAT 10 reading assessment of students in grades 4–8, was 2.34 scale score units. The statistically significant difference of 0.06 standard deviation in favor of AMSTI schools is equivalent to a gain of 2 percentile points on the SAT 10 reading assessment for the average control group student had the student received AMSTI. This difference can be translated into an estimated 40 days of additional student progress over students receiving conventional reading instruction.

The effect of AMSTI on teacher-reported content knowledge after one year was not statistically significant in either mathematics or science. AMSTI did have a positive and statistically significant effect on student engagement after one year, measured on a 5-point scale ranging from "not engaged" to "fully engaged." AMSTI teachers were more likely than control teachers to rate their students as achieving higher levels of engagement.

An exploration of the differential effects of AMSTI on student achievement for subgroups of students found no statistically significant differential effects on student achievement in mathematics or science based on racial/ethnic minority status, eligibility for free or reduced-price lunch, gender, or pretest level. In reading, however, AMSTI had a statistically significant differential effect for minority and White students of 3.04 scale score points ($p < .001$). This difference can be translated into days of student progress, where progress is measured as the average gain in test scores over the course of the school year by the control group using conventional reading instruction. In this metric, White students in AMSTI made an estimated 52 more days of progress than minority students in AMSTI. The effect of AMSTI on reading achievement for minority students was not statistically significant ($p = .294$); for White students, there was a statistically significant positive effect of AMSTI on reading achievement ($p < .001$).

---

[9] The analysis of the two-year impact of AMSTI on student achievement is exploratory. Readers should exercise caution in interpreting the results. For instance, we remind the reader that with exploratory analyses we do not perform multiplicity adjustments. As a consequence, a less strict criterion is used with exploratory analyses for deciding whether a particular result achieves statistical significance, with the drawback that it increases the probability of finding a spurious impact. For the two-year impact on mathematics problem solving ($p = .030$) and science ($p =.038$) the results reach statistical significance under the less strict criterion (alpha = .05). Under the more strict criterion used with the primary confirmatory analyses (alpha = .025) these results would not have been considered statistically significant.

1: Introduction and study overview

This report presents the results of an experiment conducted in Alabama beginning in the 2006/07 school year, to determine the effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI), which aims to improve mathematics and science achievement in the state's K–12 schools. This chapter first describes the theoretical underpinnings of AMSTI, identifies its components, and reviews prior research on the initiative. It then describes the study design and presents the research questions. Subsequent chapters detail the research methods, the implementation of the program, and the effects on student achievement and classroom practices.

Strengthening skills of mathematics and science teachers nationwide

Strengthening the instructional skills of mathematics and science teachers nationwide is an essential step in adequately preparing American students to compete globally. In testimony to Congress in 2005, representatives of the National Academies of Science pointed to mounting concern that the United States is not producing an adequate number of science, technology, engineering, and mathematics graduates prepared to meet the global demands of the 21st century.[10] To address the problem, they recommended strengthening the skills of mathematics and science teachers (Augustine, Vagelos, and Wulf 2005).

Providing on-site direct professional development, coaches, and technical assistance to schools is one of several state-level strategies, or levers of change, used to strengthen the skills of mathematics and science teachers (Edmunds and McColskey 2007). AMSTI is an example of this strategy. The creation of AMSTI was motivated by the understanding that "a major challenge that America faces as it moves into the 21st century is assuring that its citizens have the mathematical, scientific and technological skills and knowledge necessary to be productive members of society" (AMSTI Committee 2000, p. 20).

The statewide effort was further motivated by the 1996 National Assessment of Educational Progress (NAEP), on which Alabama's grade 4 and grade 8 students scored below the national average in mathematics. Grade 8 students—the only grade in Alabama whose science test scores were reported in 1996—also scored below the national average in science (O'Sullivan, Jerry, Ballator, and Herr. 1997). In response, policymakers in Alabama undertook a statewide effort to raise students' achievement levels in mathematics and science.

---

[10] On a 2003 international assessment of 15-year-old students, the United States ranked 28th in mathematics literacy and 24th in science literacy (Lemke et al. 2004).

The Alabama Math, Science, and Technology Initiative (AMSTI): A state-level strategy to strengthen skills of mathematics and science teachers

In November 1999, the Alabama State Department of Education (ALSDE) appointed a 38-member blue ribbon committee of K–12 educators, university professors, administrators, and business and industry leaders to recommend and formulate an action plan to improve mathematics, science, and technology education throughout the state. Based on a multistep process of reviewing the research literature; examining international, national, and state assessment data; investigating national standards; and identifying the needs of Alabama teachers through a statewide survey of K–12 public school mathematics and science teachers,[11] the committee released the following findings (ALSDE 2000):

- The two greatest needs identified by teachers were access to technology and integration of technology within mathematics and science instruction. Fifty-six percent of responding mathematics teachers and 54 percent of responding science teachers identified "incorporating technology into the classroom" as one of their four greatest needs.
- Teacher instructional strategies were not aligned with national standards in mathematics and science. The most frequent instructional strategies appeared to be lecture and whole-group discussion, with the more innovative techniques endorsed by the National Council for Teachers of Mathematics and the National Research Council (for example, working on projects, using hands-on materials) used less often. More than half of science teachers identified lecture and whole-group discussion as their primary forms of instruction.
- Other identified needs included more planning time for teachers who taught the same subject at different grade levels (cited by 39 percent of mathematics teachers and 42 percent of science teachers), better assessment approaches than paper and pencil tests (cited by 55 percent of mathematics teachers), and more involvement in professional development activities directly related to mathematics and science instruction (cited by more than 70 percent of both mathematics and science teachers). Twenty-one percent of both mathematics and science teachers said they "almost never" had the opportunity to participate in such training.

---

[11] For each subject, survey packets with return stamped envelopes were sent out to 250 teachers in kindergarten through grade 5, 125 teachers in grades 6–8, and 125 teachers in grades 9–12 for a total sample size of 500 mathematics and 500 science teachers. The response rate was 54.6 percent for mathematics teachers and 37.4 percent for science teachers (ALSDE 2000).

The committee then developed recommendations and action plans based on these findings and upon a year-long extensive review of research literature (AMSTI Committee 2000), including the recommendations of the American Association for the Advancement of Science (1990) and the National Council of Teachers of Mathematics (2000). The committee made five final recommendations:

- Classroom practice should incorporate hands-on, inquiry-based instruction.
- Mathematics and science curricula should focus on a reduced number of topics, emphasizing depth versus breadth of knowledge.
- Performance-based assessments should complement standardized testing strategies.
- Content-specific and ongoing professional development must be provided to teachers.
- Adequate and accessible technological resources and classroom materials, from hand-held calculators to computers, are required for effective classroom instruction.

Once these recommendations were adopted, ALSDE charged two committees, one for science and one for mathematics, with turning the recommendations into a program. Under the guidance of the department, writers were hired to develop grade- and subject-specific modules and implementation guides for teachers in line with the recommendations. Once the curricula were developed, trainers were hired to build a pool of individuals capable of teaching the modules. According to the initial design, participating schools were to send their staff for training for two consecutive summers. During the first summer, teachers would receive training in the first half of the program and implement those units during their school's first year of participation. The following summer, teachers would be trained on the second half of the program and teach the full AMSTI curriculum starting the next school year. The program staff has continued to add and revise modules based on feedback and to maximize alignment with the state's content standards, the Alabama Course of Study.

The first group of 20 participating schools started AMSTI in 2002. Each ensuing year, the program expanded to additional sites. In 2009, about 40 percent of Alabama's 1,518 public schools were designated as AMSTI schools.[12] Funding for AMSTI, which comes from the state legislature as part of the education budget, was $46 million in 2009.

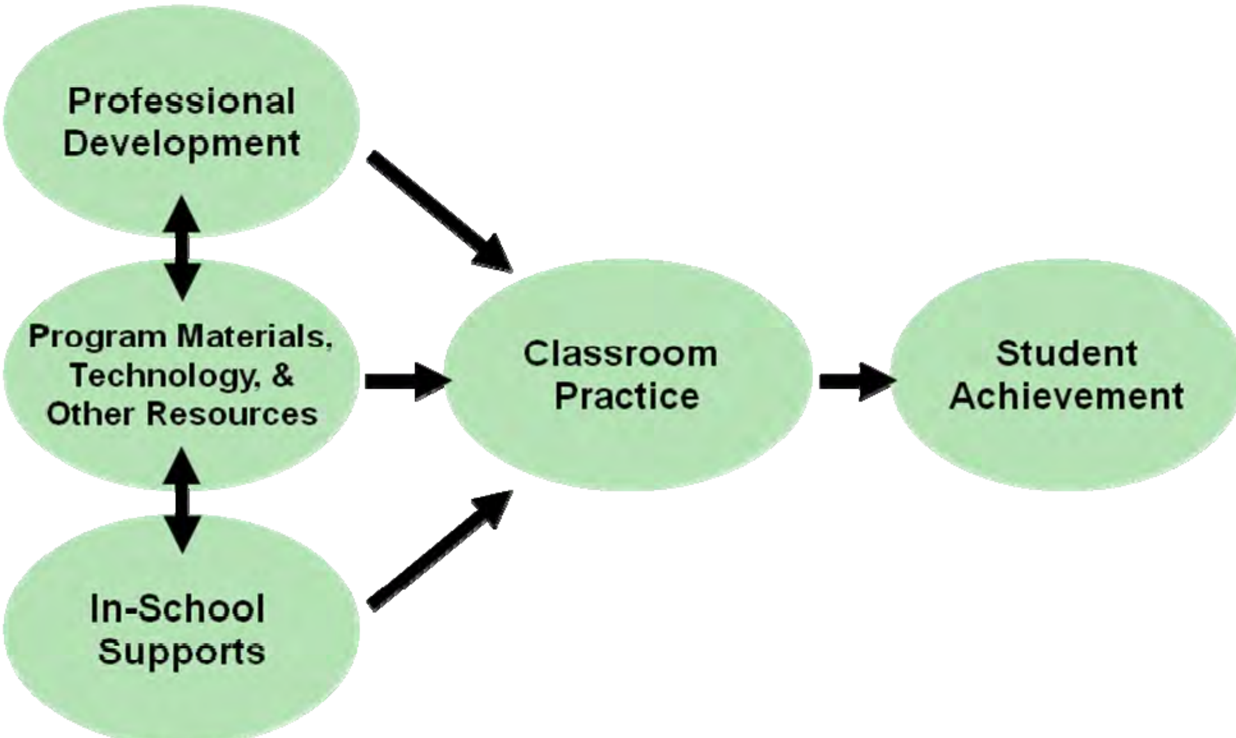
Alabama Math, Science, and Technology Initiative (AMSTI) theory of action

AMSTI developers posited that teacher quality and effectiveness were the keys to improving student test scores in mathematics and science. They believed that the most direct way to improve teacher quality was to create an in-depth, comprehensive professional development program reflective of national standards in mathematics, science, and technology and to provide teachers with resources to support what they learned in that program. The AMSTI model is based on the hypothesis that intensive, comprehensive professional development; in-school support (for example, teacher coaching provided by technical assistance staff); and associated resources and materials (for example, curricular materials, manipulatives, and microscopes) will lead to

[12] Accessed on May 8, 2010, from (http://www.alsde.edu/general/quick_facts.pdf).

teachers' use of effective instructional strategies that are aligned with statewide and national standards. These changes in instructional strategies were hypothesized to lead to improved student achievement in mathematics and science (figure 1.1).

Figure 1.1 Alabama Math, Science, and Technology Initiative (AMSTI) theory of action



The AMSTI theory of action implies two fundamental mechanisms of change, which are supported by research. The first mechanism concerns higher levels of hands-on, inquiry-based instruction, which are thought to lead to in-depth mastery of core mathematics and science concepts and higher achievement scores. The second mechanism involves intensive, comprehensive professional development, ready access to instructional materials and technology, and in-school supports for teachers, which were hypothesized to lead to increases in hands-on, inquiry-based instruction within the classroom. Research shows that the first mechanism leads to significant improvements in a variety of student learning outcomes, including higher-order critical thinking and process skills, problem solving abilities, and achievement scores (Romberg, Carpenter, and Dremock 2005; Chang and Mao 1999; Battista 1999). The foundation for the second mechanism is research on effective practices and programs that help reform teaching practices in the classroom (Gamoran et al. 2003; Glennan and Resnick 2004; Loucks-Horsley, Stiles, and Hewson 1996; Lord and Miller 2000).

These three key factors—intensive, comprehensive professional development; ready access to instructional materials and technology; and in-school supports for teachers—were hypothesized to work interactively (as indicated by the double-headed arrows in figure 1.1) to increase hands-on, inquiry-based classroom instructional strategies. These classroom practices, in turn, were hypothesized to lead to higher student achievement.

Alabama Math, Science, and Technology Initiative (AMSTI) program components and elements

The AMSTI program consists of three components that map directly onto each key factor hypothesized to influence classroom practice within the theory of action: professional development; program materials, technology, and other resources; and in-school supports (table 1.1; see chapter 3). The components are delivered over two years to at least 80 percent of mathematics and science teachers in participating middle schools; in elementary schools, they are provided to regular classroom teachers.[13]

Table 1.1 Alabama Math, Science, and Technology Initiative (AMSTI) components and corresponding elements

Component	Elements
Professional development	• Teachers and principals participate in two-week summer institutes at which they are trained in the AMSTI curriculum, which focuses on increasing teacher content-knowledge and hands-on, inquiry-based instruction. • Elementary school teachers attend one week of training for AMSTI mathematics and one week for AMSTI science. Secondary school teachers attend two weeks of AMSTI training for their subject area and grade level. • Teachers attend summer institutes for two consecutive years, during the summers before and after their first year of classroom implementation. • Summer institutes are taught by master trainers who are AMSTI certified at one of the AMSTI sites. • Trainers use content- and grade-specific instructional methods. For example, trainers of grade 8 mathematics teachers model lessons that align with the Alabama Course of Study for grade 8 and use instructional methods appropriate for grade 8. • Trainers model lessons of hands-on, inquiry-based instruction. • Teachers receive follow-up or on-site professional development during the school year.
Program materials, technology, and other resources	• Teachers are provided with all materials needed to deliver hands-on, inquiry-based instruction. • Program materials include teacher guides, student guides, participant manuals (grade and subject specific and aligned with the Alabama Course of Study), student assessments, software, and various toolkits composed of manipulatives and hands-on activities. • Hands-on materials and manipulatives are rotated among schools in bins or “kits” delivered to schools by AMSTI sites and picked up for complete refurbishment.
In-school supports	• AMSTI site specialists are available for on-site mentoring to help teachers implement lessons throughout the school year.

[13] In an attempt to bring about institutional change at participating schools, ALSDE limits participation to schools that demonstrate staff buy-in by submitting signatures from the school principal and at least 80 percent of the subject area teachers. After the first two years, support is reduced. However, AMSTI continues to train new teachers and to replenish curriculum and supplies at program schools.

Component	Elements
	• Schools designate one teacher to receive additional training and serve as the school-based AMSTI lead teacher and AMSTI liaison. The lead teacher provides mentoring to newly trained faculty and serves as one conduit for communication with AMSTI sites. • AMSTI schools hold regularly scheduled sessions at which “learning teams” or “study groups” of AMSTI teachers within a school meet to plan and discuss AMSTI-related issues and student responses.

There are 11 regional AMSTI sites, each partnered with a local university or college. The sites are responsible for delivering AMSTI program elements, including the summer training, ongoing professional development during the school year, and the training of any new hires at schools. The sites are also responsible for storing, refurbishing, and distributing materials and kits to all schools within their region. Each site is staffed with a director, mathematics and science specialists, a materials manager, and materials staff. ALSDE oversees the program throughout the state, ensuring consistent quality across schools. It certifies trainers to teach at the summer institutes, sets the curricula and activities, and provides oversight support to each AMSTI site.

Prior research on the effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement

The Institute for Communication and Information Research at the University of Alabama, an external evaluator, has evaluated AMSTI for four school years since its inception in Alabama schools in 2002 using a quasi-experimental design (Miron and Maxwell 2004; 2005a, b; 2006a, b; 2007). All evaluations used a similar methodology and data sources, pairing AMSTI schools with other schools from the same districts that did not use the program and comparing school-level standardized test scores.[14]

Evaluations focused on the impact of AMSTI on student achievement only; implementation data were not collected. The primary impacts of interest were mathematics and science test scores, but researchers also investigated possible secondary effects on reading and writing scores. Elementary (kindergarten through grade 5), middle school (grades 6–8), and high school (grades 9–12) levels were analyzed separately. The most recent evaluation (Miron and Maxwell 2007) analyzed the school-level mean percentile ranks on the 2007 Stanford Achievement Test Series, Tenth Edition (SAT 10) for the 195 schools that have adopted AMSTI since 2002 and the 576 schools from the same school districts that did not use the program. It reported statistically significant findings at the elementary school level, with grade 5 students in AMSTI schools outperforming their counterparts in non-AMSTI schools on the SAT 10 by 3.51 percentile points in mathematics ($p = .037$) and 6.35 percentile points in science ($p = .001$). The

[14] For 2004–07, the external evaluator examined results from the following standardized tests: the Stanford Achievement Test Series, Tenth Edition; the Alabama Reading and Math Test; and the Alabama High School Graduation Exam. Starting in 2005, the evaluation also included results from the Alabama Direct Assessment of Writing.

evaluators also found that AMSTI had statistically significant spillover effects on SAT 10 reading outcomes of 4.05 percentile points in grade 4 ($p = .020$) and 3.51 percentile points in grade 5 ($p = .037$). AMSTI was not shown to have statistically significant effects on students in grades 6–8 on any of the SAT 10 tests.

As with other quasi-experimental studies, results may have been subject to substantial bias, particularly because schools must apply to participate in AMSTI and so may already be more motivated to improve their mathematics and science programs than schools that do not apply. In addition, none of the evaluations examined whether there was baseline equivalence on possible confounders, including pretest measures, nor did they attempt to adjust for potential bias because of nonequivalence between AMSTI and non-AMSTI schools. These issues raise concerns about the internal validity of study findings—specifically, whether the observed increase in AMSTI students' achievement can be explained only by their participation in the program and not by other plausible explanations, such as prior achievement.

Overview of study design

Given the policy relevance of AMSTI and the state's investment in the program, the Regional Educational Laboratory Southeast mounted a longitudinal, cluster randomized controlled trial to determine the effectiveness of AMSTI, as implemented in five regions of Alabama. Random assignment of schools controlled for potential explanations other than AMSTI as the cause of observed differences in student achievement. Removing sources of potential bias present in the previous evaluations enhanced the level of confidence in the evaluation of the effect of AMSTI on student outcomes.

The research team selected a sample of 82 schools, over a period of two school years, from a larger pool of qualified schools that expressed interest in AMSTI. Because in each region a larger number of qualified schools applied to participate in the AMSTI program than resources could accommodate, the study was able to randomly assign the schools to either the AMSTI or the control condition.

To reach the number of schools required for the AMSTI evaluation the study set up two experiments. In Subexperiment 1, 40 schools were randomized to conditions in 2006. In Subexperiment 2, another 42 schools were randomized to conditions in 2007. In all, 40 districts and about 780 teachers and 30,000 students were eligible for participation in the study.[15] For both subexperiments, the intervention group attended the summer institutes and began implementing AMSTI in the first year after random assignment, and the control group continued to use the existing mathematics and science programs. Schools assigned to the control group were accepted to the program but with a one-year delay, so that in the second year of each

[15] These numbers represent the approximate number of unique teachers and students in the 82 study schools in Year 1 and Year 2 (of Subexperiment 1 and Subexperiment 2). For precise numbers of teachers and students included in each analysis, see chapter 2.

subexperiment, the sample consisted of one group in its second year of AMSTI implementation (the first year intervention group) and another group in its first year of AMSTI implementation (the first year control group); therefore, there was no control group that did not receive AMSTI intervention.

The principals of 17 AMSTI schools and 13 control schools indicated that members of their staff had participated in the Leadership Academy for Math, Science, and Technology (LAMST) the year before the study began (2005/06 for Subexperiment 1, 2006/07 for Subexperiment 2). LAMST provided grade-and subject-specific materials and one week of training based on the AMSTI model to school teams composed of the principal, one mathematics teacher, and one science teacher in kindergarten through grade 5.[16] LAMST-trained teachers were expected to share what they learned with the rest of the teachers in their schools. ALSDE's expectation was that exposure to LAMST would encourage the schools to participate in the full AMSTI program. LAMST did not provide text or materials for teachers other than to the two who attend the training. This study examined the effect of AMSTI over and above this introductory exposure to the program.

ALSDE oversaw all aspects of AMSTI program implementation. It also led the professional development, distributed the program materials, facilitated the in-school supports, and provided data, including demographic information and standardized test scores. Individual school districts provided data on classroom teacher assignment, school enrollment status, and additional student demographic information. Researchers gathered data from teachers and principals using surveys, classroom observations, and interviews, to understand more fully the nature of training and resources, instructional strategies, and possible changes in student academic achievement.

A cluster randomized controlled trial research design for AMSTI is an advance over the previous research on the intervention, because it eliminates selection bias from the estimate of the effect. Randomization, however, does not ensure the generalizability of the findings to other regions or implementations that may use different levels of resources. AMSTI schools not in the study were not observed; whether the observed implementation of AMSTI was typical cannot be verified.

The study took advantage of ALSDE's rollout of AMSTI to specific regions during the study years. From the eligible schools that applied to participate in the program, researchers tried to select a sample that was representative of the population of schools in the regions chosen. Similarity between the sample and the reference populations was determined by school dimensions that regional experts thought were important. This process did not ensure the kind of external validity that would have been obtained through a formal random sample. The description of the selection process and the characteristics of the resulting sample (in chapter 2) allow readers to judge whether the findings might be extrapolated to their particular cases.

---

[16] The number of teachers who participated in this study and received LAMST training is not known.

This report describes the conditions of implementation to help ALSDE strengthen its program. Its key contribution is to provide rigorous estimates of whether this effort improved student achievement in mathematics and science in schools that adopted AMSTI.

Research questions

The analyses of the effect of AMSTI on key study outcomes are referred to as *confirmatory analyses*. Confirmatory analyses are those for which the evaluation was specifically designed and for which the design provides a strong basis for causal inference. *Exploratory analyses* were also conducted. The evaluation was not specifically designed to address these questions. It thus can produce only suggestive evidence and ideas for future research.

- ***Primary confirmatory research question: effect on student achievement after one year***
  - What is the effect of AMSTI on:
    a. student achievement in mathematics problem solving *after one year*?
    b. student achievement in science *after one year*?

- ***Secondary confirmatory research question: effect on classroom practice after one year***
  - What is the effect of AMSTI on:
    a. the use of active learning instructional strategies by mathematics teachers *after one year*?
    b. the use of active learning instructional strategies by science teachers *after one year*?

- ***Exploratory research question: effect on student achievement after two years***
  - What is the effect of AMSTI on:
    a. student achievement in mathematics problem solving *after two years*?
    b. student achievement in science *after two years*?

- ***Exploratory research question: effect on student achievement in reading after one year***
  - What is the effect of AMSTI on student achievement in reading *after one year*?

- ***Exploratory research questions: effects on teacher content knowledge and student engagement after one year***
  - What is the effect of AMSTI on:
    a. mathematics teachers' reported level of content knowledge *after one year*?
    b. science teachers' reported level of content knowledge *after one year*?
  - What is the effect of AMSTI on:
    a. mathematics teachers' reported level of student engagement *after one year*?
    b. science teachers' reported level of student enagement *after one year*?

- ***Exploratory research questions: variations in effects on student achievement for specific subgroups of students after one year***
    - Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by student pretest scores? What is the effect of AMSTI on these outcomes after one year for students with pretest scores that fall in the low, middle, and high ranges?
    - Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by low-income status, proxied by enrollment in the free or reduced-price lunch program (as part of the National School Lunch Program)? What is the effect of AMSTI on these outcomes after one year for students who were enrolled in the free or reduced-price lunch program and students who are not enrolled?
    - Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by racial/ethnic minority status? What is the effect of AMSTI on these outcomes after one year for racial/ethnic minorities and White students?
    - Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by gender? What is the effect of AMSTI on these outcomes after one year for boys and for girls?

The AMSTI theory of action provided the theoretical basis for selecting confirmatory outcomes (see figure 1.1). The primary outcomes selected for evaluation are measures of student achievement in mathematics problem solving and science. These outcomes were designated as primary confirmatory outcomes because the basic purpose of AMSTI is to increase student achievement in these content domains.[17]

The secondary outcomes are the classroom practices that, according to the theory of action, are the mediating link between the intervention and student achievement. The classroom practice outcomes were designated as secondary confirmatory outcomes because the main mechanism through which AMSTI is expected to affect student achievement is the set of classroom practices used by the mathematics and science teachers who implement the program.[18]

Because AMSTI was designed to increase student achievement, its success was judged by whether its estimated achievement effects were positive and statistically significant. For the confirmatory analysis, two outcomes were measured—one in mathematics and one in science. At the same time, it is useful to make summary statements about the findings that span multiple outcomes. Summary statements about the effects of AMSTI are based on whether the estimated

---

[17] The decision to examine these as separate outcomes was further warranted by the program design elements (see table 1.1). During the professional development, trainers use content- and grade-specific instructional methods; there are separate mathematics and science specialists and separate curriculum modules.

[18] These practices were measured in the surveys and used to create a composite variable, the active learning scale (see appendix A).

effect is significant for either of the two confirmatory outcomes. A multiplicity adjustment using the Bonferroni procedure was used to provide rigorous support for these statements.

Although AMSTI is a two-year program, the confirmatory analyses address the effect of only the first year of implementation (across both subexperiments). The effect of two years of exposure cannot be estimated without additional assumptions because, as detailed in chapter 2, the control group entered the program in the second year and was no longer a pure control group.[19] Researchers selected an appropriate method to estimate the two-year effects, but due to the uncertainty introduced by the assumptions required by the analysis, the findings are considered exploratory.

The additional exploratory questions pertain only to the first year of AMSTI and could be answered without additional assumptions. These additional analyses addressed the effect of AMSTI on student achievement in reading, teacher content knowledge, and student engagement, as well as variations in effects on student achievement for various subgroups of students. These questions are important to understanding the full effect of AMSTI. The rationale for selecting these questions arose from several sources: the AMSTI theory of action; interest from program developers; prior research on AMSTI and the fields of science, technology, engineering, and mathematics; and measured achievement gaps in Alabama (appendix B).

The study investigates whether participation in AMSTI mathematics and science instruction had an effect on reading achievement. This was considered an important investigation by the AMSTI developers; previous quasi-experimental evaluations have found positive spillover effects, prompting the study's technical working group to suggest this question as an area of investigation (Miron and Maxwell 2007). The AMSTI theory of action provides the primary motivation for examining the intermediate effect on teacher content knowledge and student engagement. AMSTI's professional development focuses on changing teacher content knowledge. AMSTI's emphasis on teaching methods that promote active learning in the classroom is designed to motivate and engage students in learning (University of Alabama n.d.; AMSTI Committee 2000). Examining the variations in effects on student achievement for specific subgroups of students is important to identify particular subgroups for which AMSTI is or is not working, to determine whether the intervention narrows or expands the achievement gap across these subgroups.

To inform an understanding of the primary and secondary confirmatory outcomes, the report also provides brief descriptions of implementation during the first year. The theory of action suggests that if implemented correctly, the three components of AMSTI will lead to changes in classroom practice and, consequently, to changes in student outcomes, as examined in the primary analysis.

[19] Two years after random assignment, the original control schools had participated in AMSTI for one year; intervention schools had participated in AMSTI for two years.

Structure of the report

Chapter 2 details the study design and methods, including the study setting and randomization procedures as well as descriptions of the sample, data sources and collection procedures, and analysis methods; and chapter 2 also provides enough information for assessing the internal and external validity for all analyses. Chapter 3 describes each component of the intervention as articulated through the AMSTI theory of action and corresponding implementation analyses to provide a sense of the average receipt of both AMSTI and control group components in both the intervention and control conditions. The implementation findings, which draw mainly on quantitative data, are descriptive and provide context for interpreting confirmatory findings. Chapter 4 presents the primary confirmatory results of the effect of AMSTI on student achievement outcomes and the secondary confirmatory results of the effect of AMSTI on classroom instructional practices after one year. Chapter 5 presents the exploratory results of the effect of AMSTI on student achievement after two years. Chapter 6 presents the exploratory results of the effect of AMSTI on student achievement in reading, the effect of AMSTI on teacher content knowledge and student engagement, and variations in effects on student achievement for various subgroups of students after one year. Chapter 7 summarizes and discusses the report’s main findings. Appendixes provide additional technical detail.

2: Study design and methodology

To examine the effect of AMSTI on the mathematics problem solving and science achievement of students in grades 4–8 in 82 public schools in five regions of Alabama, researchers randomly assigned a volunteer sample of schools to either implement AMSTI or continue using their school districts' existing mathematics and science programs. This chapter describes the methods used to assess the differences in student and classroom outcomes between the two groups of schools. It begins by providing the rationale for the experimental design and describing the target population, recruitment, and randomization of schools. It then details the study's data sources and collection procedures, addresses attrition, describes the composition of the experimental groups for the confirmatory outcomes, and explains the statistical methods used to generate the findings on the effect of AMSTI on students and teachers.

Rationale for experimental design

The research team chose a longitudinal, cluster randomized controlled trial design, allowing it to improve on results from earlier quasi-experimental evaluations, which could not disentangle the program's effect from possible bias caused by participant selection. The design of the current study ensures that its findings are less susceptible to bias.

Because the goal was to measure AMSTI's impact, if any, the experimental design needed to isolate the program's effect on outcomes of interest from the effects of other factors. Randomization increases the likelihood that, at the outset of the study, factors other than the program that could affect the outcomes of interest are equally distributed, on average, between the AMSTI and control schools. This is true both for observed and measured factors, as well as for variables that are unmeasured, unobserved, or unanticipated. Randomization prevents confounding of the intervention with other factors that affect outcomes, preventing bias in the results. For example, randomization increases the likelihood that lower-achieving schools are not selectively assigned to either the AMSTI or control group. The benefits of randomization are preserved as long as the trial is well implemented with safeguards against disruptions, such as differential attrition.

Unit of randomization

This study was randomized at the school level, because AMSTI uses a whole-school implementation model and requires at least 80 percent of the mathematics and science teachers in an AMSTI school to participate in the program. AMSTI requires that teachers work together in professional development groups, that lead teachers mentor others in mathematics and science, and that all of a teacher's students participate in AMSTI. A design that randomizes within schools, such as at the class, teacher, or student level, might disrupt key components of the intervention (such as in-school supports and teacher collaboration). School-level randomization was made possible because in each of the five study regions, greater number of qualified schools applied to participate in the AMSTI program than resources could accommodate.

Matched-pairs design

A matched-pairs design was selected for school-level randomization because it can increase the precision of estimates by removing between-pair variation in the outcome as a source of error variance in the standard error of the impact estimate. However, the benefits of a matched-pairs design can be offset by a loss in degrees of freedom resulting from estimating pair effects. (For a discussion of the tradeoffs associated with using a matched-pairs design, see Bloom [2005] and Raudenbush, Martinez, and Spybrook [2007].)

School selection

Eighty-two study schools were selected from a sample of 144 eligible schools that applied to participate in AMSTI. Although schools were not selected from a random sample, researchers did select a set of applicant schools that, in aggregate, were similar to the population of eligible schools within their regions. (See appendix C for a description of the selection process.)

Recruitment, selection, and random assignment of schools

Analysis of sample sizes and statistical power

The study was designed to achieve a minimum detectable effect size of 0.20 for estimating the impact for a single student outcome. The statistical power analysis conducted at the planning stage of this experiment showed that 66 schools were required to detect an impact of at least 0.20 standard deviation with .80 power for the mathematics problem solving outcome.[20] The value of the minimum detectable effect size was chosen to be consistent with other evaluations sponsored by the National Center for Education Evaluation and Regional Assistance (see appendix D for citations). If a 20 percent attrition rate for schools is assumed, then 82 schools would be required. The estimates also suggested that a final analytic sample of 66 schools would detect an effect as small as 0.38 standard deviation for teacher outcomes.[21]

---

[20] The science outcome is assessed only in grades 5 and 7. Therefore, the student sample size for the analysis of the science outcome was assumed to be two-fifths of the student sample size for math (112 students per school instead of 280 students per school), yielding a minimum detectable effect size for science of 0.22.

[21] As shown in chapter 4, the analyses ultimately achieved minimum detectable effect size values for one-year impacts of 0.063 for mathematics and 0.126 for science. These differed from the planned minimum detectable effect size values for two reasons. First, estimates of the key power parameters differed from the values assumed in designing the study. Second, technical guidance received from outside reviewers (the Analytic and Technical Support team) after the experiment was conducted led researchers to adopt the Bonferroni correction for multiple outcomes, which reduced the effective alpha level for the analysis. Two independent adjustments were conducted. The first was carried out on analyses of impacts on student performance in mathematics problem solving and science. The alpha level was reduced from .050 to .025 for each of these tests. The second was carried out on analyses of impacts the use of active learning instructional strategies in mathematics and science classrooms. The alpha level was reduced from .050 to

Rollout of the Alabama Math, Science, and Technology Initiative (AMSTI) program and study

Since AMSTI's inception, the Alabama State Department of Education (ALSDE) has been slowly rolling out the program, offering it to a specified number of schools within newly added regions each year. Each of the 11 AMSTI regions is named after its lead institution, or AMSTI site—the educational institution that takes responsibility for teacher training, support, and distribution of AMSTI materials within that region.[22]

The study began in the 2006/07 school year, at which time ALSDE planned to introduce the AMSTI program to schools in three regions. That year, the state could afford to offer the program to a total of 20–25 new schools in these regions and to promise the control schools that they would receive the program the following year. In the 2007/08 school year, ALSDE introduced the AMSTI program in two additional regions. The state could afford to offer the program to an additional 20–25 schools in those two regions and to promise the control schools that they would receive the program in the 2008/09 school year. Because the state did not plan to introduce the program in the number of schools required for the study in a single year, this study was conducted in two phases (or subexperiments).

In February 2006, 40 schools from the first three regions were selected to participate in the study, and researchers randomly assigned 20 schools to the AMSTI condition and 20 schools to the control condition. (The process of identifying the sample of 40 schools is described in the section on selection and random assignment of schools below.) The 20 AMSTI schools began implementation in August 2006. In January 2007, a similar process was followed for the second subexperiment. Researchers selected 42 schools from the two additional regions to participate, with 21 randomly assigned to AMSTI and 21 randomly assigned to the control condition. The 21 AMSTI schools began implementation in August 2007. Data from both subexperiments were pooled and analyzed together (table 2.1).[23]

.025 for each of these tests as well. For the one-year impacts on active learning instructional strategies, the achieved minimum detectable effect size values were 0.37 for mathematics and 0.31 for science. (See chapter 4 for values assumed for the original power analysis as well as estimates of the parameters used for the sample-based power calculations.)

[22] As explained in chapter 1, the 11 AMSTI regions follow the same boundaries as the 11 regional inservice centers established by the Alabama legislature in 1984.

[23] The decision to combine data across subexperiments was made for practical reasons. If we use the parameter assumptions in the original power analysis, dividing the sample would result in two underpowered analyses (i.e., two analyses, neither of which by itself would be powered to detect impacts that are as small as expected). There is no a priori reason to expect the impact to be different for the two subexperiments. Furthermore, given that the goal of AMSTI is to improve student math and science student achievement for all Alabama students, the study was designed to cover as many regions of Alabama as possible, to make the results more generalizable to the state as a whole, and to provide relevant information to ALSDE. This further justified not dividing up the sample and running separate analyses for each group.

Table 2.1 Regions, schools, and school year of Subexperiment 1 and Subexperiment 2

Subexperiment	Number of regions	Number of schools randomized	Year 1 (school year)	Year 2 (school year)
1	3	40	2006/07	2007/08
2	2	42	2007/08	2008/09

Throughout the report, Year 1 signifies the first year of AMSTI implementation for the schools randomized to the AMSTI group in each subexperiment (2006/07 for Subexperiment 1, 2007/08 for Subexperiment 2). Year 2 signifies the second year of AMSTI implementation for the schools randomized to the AMSTI group in each subexperiment (2007/08 for the Subexperiment 1, 2008/09 for Subexperiment 2).

Recruitment, selection, and random assignment of schools

During the winter of the school year before the study began (2005/06 for Subexperiment 1, 2006/07 for Subexperiment 2), ALSDE invited 613 schools (352 from the three Subexperiment 1 regions and 261 from the two Subexperiment 2 regions) to participate in the AMSTI program; 190 schools returned the application forms (101 from Subexperiment 1 and 89 from Subexperiment 2). Schools were eligible for selection into the study if they met two criteria:

- *Program criterion.* At least 80 percent of the mathematics and science teachers signed the application form indicating that they wanted to participate in AMSTI. Eleven schools from Subexperiment 1 and eight from Subexperiment 2 did not meet this criterion.
- *Study criterion.* The school housed at least one of grades 4–8. Grades 3 and below were ineligible for the study because of lack of student pretest measures; grades 9–12 were ineligible because the AMSTI high school program uses a separate curriculum. Twenty-seven schools (16 from Subexperiment 1 and 11 from Subexperiment 2) did not meet this criterion.[24]

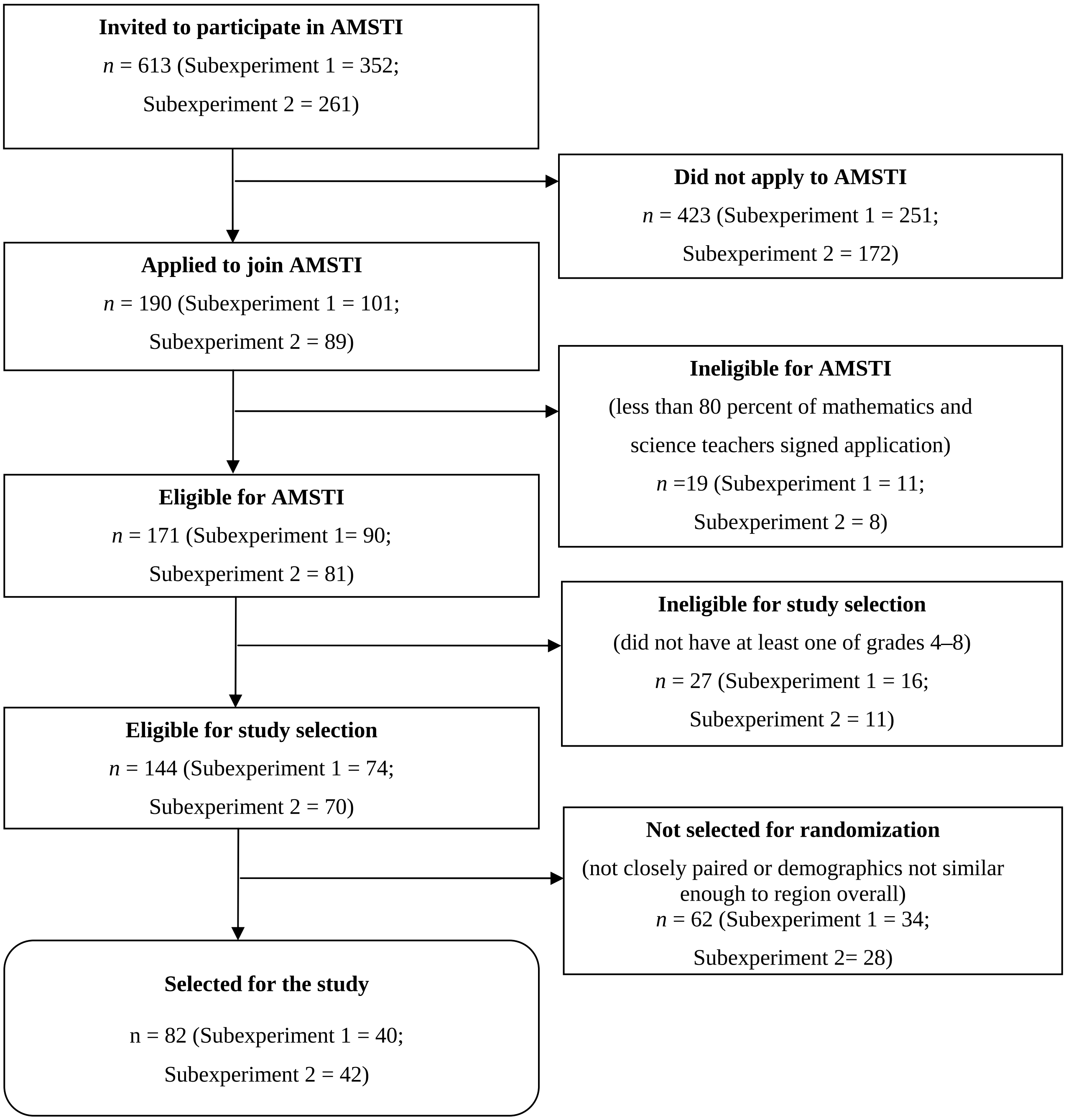
After schools that did not meet the AMSTI program and study criteria were removed, 144 schools remained (74 in Subexperiment 1 and 70 in Subexperiment 2).

From those 144 eligible schools, a sample of 40 schools was selected for Subexperiment 1 (figure 2.1). One at a time, pairs of schools were identified and selected until the desired number was reached in each region. This sample consisted of six matched pairs in one region and seven matched pairs in each of the other two regions. For Subexperiment 2, a sample of 42 schools was selected, consisting of 10 pairs in one region and 11 pairs in the other.

[24] ALSDE did admit grades K–3 into the program for schools that were admitted into the study, even though these grades did not participate in the study. Although researchers did not track admittance of non-study schools into the program, a number of schools were admitted before the selection of the study schools and randomization.

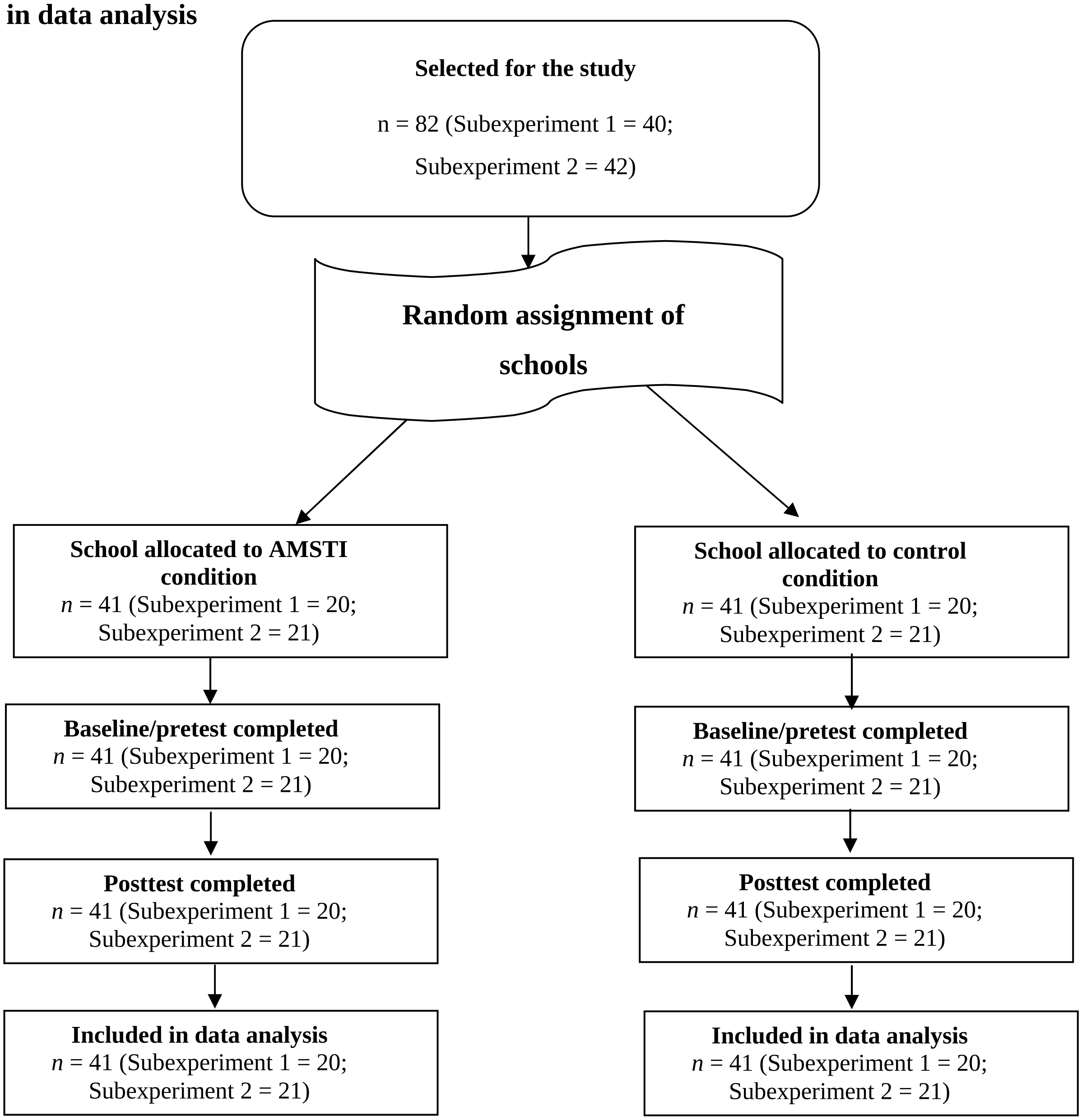
For both subexperiments, researchers met with AMSTI staff and regional site directors to select schools to participate in the study. To select the pairs that would then be randomized within each region, researchers paired schools first on the basis of similarity of grade configuration, then on mathematics scores, on percentage of racial/ethnic minority students, and finally (when possible) on percentage of students from low-income households (students enrolled in the National School Lunch Program). AMSTI regional directors provided input on the appropriateness of the pairings, based on their local knowledge of similarities that went beyond those captured in the formal criteria, and corrected or updated data. As pairs were selected, researchers used a spreadsheet to maintain a running tally of the demographics of the combined pairs in each region to ensure they were similar to the demographics of that region's schools. In some cases, this consideration was the deciding factor in choosing a school from two closely matched pairs.

Figure 2.1 Flow chart of school sample selection from invitation to participate in the Alabama Math, Science, and Technology Initiative (AMSTI) to selection for the study



Within each pair, a senior state official tossed a coin to determine assignment of each school to AMSTI or the control condition. The sample was allocated through a process that began with selection and random assignment and ended with inclusion in data analysis (Figure 2.2; see appendix C).[25]

Figure 2.2 Flow chart of school sample allocation from random assignment to inclusion in data analysis



[25] Data from all 82 schools were included in the first-year impact analysis of the SAT 10 mathematics problem solving outcome. For the numbers of schools used in other analyses, see the section on attrition rates and equivalence of baseline and analyzed sample in the Threats to Internal Validity section of Chapter 2.

After randomization, control schools continued to use their regular instruction program for mathematics and science for one year. Programs in use met the objectives specified in the Alabama Course of Study and used comprehensive and supplementary texts recommended by the Alabama State Textbook Committee. More than 50 sets of curriculum materials have been approved for mathematics and science, some of which are included in the AMSTI curriculum. Activities at control schools were not linked specifically to AMSTI, however, and were considered part of the natural evolution of pedagogy, or "business as usual."

Control schools were expected to adopt AMSTI the following school year (2007/08 for control schools in Subexperiment 1, 2008/09 for control schools in Subexperiment 2). Although two years without AMSTI would have been preferable for enabling the study design to detect the two-year effect of AMSTI, AMSTI staff believed that offering delayed acceptance into the program after one year was necessary to encourage the schools to serve as controls. A delay of more than one year would have served as a disincentive to participate in the study, as schools not participating in the study would be allowed to join AMSTI before the participating control schools. With no guarantee of continued funding for program expansion, assignment to the control group was associated with a greater risk of never being able to join the program.

Incentives to participate

The main incentive to participate in the study was the opportunity to participate in AMSTI. All mathematics and science teachers in schools assigned to the intervention group received AMSTI professional development as well as AMSTI support, materials, and technology. Schools assigned to the control group were assured that they could participate in AMSTI the following year.

Teachers were offered incentives to participate fully in both the AMSTI program and the study. Teachers attending the AMSTI summer institute on noncontract hours received stipends of $100 a day from AMSTI (which certain districts supplemented with their own funds). In addition, teachers in both the AMSTI and control group who participated in the study's web-based surveys, described below, received stipends of up to $100 from the Regional Educational Laboratory Southeast.

Defining the term "Alabama Math, Science, and Technology Initiative (AMSTI) teacher"

Because AMSTI is considered a schoolwide program and randomization was performed at the school level, all teachers who matched the subject matter and grade criteria were labeled "AMSTI teachers." However, the AMSTI program requires that at least 80 percent of the mathematics and science teachers at a school participate. Therefore, the term "AMSTI teacher" does not necessarily signify that the teacher actively implemented the AMSTI program. The percentage of Year 1 AMSTI teachers who reported participating in some AMSTI training was 86 percent in Subexperiment 1 and 87 percent in Subexperiment 2. It would not have been possible to maintain the benefits of the randomized design if nonimplementers were removed from the study sample, because there was no way to identify equivalent teachers at the control schools.[26] Furthermore, removing nonimplementers would be inconsistent with the intent-to-treat aspect of this experiment, in which the effect estimate is based on the sample of participating teachers in the intervention group who were offered (but did not necessarily use) AMSTI.

Data sources, collection, and cleaning

Data were collected from multiple sources across four school years: 2006/07, 2007/08, 2008/09, and 2009/10 (table 2.2). Summer professional development trainer logs and observations, professional development teacher surveys, classroom observations, teacher interviews, and principal interviews were collected only for Subexperiment 2, because researchers did not receive approval from the Office of Management and Budget to administer these implementation measures in Subexperiment 1. All other measures described were collected for both subexperiments. Student-level data and teacher and principal surveys collected for Year 1 of Subexperiment 1 were collected through a research grant from the Institute for Education Sciences (IES) (#R305E040031) to Empirical Education Inc., with permission from the IES program officer. Student-level outcome data were collected for two academic years for each subexperiment. This section describes all the data collected. The report analyzes only the data from the first and second year of each subexperiment.

Table 2.2 Data collection activities

Level of data	Type of data	Type of analysis
State and regional	Document search and informational interviews with key decision makers	Background information
Student	Class rosters, student demographics, and achievement measures	Primary confirmatory and exploratory analysis
Professional development trainer	Summer professional development trainer logs[a]	Implementation analysis
Teacher	Teacher interviews[a] Teacher web-based surveys (four surveys administered between January and April)	Implementation analysis, secondary confirmatory, exploratory, and implementation analysis
School	Interviews with principals[a]	Implementation analysis

a. Data were collected only for Subexperiment 2, because researchers did not receive approval from the Office of Management and Budget to administer these implementation measures in Subexperiment 1. Professional development teacher surveys, professional development observations, classroom observations, and web-based surveys of principals were also collected; however, they were not used in this report (see additional information in the program implementation data section).

[26] Although arguably the teachers who signed the original AMSTI form agreeing to participate in the program might have been considered “AMSTI teachers” for that school, the program did not focus exclusively on those teachers. Instead, the focus was on the whole school program. Because schools experience changes in teaching staff and changes in teaching assignment from year to year, the teachers who actually participated may not have signed the agreements, which were signed before implementation.

Consent process

To collect data at the state, district, and teacher levels, researchers completed multiple formal consent procedures. Before randomization, the Alabama superintendent of public education signed formal agreements to support the study and its data collection in May 2006 for Subexperiment 1 and in February 2007 for Subexperiment 2. The school district for each of the study schools then signed an agreement and appointed a point of contact to facilitate communication between researchers and district personnel, including principals.

After randomization (fall 2006 for Subexperiment 1, fall 2007 for Subexperiment 2), researchers contacted all principals by phone and email to remind them about the study, answer questions, and solicit a list of the teachers who taught the relevant subject matter and grade range during the study year. Researchers then requested that principals distribute and collect signed teacher study participation consent forms that served as the formal agreement for individual teachers to take part in the web-based surveys. In Year 1, the first year AMSTI was implemented, 99.1 percent of eligible teachers in Subexperiment 1 and 94.4 percent in Subexperiment 2 consented to participate in the web-based surveys. Non-special education mathematics and science teachers who taught grades 4–8 were considered eligible to complete the web-based surveys. Parental consent was not required; therefore, the sample included data on all eligible students.

State and regional data

To learn about the goals, design, and implementation of AMSTI at the state and regional levels, researchers reviewed ALSDE documents and conducted semistructured interviews with ALSDE staff and other state officials. Interviews took place during the summer 2006 professional development training sessions and the January 2007 training workshops provided by ALSDE. Researchers also consulted frequently with ALSDE on types of materials (including print materials) provided to teachers, the alignment of program content with state standards, expectations of and requirements for AMSTI teachers, similarity of implementation across regions and over time, and school year professional development offered to the schools. These data were not used in the analysis but to inform development of the web-based surveys and provide researchers with a comprehensive understanding of the AMSTI theory of action and program implementation expectations.

Student data

Researchers collected all available class rosters, demographic data, and student achievement data for all students in grades 4–8 in the randomized schools in each subexperiment. See appendix E for a timeline of data collection procedures used in the confirmatory and exploratory analyses. The procedures for Year 2, identical to those for Year 1, took place on about the same schedule.

Classroom rosters. Classroom rosters served two purposes. First, the data were used to populate the request submitted to ALSDE for data on student achievement and additional student demographics. Second, the data were used to connect measures of teacher and classroom characteristics to student outcomes, because the rosters linked each student to one mathematics teacher and one science teacher. (In most elementary schools, this was a single classroom teacher.) School districts provided classroom rosters for students enrolled in all regular education mathematics and science classes in grades 4–8 in study schools. The rosters contained the following data for each student: state student identification, district student identification, first and last name, grade, race/ethnicity, date of birth, gender, school name, district name, name of mathematics teacher, name of science teacher, name of mathematics course, and name of science course.

Student achievement measures and demographics. Researchers prepared and submitted a data request to ALSDE asking for student achievement and additional demographic data for all study students. ALSDE supplied the following student data requested by researchers: results of the Stanford Achievement Test Tenth Edition (SAT 10), mathematics problem solving, mathematics procedures, science, and reading subtests; results of the Alabama Reading and Mathematics Test reading and mathematics subtests; disability status; English language status; and eligibility for free or reduced-price lunch.

The SAT 10 is a norm-referenced test that compares individual and group performance with a norming group. All schools in Alabama administer the SAT 10 each April to comply with state accountability requirements. The mathematics and reading subtests are required in grades 3–8. The science subtest is required in grades 5 and 7.

The Alabama Reading and Mathematics Test is a criterion-referenced test administered in grades 3–8. It contains items from the SAT 10, as well as items developed to cover Alabama state content standards that are not tested by the SAT 10.

Before conducting the analysis of the SAT 10 mathematics problem solving, science, and reading subtests, researchers limited the number of student-level hypotheses. Reading scores were used as pretest measures for the science outcomes. Science tests are required only in grades 5 and 7; results were therefore not available for the year immediately before the outcome measure.[27] Because the Alabama Reading and Mathematics Test mathematics and the SAT 10 mathematics procedures subtests were not used in analyses or as pretests, they were not examined in this report.[28] Pretest and outcome measures were used for each subexperiment (table 2.3).

---

[27] During the design phase, researchers selected the SAT 10 reading score as the pretest measure for the science outcome. This measure was selected over the mathematics measure in part because it was hypothesized that science instruction in grades 4–8 would be more focused on process skills and science concepts than on mathematics skills. Therefore, student achievement in reading was hypothesized to be a better predictor of achievement in science in these grades. A sensitivity test was conducted to estimate the effect on science test scores of using a model that includes both the mathematics problem solving pretest and reading pretest, to examine whether the effect estimates and the corresponding *p* values are sensitive to the inclusion of an additional covariate. For the findings of this analysis, see chapter 4.

[28] As explained in appendix A, researchers hypothesized and the AMSTI coordinator confirmed that the problem solving scale was a better test of the AMSTI model than the procedures scale or the content on

Table 2.3 Student achievement data analyzed in Year 1

Subexperiment	Pretest measure	Outcome measure
1	SAT 10 mathematics problem solving achievement test scores from spring 2006	SAT 10 mathematics problem solving achievement test scores from spring 2007
	SAT 10 reading achievement test scores from spring 2006	SAT 10 science achievement test scores from spring 2007
	SAT 10 reading achievement test scores from spring 2006	SAT 10 reading achievement test scores from spring 2007
2	SAT 10 mathematics problem solving achievement test scores from spring 2007	SAT 10 mathematics problem solving achievement test scores from spring 2008
	SAT 10 reading achievement test scores from spring 2007	SAT 10 science achievement test scores from spring 2008
	SAT 10 reading achievement test scores from spring 2007	SAT 10 reading achievement test scores from spring 2008

Note: For data used in the exploratory analysis of the two-year effects of AMSTI on mathematics problem solving and science, testing dates occurred the following year.

Program implementation data

Data from summer professional development training logs and interviews with teachers and principals (from Year 1 of Subexperiment 2) were used to inform implementation analyses (chapter 3). Data from teacher surveys were used to inform implementation analyses (chapter 3), confirmatory analyses (chapter 4), and exploratory analyses (chapter 6) that focus on Year 1 (that is, the first year of AMSTI implementation for the schools randomized to the AMSTI group in each subexperiment). Data from these sources were used to:

- Assess, with descriptive statistics, the extent to which the intervention components of AMSTI (summer professional development, access to equipment and materials, and in-school support) were implemented in Year 1.
- Assess, with inferential statistics, differences between the AMSTI and control conditions in the extent to which the three main intervention components are present in schools in Year 1.
- Address the second confirmatory research question regarding the effect of one year of AMSTI on teachers' use of active learning instructional strategies and address the exploratory research questions pertaining to the effect of one year of AMSTI on teacher content knowledge and student engagement.

Data from teacher surveys from Year 1 and Year 2 were also used to check the assumptions behind the Bell-Bradley methodology (discussed later in this chapter), which was used in an exploratory analysis of the effect of two years of AMSTI. Data from professional development teacher surveys, professional development observations, classroom observations,

---

the Alabama Mathematics Test.

and principal surveys were not used in the analyses; the purposes and details of these data are explained in appendix AG.

Professional development training logs. Professional development training logs were completed daily to reduce recall bias and obtain cost-effective, first-person accounts of the topics and instructional methods used by AMSTI trainers over the course of the institutes. Before the AMSTI summer institute in Year 1 of Subexperiment 2, researchers provided all AMSTI trainers of grades 5 and 7 mathematics and science with training logs and asked them to complete the daily logs after each session. At the end of the summer institute, the AMSTI site coordinator collected the logs and mailed them to the research staff. For grade 5 mathematics and science, there were 5 logs per trainer (5 training days); the grade 7 mathematics and science training courses were 10 days each, yielding 10 logs per trainer.

Twelve trainers taught grade 5 or 7 mathematics or science in the AMSTI summer institute (four grade 5 mathematics, four grade 5 science, two grade 7 mathematics, and two grade 7 science). All agreed to participate in the evaluation without compensation. All 12 trainers completed all the daily logs provided. The final sample for this data source thus consisted of 80 logs from 12 AMSTI trainers.

Training logs asked trainers to rate how extensively the topics and content areas in the AMSTI curriculum were covered, based on a 5-point Likert scale (not at all, limited extent, moderate extent, large extent, full extent). Trainers were also asked to use a 5-point Likert scale to record the percentage of time spent on teaching methods (none, 1–25 percent, 26–50 percent, 51–75 percent, 76–100 percent). Two open-ended questions prompted trainers to reflect on the most effective part of the training that day and on anything they would change. The logs also collected information about the trainers: years of classroom experience, teaching experience in an AMSTI school, overall experience as a trainer in mathematics or science, and prior experience as an AMSTI trainer.

Professional development training logs were collected for Years 1 and 2 (of Subexperiment 2 only). Data from Year 1 logs were used to assess the extent to which the intervention components of AMSTI were implemented in Year 1 (see chapter 3).

Teacher interviews. Interviews were used to learn about AMSTI teachers' experiences in implementing AMSTI in their classrooms and control teachers' overall experiences in math and science instruction. Researchers selected Subexperiment 2 AMSTI and control teachers in seven grade/subject-level strata to participate in on-site interviews. The seven strata included mathematics teachers in grades 4–8 and science teachers in grades 5 and 7. Researchers randomly selected two AMSTI teachers from any of the strata in each of the 21 AMSTI schools and asked them to participate, and all agreed. This process was repeated for all 21 schools, chosen in random order, until two teachers had been selected from each AMSTI school. Once six teachers for a particular stratum had agreed to participate, no more teachers were selected for that stratum. This selection process was repeated for teachers in all 20 control schools. If a selected teacher was not available (for example, on leave, on vacation, or retired), researchers randomly selected another teacher for that school from that strata. In two cases, there were no more AMSTI teachers within a stratum within a school. In these cases, no teacher interview was conducted. The final sample for the teacher interviews consisted of 40 teachers in 21 AMSTI schools and 41 teachers in 20 control schools.

The interviews, lasting about 20–30 minutes, were conducted as soon as possible following a classroom observation of the teacher. (Details of classroom observations are in appendix F.) Two researchers attended the interview—one conducting the interview and the other taking notes. Interviewers were instructed to adhere strictly to the interview protocol.

The interview protocol for AMSTI teachers included questions about teachers' use of the AMSTI program and materials and their perspectives on their experiences implementing AMSTI, students' responses to AMSTI, and any additional training or assistance needed to implement AMSTI. The protocols for interviews with control school teachers were similar but referenced general science and mathematics materials, curricula, and instruction. Both AMSTI and control interview protocols included questions that followed up on the classroom lesson observed by the researchers.

Teacher interviews were conducted in Years 1 and 2 (of Subexperiment 2 only). Data from Year 1 AMSTI teacher interviews were used to assess the extent to which the intervention components were implemented in Year 1 (see chapter 3).

Interviews with principals. Interviews were conducted with AMSTI principals to obtain the school leadership perspective on the AMSTI initiative; interviews were conducted with control principals to obtain the school leadership perspective on mathematics and science instruction. Researchers scheduled these interviews to follow the classroom observations and teacher interviews to facilitate discussion with the principal about mathematics and science instruction in the school. Interviews were conducted at all 21 Subexperiment 2 AMSTI schools and at 19 control schools.

The interviews lasted about 20 minutes and were conducted by two researchers—one conducting the interview and the other taking notes. Interviewers were instructed to adhere strictly to the interview protocol. All 21 AMSTI principals contacted agreed to participate in the interviews. Nineteen control principals contacted agreed to participate in the interviews (one control principal did not respond to schedule an interview). Principals were not compensated for their participation.

Interviews with AMSTI and control principals allowed for open-ended answers to questions about the principal's role in the school's mathematics and science instruction, the training the principal received, and technical assistance and coaching for teachers on mathematics and science instruction. AMSTI principals were also asked about the extent of AMSTI implementation in the school, teachers' preparedness to teach AMSTI, and the availability of AMSTI materials.

Interviews with principals were conducted in Years 1 and 2 (of Subexperiment 2 only). Data from Year 1 AMSTI teacher interviews were used to assess the extent to which the intervention components were implemented in Year 1 (see chapter 3).

Teacher surveys. Web-based teacher surveys are a cost-effective way to gather information on professional development, instruction, support, and materials. The four surveys were developed by Empirical Education researchers, based on the information needs of ALSDE. The items were adapted from items used in previous Empirical Education studies. The surveys were reviewed for content validity by program experts at ALSDE, including the AMSTI director, the math coordinator, and regional site directors. The four surveys were not identical; however, some questions from specific survey domains were repeated in all four surveys in order to 1) calculate averages of those domains over time, and 2) reduce measurement error by averaging responses over multiple occasions. This allowed researchers to construct more reliable measures of teacher practices than could have been constructed from any one survey.

Teachers from both Subexperiment 1 and Subexperiment 2 who submitted signed consent forms received an e-mailed survey invitation each month, from January through April during Year 1 and Year 2. Nonrespondents received first an e-mail, then a fax, and then telephone calls to achieve acceptable response rates. If needed, teachers had the option of responding by means of a paper survey. Teachers who completed all four surveys received a yearly stipend of no more than $100.[29] Response rates ranged from 83 to 96 percent (table 2.4).

Table 2.4 Year 1 teacher survey response rates

Survey	Overall		AMSTI		Control	
	Subexperiment 1	Subexperiment 2	Subexperiment 1	Subexperiment 2	Subexperiment 1	Subexperiment 2
1 (January)	306/320	236/254	161/169	122/129	145/151	114/125
	(95.6)	(92.9)	(95.3)	(94.6)	(96.0)	(91.2)
2 (February)	304/320	228/254	163/169	119/129	141/151	109/125
	(95.0)	(89.8)	(96.5)	(92.3)	(93.4)	(87.2)
3 (March)	298/320	216/254	160/169	107/129	138/151	109/125
	(93.1)	(85.0)	(94.7)	(83.0)	(91.4)	(87.2)
4 (April)	295/320	218/254	158/169	109/129	137/151	109/125
	(92.2)	(85.8)	(93.5)	(84.5)	(90.7)	(87.2)

Note: Numbers in parentheses are percentages.
Source: Teacher survey data.

Teacher surveys were conducted in Year 1 and Year 2. They included the same questions in both years in the following domains:

- Professional development (types, frequency, impact on learning).
- Instructional time.
- Student assessment.
- Technology (availability, support, comfort).
- Teacher background.

[29] Teachers received stipends for completing fewer than four surveys if, for example, they were on leave for part of the year.

- Equipment and materials (availability, use, satisfaction).
- Instructional strategies (inquiry, hands-on, higher-order thinking skills).
- Planning.
- Collaboration and support.
- Student engagement.
- Self-rating of teacher content knowledge and implementation.

Only some of these domains pertained to the analyses addressed in this report. Relevant domains included questions about the three main intervention components (summer professional development, access to equipment and materials, and in-school support) during Year 1; classroom instructional strategies (used in the secondary confirmatory analyses) during Year 1; and reported levels of student engagement and teacher content knowledge (used in the exploratory analyses) during Year 1. Data from teacher surveys from Year 2 (2007/08 for Subexperiment 1, 2008/09 for Subexperiment 2) were used to check the assumptions behind the Bell-Bradley methodology for estimating the impact of two years of AMSTI participation (discussed later in this chapter).

The first teacher survey asked both AMSTI and control teachers to report the number of hours of mathematics and science professional development they had received during the summer before the school year. The third survey measured AMSTI and control teachers' reported access to materials and equipment in their classrooms. Surveys two, three, and four, asked teachers the number of times they requested support for instruction and the number of times they received support for instruction.[30]

The teacher survey data were the basis for addressing the secondary confirmatory research question. The theory of action hypothesized that, after receiving the professional development, teachers would change their classroom practice, using the most effective strategies to emphasize hands-on, inquiry-based learning. In particular, the AMSTI program focuses on three types of instruction:

- *Inquiry-based instruction.* This is defined as having students do all of the following activities as part of the learning process: make observations; pose questions; examine books and other sources of information to see what is already known; plan investigations; review what is already known in light of experimental evidence; use tools to gather, analyze, and interpret data; propose answers, explanations, and predictions; and communicate the results.
- *Hands-on instruction.* This is defined as having students participate in activities involving active participation and application, as opposed to a theoretical discussion.
- *Instruction using higher-order thinking skills.* This requires students to advance from skills such as focusing and information gathering to integrating and evaluating.

---

[30] Access to support provided through the AMSTI program was hypothesized to result in AMSTI teachers both requesting and receiving more support than control teachers.

The use of these strategies has been shown to support development of higher-order critical thinking and processing skills and to engender positive attitudes toward mathematics and science (AMSTI Committee 2000; Anderson and Krathwohl 2001; Bonwell and Eison 1991; Haury 1993). On each survey, teachers reported the number of minutes students spent on each of the three types of mathematics and science instruction during the previous two weeks of instruction. These data were used to construct the composite variable for active learning as the classroom practice outcome (see section on data analysis methods later in this chapter; see appendix G for a copy of the survey). The definitions of each type of instruction (as defined above) were provided in the surveys. Internal consistency reliability of the active learning score was calculated using Cronbach's alpha (see results in appendix L). Additional reliability checks were not explicitly conducted because the questions were low-inference questions concerning the amount of time spent rather than questions about complex constructs underlying instructional strategies.

Teacher content knowledge and student engagement. One of the primary objectives of the AMSTI professional development is to increase teacher content knowledge in the subjects they teach. Therefore, on the final teacher survey of the year, both AMSTI and control teachers were asked to rate their content knowledge for teaching mathematics or science at their current grade level.[31] Teachers responded on a 5-point Likert scale (very low, low, moderate, high, very high), with a sixth option for "not applicable." Teacher perceptions of student engagement were also solicited in this survey, because AMSTI is expected to increase student engagement as students shift to more hands-on, inquiry-based learning. Teachers were asked to rate the average level of student engagement during the school year on a 5-point Likert scale (not engaged, slightly engaged, moderately engaged, almost fully engaged, fully engaged).[32] The question included instructions to respondents that students should be considered fully engaged if they paid full attention, participated fully, and completed all assignments.

Measures of teacher content knowledge and student engagement are from teacher self-reports, which may be susceptible to response bias. These outcomes are nevertheless important to investigate, because they are part of AMSTI's theory of action, prior research has found them to be related to student achievement, and they may suggest areas for further research.

Data cleaning procedures

A description of the data cleaning procedure and construction of the data files for analysis appears in appendix H.

---

[31] Survey questions were asked separately for mathematics and science teachers.
[32] Survey questions were asked separately for mathematics and science teachers.

Threats to internal validity

This section defines the baseline and analytic samples used in the one-year confirmatory impact analysis of pooled data from Subexperiment 1 and Subexperiment 2; describes differential attrition rates and presents the baseline and analytic samples of students, teachers, and schools; examines whether random assignment resulted in statistically equivalent groups by comparing AMSTI and control groups according to a set of background characteristics; and describes potential spillover effects of the intervention. The same type of information about the exploratory analyses is presented later in the data analysis methods section.

As described in detail below, at the school level, differential attrition was 5 percentage points or less; overall attrition was 2.5 percent or less for the four confirmatory outcomes. Two statistically significant differences were found between the AMSTI and control groups on a set of covariates for the baseline and analytic samples.[33] The baseline sample association with the SAT 10 science outcome differed only by gender; the analytic sample associated with the active learning in mathematics outcome differed only by teacher degree rank. There was no indication of nonequivalence on the remaining covariates for either the baseline or analytic samples associated with the four confirmatory outcomes. The integrity of the samples, the achievement of adequate statistical power, and the internal validity of the impact inferences have all been confirmed.

Defining the baseline and analytic samples used in the confirmatory analysis

Using the student classroom rosters received by the district, researchers identified the baseline sample for the student achievement outcomes, which included all students in grades 4–8 for mathematics and all students in grades 5 and 7 for science who were not formally designated by the district as having a disability.[34] To identify the analytic sample, researchers then tracked the loss of data caused by students moving between subexperiments, students' data that were missing a student or school identifier, and students who missed posttests.

The baseline sample for the active learning outcomes included all teachers in appropriate grades and subjects who completed surveys. One of the 82 randomized schools (a control school) withdrew from the study the day after randomization. This school was excluded from the survey sample and from the analysis of both secondary confirmatory outcome measures (active learning for mathematics teachers and active learning for science teachers). To identify the analytic

[33] In the analysis of confirmatory outcomes, 90 tests of equivalence of background characteristics were conducted, for both the baseline and analytic sample: 18 characteristics for the SAT 10 mathematics problem solving sample; 13 characteristics for the SAT 10 science sample; 7 characteristics for the active learning in mathematics outcome; and 7 characteristics for the active learning in science outcome. Of the 90 tests, the null hypothesis was rejected twice. The number of observed rejections can reasonably be accounted for by chance.

[34] As noted in the section on student-level data, the classroom rosters for study schools included data for students enrolled with all regular education mathematics and science teachers in grades 4–8, including students with disabilities, in study schools.

sample, researchers then tracked the loss of data caused by missing teacher/school identifiers, and missing valid active learning scores.

Researchers also tracked the number of schools, teachers, and students through the course of the study (see appendix I) and compared attrition rates in the AMSTI and control conditions for the student- and classroom-level outcomes.

Baseline and analytic sample and rate of sample attrition for student-level outcomes. For the sample associated with the SAT 10 mathematics problem solving outcome, there was no attrition at the school level (table 2.5). At the teacher level, differential attrition was 0.8 percentage point, and overall attrition was 0.4 percent. At the student level, differential attrition was 0.1 percentage point, and overall attrition was 4.7 percent.

For the sample associated with the SAT 10 science outcome, one school was lost from the control condition (table 2.6). Differential attrition from the baseline to the analytic sample at the school level was 2.4 percentage points, and overall attrition at the school level was 1.3 percent. At the teacher level, differential attrition was 4.3 percentage points, and overall attrition was 3.0 percent. At the student level, differential attrition was 2.3 percentage points, and overall attrition was 7.8 percent.

Table 2.5 School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome after one year

	Schools			Teachers			Students		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment, from rosters	41	41	82	249	233	482	12,065	10,492	22,557
Baseline (eligible) sample	41	41	82	246	229	475	10,517	9,109	19,626
Analytic sample	41	41	82	244	229	473	10,022	8,691	18,713
Attrition from baseline (eligible) to analytic sample	0	0	0	2 (0.8)	0	2 (0.4)	495 (4.7)	418 (4.6)	913 (4.7)

Note: Numbers in parentheses are percentages.
Source: Student achievement data from tests administered as part of the state's accountability system.

Table 2.6 School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) science outcome sample after one year

	Schools			Teachers			Students		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment, from rosters	41	41	82	233	213	446	12,065	10,492	22,557
Baseline (eligible) sample	39	41	80	103	95	198	4,480	3,688	8,168
Analytic sample	39	40	79	102	90	192	4,082	3,446	7,528
Attrition from baseline (eligible) to analytic sample	0	1 (2.4)	1 (1.3)	1 (1.0)	5 (5.3)	6 (3.0)	398 (8.9)	242 (6.6)	640 (7.8)

Note: Numbers in parentheses are percentages.
Source: Student achievement data from tests administered as part of the state's accountability system.

Rate of sample attrition for classroom-level outcomes. For the sample associated with the active learning in mathematics classroom outcome, there was no attrition at the school level (table 2.7). At the teacher level, differential attrition was 2.7 percentage points, and overall attrition was 4.9 percent.

Table 2.7 School and teacher attrition associated with active learning in mathematics outcome after one year

	Schools			Teachers		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment	41	41	82	na	na	na
Baseline (eligible) sample	41	40	81	221	205	426
Analytic sample	41	40	81	213	192	405
Attrition from baseline (eligible) to analytic sample	0	0	0	8 (3.6)	13 (6.3)	21 (4.9)

na is not applicable.
Note: Numbers in parentheses are percentages. The baseline was the first point for which information from teachers was available that allowed equivalence tests to be conducted.
Source: Teacher survey data.

For the sample associated with the active learning in science classroom outcome, two control group schools were lost from the analytic sample, leading to differential attrition of 5.0 percentage points and overall attrition of 2.5 percent (table 2.8). At the teacher level, differential attrition was 4.5 percentage points, and overall attrition was 6.6 percent.

Table 2.8 School and teacher attrition associated with active learning in science outcome after one year

	Schools			Teachers		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment	41	41	82	na	na	na
Baseline (eligible) sample	40	40	80	203	192	395
Analytic sample	40	38	78	194	175	369
Attrition from baseline (eligible) to analytic sample (percent)	0	2	2	9	17	26
		(5.0)	(2.5)	(4.4)	(8.9)	(6.6)

na is not applicable.
Note: Numbers in parentheses are percentages. The baseline was the first point for which information from teachers was available that allowed equivalence tests to be conducted.
Source: Teacher survey data.

Year 1 equivalence of the confirmatory baseline and analytic samples

Tests were run to determine whether random assignment resulted in statistically equivalent groups at baseline and whether there continued to be equivalence with the analytic sample. For the samples associated with the confirmatory analysis on student achievement after one year, the equivalence between the AMSTI and control schools was examined on the following teacher and student characteristics:

Teacher characteristics:

- Proportion teaching out of field.
- Proportion with one degree in teaching content area.
- Proportion with two or more degrees in content area.[35]
- Proportion with less than four years of teaching experience.
- Proportion with less than four years of teaching experience in subject area.
- Distribution of teacher degree rank.

[35] The proportion of teachers within each category of the degree rank variable (out-of-field teaching, teachers with one degree in teaching content area, teachers with two or more degrees in the content area) was examined. The variable was created to categorize teachers' postsecondary major and minor degrees based on how closely they matched their current teaching assignment. For elementary teachers, the degree rank was based on the presence or absence of at least one degree in elementary education; for secondary teachers, the degree rank was based on whether teachers had a degree in mathematics or science content. See appendix J for descriptions of degree rank.

Student characteristics:

- School average pretest score.
- Proportion of boys.
- Proportion of racial/ethnic minority students.
- Proportion of English-proficient students.
- Proportion of students enrolled in the free or reduced-price lunch program.
- Proportion of students at each grade level (grades 4–8 for mathematics outcome and grades 5 and 7 for science outcome).
- Distribution of students across grade levels (grades 4–8 for mathematics).

For the samples associated with the confirmatory analysis of active learning outcomes after one year, the equivalence between the AMSTI and control schools was tested on the following teacher characteristics:

- Proportion teaching out of field.
- Proportion with one degree in teaching content area.
- Proportion with two or more degrees in teaching content area.
- Proportion with less than four years of teaching experience.
- Proportion with less than four years of teaching experience in subject area.
- Distribution of teacher degree rank.

For the baseline and analytic samples for each outcome, a joint significance test was conducted of all covariates combined to see whether there was an overall difference between conditions on the background variables.

Of the characteristics measured on the baseline sample associated with the SAT 10 science outcome, there was a statistically significant difference between the two groups for gender only: there were more boys in the control schools (51.13 percent) than in AMSTI schools (48.10 percent; $p = .02$). With 13 different equivalence tests, it is not unusual to find one statistically significant difference by chance alone. There were no statistically significant differences for any other background characteristics measured for the baseline samples of the four Year 1 outcomes (SAT 10 mathematics problem solving, SAT 10 science, active learning in mathematics, and active learning in science).

Of the characteristics measured on the analytic sample associated with the active learning outcome for mathematics, there was a statistically significant difference between the two groups for the distribution of teachers' degree rank ($p = .04$). There were no statistically significant differences between AMSTI and control schools for any other background characteristics measured on the analytic samples of the Year 1 outcomes. Tables displaying the full results of the equivalence tests of the baseline and analytic samples for all four confirmatory outcomes are in appendix K.

Spillover of the intervention

Spillover of the intervention, sometimes referred to as *control group crossover*, *contamination*, or *intervention diffusion*, can occur in an experiment when a control group is exposed to some of an intervention's elements. In this evaluation, the whole school was randomly assigned, and only AMSTI schools received AMSTI training and materials. However, given teacher mobility and the fact that AMSTI is a statewide initiative that has existed since 2002, teachers in control schools may have been exposed to AMSTI.[36] Of the 273 control teachers that completed the web-based surveys in Year 1, 15 mathematics teachers and 9 science teachers reported receiving AMSTI professional development during the study period.[37] In addition, 12 mathematics and 23 science teachers in the control group reported using AMSTI print materials during instruction during the study period.[38] In total, 48 control group teachers reported that they received AMSTI professional development, used AMSTI print materials, or both. Exposure of control teachers to AMSTI is likely a combination of their having become familiar with the program in the years before the study and possible spillover during the study. These two mechanisms of exposure could not be distinguished.

The principals of 17 AMSTI schools and 13 control schools indicated in the principal survey that members of their staff had participated in the Leadership Academy for Math, Science, and Technology (LAMST) in the year before the beginning of this study (2005/06 for Subexperiment 1, 2006/07 for Subexperiment 2).[39] As described in chapter 1, the LAMST training and the materials provided to the school teams (one mathematics and one science teacher in kindergarten through grade 5) are modeled after AMSTI. A chi-square test conducted on the proportion of AMSTI and control schools participating in LAMST did not reveal statistical significance (p = .56). Hence, prior exposure to LAMST did not create an AMSTI–control group imbalance, although it may suggest that the effect of AMSTI is somewhat weaker than shown, as some of the indicated effects may be attributable to LAMST rather than AMSTI.

Data analysis methods

This section describes the data analysis methods used to obtain effect estimates. It also describes the analyses used to examine the sensitivity of the results to differences in the specifications of the models used to measure effects and presents the data analysis methods used to address the exploratory research questions.

---

[36] Information on teacher and student mobility was not available. The impact analyses classified teachers and students as being in the schools to which they belonged and the conditions to which they were assigned at the time of randomization.

[37] A total of 21 control group teachers reported receiving AMSTI professional development (three teachers reported receiving AMSTI professional development in both mathematics and science).

[38] A total of 31 control group teachers reported using AMSTI print materials (four teachers reported using AMSTI print materials in both mathematics and science).

[39] Data from the principal survey from four control schools were missing.

Analysis of the effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement in mathematics problem solving and science after one year

A two-level hierarchical linear regression model (Raudenbush and Bryk 2002) was used to estimate the effects of AMSTI on student performance in mathematics problem solving and science. (See chapter 4 for results.) Students were modeled at Level 1 and schools at Level 2. The student-level covariates were pretest score, grade level, racial/ethnic minority status, eligibility for free or reduced-price lunch, proficiency in English, and gender. The pretest was decomposed into the school average pretest (a Level-2 variable) and the difference between the student score and the school average pretest (a Level-1 variable). The mathematics problem solving pretest was used as a covariate to reduce unexplained variance in the mathematics problem solving outcome. The reading pretest score was used for the science outcome, because no science pretest was available.

To account for the random assignment design, the model also included indicator variables that identified the matched pairs of schools that were randomized. It also accounted for students being nested in schools, recognizing that outcomes for students in the same school are expected to be more similar than outcomes for students in different schools. Multilevel models allowed the error structures to be correlated and to account for the effects of clustering in the estimation process. This resulted in more accurate standard errors for the effect estimates. The parameter estimate of interest is the coefficient on the school-level intervention indicator, which provided the estimate of the effect of AMSTI.

A dummy variable method was used to control for potential bias in the effect estimate arising from missing covariate values. A dummy variable was created for each covariate in the model and assigned a value of 1 if the value was missing for any student and 0 otherwise. The missing values from the original variable were replaced with 0. Puma, Olsen, Bell and Price, (2009) made the case that in the context of this type of evaluation, randomization ensures that the treatment indicator is, in expectation, uncorrelated with other independent variables, an important precondition for the dummy variable method to work. In fact, the independence of the treatment indicator and the covariates depend on the patterns of missing data. Puma et al. (2009) addressed this problem through a simulation study in which they compared levels of bias in the impact estimate and the standard error of the impact estimate under different scenarios, including a scenario in which missing data depended on membership in the treatment group. Where data are missing predominantly at the student level rather than the school level, as in this experiment, within the constraints and assumptions of the model used to carry out the simulation, the dummy variable method yielded effect estimates with less bias than the tolerance threshold set by the What Works Clearinghouse (described in Puma et al. 2009). The method fared no worse and in some cases performed better than other standard approaches, including case deletion, nonstochastic, and several stochastic regression imputation methods.

The SAT 10, used to assess mathematics problem solving and science achievement, is vertically scaled, meaning that outcomes from different grades are measured along a common scale and can be compared meaningfully without having to be rescaled or normalized. Grades were combined in the analysis; effects were reported in the metric of the test.

The MIXED procedure (SAS Institute 2006) was used to estimate the result. (See Singer 1998 for a description of the procedure for conducting mixed-model analyses of hierarchical datasets using Statistical Analysis Software.) The model assumes a constant intervention effect of AMSTI but includes random school effects.[40]

Level 1 (student level):

$$y_{ij} = \beta_{0j} + \sum_{p=1}^{12} \beta_{pj} COV_{pij} + e_{ij}$$

where y_{ij} is the outcome for student i in school j, and COV_{pij} is the value of the covariate p. There are 12 covariates. The first six represent attributes of students:

- The difference between the student's pretest score and the school mean of the pretest.
- Racial/ethnic minority status (coded 0 for White students and 1 for minority students).
- Free or reduced-price lunch status (coded 0 if the student was not enrolled in the free or reduced-price lunch program and 1 otherwise).
- Proficiency in English (coded 0 if the student was not proficient in English and 1 otherwise).
- Gender (coded 0 for girls and 1 for boys).[41]
- Member of grade 5 (coded 0 if the student was in grade 7 and 1 if a student was in grade 5). Only these grades were included in the analysis of science outcomes; the reference grade was grade 7.

These covariates are consistent with those used in several educational impact studies (see Garet et al. 2010; James-Burdumy et al. 2009; James-Burdumy et al. 2010). They were included in the impact model to obtain additional precision (based on findings in Bloom, Richburg-Hayes, and Black 2007). Several of the covariates were selected because they represent important designations in the No Child Left Behind Act of 2001, which requires that states disaggregate student achievement data for specific subgroups of students, including students from major racial/ethnic groups, students with limited English proficiency, students from economically disadvantaged households, and girls. (These covariates also allow a straightforward extension of the model to examine differential impacts for the subgroups indicated by the covariates, described in chapter 6.)

[40] The model presented was used to estimate the effect of AMSTI on student performance in science. The effect on mathematics problem solving was estimated using a similar model. It contained a larger number of fixed effects, because grades 4–8 were included in that analysis. Four dummy variables were used to indicate grade, with grade 6 serving as the reference grade.

[41] The majority of students in the study were either Black (39 percent) or White (57 percent); 56% were enrolled in the free or reduced-price lunch program; 98% were proficient in English, and 49% were boys (percentages are from the analytic sample associated with the SAT 10 mathematics problem solving outcome after one year).

The next six covariates are dummy variables that correspond to the six attribute variables listed above.[42] The dummy variable that corresponds to a given covariate indicates whether the value of that covariate is missing. If the value is missing, the dummy variable takes a value of 1; otherwise, the value is 0.

Level 2 (school level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\overline{X}_j + \gamma_{02}T_j + \sum_{m=3}^{42}\gamma_{0m}I_{(m-3)} + u_{0j} \qquad p = 1,\ldots,12$$

$$\beta_{pj} = \gamma_{p0}$$

where the other variables are defined as follows:

- $\overline{X}_j$ is the school mean of the pretest.
- T_j indicates whether a school is assigned to the AMSTI or the control condition (coded 1 if the school was assigned to AMSTI and 0 if the school was assigned to control).
- $I_{(m-3)}$ indicates the matched pair to which a school belongs. It takes on a value of 0 or 1. There are 40 indicators for the 41 pairs. The effect γ_{0m} represents the average difference in outcome between pair *m* and the reference pair, controlling for the other effects in the model.
- u_{0j} is the random effect of school *j*, conditioning on the other effects in the model.
- e_{ij} is the random effect associated with student *i* in school *j*, conditioning on the other effects in the model.

Analysis of the effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on classroom practice after one year

Data from the teacher web-based survey were used to create the active learning outcome for the second research question regarding the effect of AMSTI on the use of active learning instructional strategies by mathematics teachers and by science teachers after one year. The results are presented in chapter 4.

A teacher's active learning score was calculated by summing the teacher's responses to the three survey items that asked about time spent on inquiry-based instruction, hands-on instruction, and instruction for higher-order thinking skills within the previous two weeks (table 2.9). A response was considered valid if the number of minutes reported for each instructional

[42] If there were no missing values for a given covariate, the corresponding dummy variable was not included in the model. (This holds for all analytic models that utilize the dummy variable approach to handling missing data in this report.)

method was equal to or less than the total instructional time the teacher reported and the survey was completed within two weeks of being administered, so that the time periods of focus between surveys did not overlap. The mean of these responses was then calculated over the 12 survey items (3 items per survey multiplied by 4 surveys). If a teacher was missing a response to one or more items on any of the surveys, the mean was calculated without the response. If the survey included no valid responses, it was treated as missing and the teacher was removed from the analysis of the outcome. As appendix I (tables I3 and I4) indicates, data from 21 mathematics teachers (4.9 percent of available cases for mathematics) and 26 science teachers (6.6 percent of available cases for science) were lost because the teachers had no valid data to contribute to their active learning score.

Table 2.9 Distribution of teachers by number of valid responses to active learning score items in mathematics and science

	1–3 items			**4–6 items**			**7–9 items**			**10–12 items**		
Content area	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Number of Mathematics Teachers	22	23	45	35	27	62	55	59	114	101	83	184
Percent of Mathematics Teachers	(10.3)	(12.0)	(11.1)	(16.4)	(14.1)	(15.3)	(25.8)	(30.7)	(28.2)	(47.4)	(43.2)	(45.4)
Number of Science Teachers	15	18	33	44	30	74	55	56	111	80	71	151
Percent of Science Teachers	(7.7)	(10.3)	(8.9)	(22.7)	(17.1)	(20.1)	(28.4)	(32.0)	(30.1)	(41.2)	(41.0)	(41.0)

Note: Numbers in parentheses are percentages. Numbers may not sum to 100 percent because of rounding. The total analytic sample size for active learning in mathematics outcome was 405 teachers (213 AMSTI teachers and 192 control teachers); 142 teachers (35.1 percent of the total analytic sample) had valid responses for all 12 items. The total analytic sample for active learning in science outcome was 369 (194 AMSTI teachers and 175 control teachers); 119 (32.3 percent of the total analytic sample) had valid responses for all 12 items.
Source: Teacher survey data.

Correlations among inquiry-based instruction, hands-on instruction, and instruction for higher-order thinking skills in the active learning in mathematics and active learning in science scales were examined, along with their factor loadings, separately for AMSTI and control schools.[43] The correlations were 0.52 or higher and statistically significant for both groups for both mathematics and science (see appendix L).

For active learning in mathematics for AMSTI schools (n = 213), factor loadings were 0.92 for inquiry-based instruction, 0.91 for hands-on activities, and 0.91 for teaching for higher-order thinking skills. For the control schools (n = 192), factor loadings were 0.86 for inquiry-based instruction, 0.87 for hands-on activities, and 0.80 for teaching of higher-order thinking skills. For active learning in science in AMSTI schools (n = 194), factor loadings were 0.92 for

---

[43] These tests were performed separately by condition to preclude the possibility that the correlations and factor loadings reflected an impact of AMSTI.

inquiry-based instruction, 0.91 for hands-on activities, and 0.87 for teaching for higher-order thinking skills. For control schools (n = 175), factor loadings were 0.91 for inquiry-based instruction, 0.94 for hands-on activities, and 0.95 for higher-order thinking skills. These results suggest that a single latent dimension, active learning, accounted for the variance in outcomes for all three items. The three items were therefore combined into a single measure of the latent variable. The criterion for a high factor loading is arbitrary, but 0.35 is commonly used as a minimum cutoff. According to Hair, Anderson, Tatham, and Black, (1998), factor loadings are considered high if they exceed 0.60. The factor loadings for both scales exceeded this criterion.

Hierarchical linear regression models were used to estimate the effects of AMSTI on classroom practice measures of active learning (separately for mathematics and science). The models used to estimate the effects parallel the ones described above for estimating the effects on students. To adjust for random imbalances between AMSTI and control groups, the model included the following teacher-level covariates in the model: degree rank (see appendix J), total years of teaching experience, and total years of experience teaching the subject. To account for the random assignment design, the model also included indicator variables that identified the matched pairs of randomized schools. Because teachers were nested in schools, multilevel models were used to allow the error structures to be correlated. The dummy variable method, described in the previous section, was used to address missing values for covariates. For each analysis, the p-value that corresponds with a two-tailed test of the hypothesis that the effect is zero was reported. The analyses from which the regression-adjusted effect estimates were obtained are considered confirmatory, albeit of secondary importance.

The effects of AMSTI on classroom practice outcomes were estimated using a two-level model, with teachers at Level 1 and schools at Level 2. The model assumed a constant intervention effect of AMSTI but allowed for random school effects.

Level 1 (teacher level):

$$y_{ij} = \beta_{0j} + \sum_{p=1}^{6} \beta_{pj} COV_{pij} + e_{ij}$$

where y_{ij} is the outcome for teacher i in school j, COV_{pij} is the value of the covariate p, and e_{ij} is the random effect associated with teacher i in school j, conditioning on the other effects in the model.

There are six covariates. The first four represent attributes of teachers:

- Degree rank. Two dummy variables were used to indicate two out of the three mutually exclusive categories of this variable. (The highest rank category was used as the reference category; see appendix J for details.)
- Total years of teaching experience.
- Total years of teaching experience in the relevant subject.

The next two covariates are dummy variables that correspond to the attribute variables listed above. (If a teacher was missing a value for total years of teaching experience, that teacher was also missing a value for total years of teaching experience in the relevant subject area, and vice versa; therefore, a single dummy variable was used for both teaching experience covariates. A single dummy variable was also used to indicate a missing value for degree rank.) If the value was missing, the dummy variable took a value of 1; otherwise, the dummy variable took a value of 0.

Level 2 (school level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} T_j + \sum_{m=2}^{41} \gamma_{0m} I_{(m-2)} + u_{0j} \qquad p = 1,\ldots,6$$

$$\beta_{pj} = \gamma_{p0}$$

- T_j indicates whether a school was assigned to the AMSTI or control condition (coded 1 if the school was assigned to AMSTI and 0 if the school was assigned to control).
- $I_{(m-2)}$ indicates the matched pair to which a school belongs, taking on a value of 0 or 1. There are 40 indicators for the 41 pairs. The effect γ_{0m} represents the average difference in outcome between pair *m* and the reference pair, controlling for the other effects in the model.
- u_{0j} is the random effect of school *j*, conditioning on the other effects in the model.

Description of the extent of Alabama Math, Science, and Technology Initiative (AMSTI) implementation in Year 1

To describe the extent to which the main intervention components of AMSTI were implemented in Year 1, researchers conducted descriptive analyses for the AMSTI condition only, using data from professional development training logs, principal interviews, and teacher interviews for Subexperiment 2. No comparisons or statistical tests were performed. The results are reported in chapter 3.

Professional development. To characterize how comprehensively the professional development component of AMSTI was implemented in Year 1, researchers assessed the coverage of training topics and the instructional methods used. Training topic coverage was determined from the professional development training logs that reported daily on a 5-point scale (from 0, no coverage, to 4, full extent) how fully each topic outlined in the AMSTI manual was covered. Researchers calculated the average number of days each topic was covered to at least a “moderate” extent (a score of 2 on the scale) for each of the four grade/subject levels and reported the averages for each grade and subject level.

Researchers also used the professional development training logs to assess the extent to which trainers used particular instructional methods. Trainers reported the time spent daily on each method on a 5-point scale (from 0, no time, to 4, 76–100 percent of the time). For each instructional method, researchers calculated the average number of days trainers reported using a method more than 25 percent of the time and reported the averages for each grade and subject level.

Program materials, technology, and other resources. To characterize the extent to which the program materials, technology, and other resources component of AMSTI were implemented in Year 1, researchers assessed the availability of AMSTI materials based on teachers' reports during the school year. During interviews, teachers were asked an open-ended question about the accessibility of AMSTI materials. Two researchers independently reviewed responses to the question and searched for recurring words or themes in the text (Patton 2002) and then developed codes for each theme. The researchers then agreed on appropriate codes, deciding to categorize access as "full," "partial," or "none." The number and percentage of AMSTI teacher responses that fell into each category is reported in chapter 3.

*In-school supports***.** To characterize the extent to which the in-school supports component of AMSTI was implemented in Year 1, researchers assessed the availability of follow-up support throughout the school year. They used data from interviews in which principals were asked an open-ended question about whether support from the AMSTI site was provided to teachers when they needed it.[44] Responses were analyzed using the content analysis method described above, with supports categorized as either available or not available. The number and percentage of "Yes" and "No" responses by AMSTI principals were then reported.

Estimation of differences in implementation between Alabama Math, Science, and Technology Initiative (AMSTI) and control groups in Year 1

Teacher survey data for both subexperiments were used to describe the differences between the AMSTI and control conditions for each of the three components of the AMSTI intervention. Tests of the statistical significance of the differences were then conducted.

Professional development. In the first teacher survey (conducted in January of Year 1), both AMSTI and control teachers were asked to report the number of hours of professional development they received during the summer before the school year. The main source of variation in responses was whether or not teachers received professional development rather than how much professional development they received (see details in chapter 3 and relevant appendixes).[45] Because the large number of zero scores, researchers recoded the data as 0 for zero hours and 1 for more than zero hours. They estimated the intervention effect using a logit model that accounted for the clustering of teachers within schools. The *p*-value associated with the treatment effect—that is, the result of the statistical test of the null hypothesis that the AMSTI and control teachers have the same odds of reporting more than zero hours of summer

---

[44] Additional details on AMSTI site regions are in chapter 1.

[45] Among mathematics teachers, 13 percent of AMSTI teachers and 76 percent of control teachers reported receiving zero hours of summer professional development. Among science teachers, 16 percent of AMSTI teachers and 76 percent of control teachers reported receiving zero hours of summer professional development. Researchers tested ways of capturing the variance in responses that were greater than zero—for example, by creating a 5-level ordinal scale on which all zero responses were put into one category and the remaining responses ordered and divided into quartiles. The test of whether there was a difference between conditions in the distribution of responses across categories did not converge, because of the small number of responses greater than zero.

professional development—is reported separately for mathematics and science teachers. Accompanying these results in chapter 3 are bar graphs (figures 3.6 and 3.7) of the percentages of teachers in each condition who selected the response "more than zero hours." In addition, the parameter estimates on the probability scale are in appendix V: estimates of the marginal probabilities and their associated standard errors of responding "more than zero hours" are reported for the AMSTI and control groups. For descriptive purposes, the average number of professional development hours received by AMSTI and control teachers who reported taking part in professional development is also reported.

Program materials, technology, and other resources. In the third teacher survey (conducted in March of Year 1), teachers were asked how well equipped their classrooms were with the materials and equipment needed for mathematics and science instruction. Teachers used the following 4-point Likert scale to respond: I have all of the materials/manipulatives that I need; I have most of the materials/manipulatives that I need; I have some of the materials/manipulatives that I need; I do not have any of the materials/manipulatives that I need. The intervention effect was estimated using a multilevel ordinal logit model that accounts for the clustering of teachers in schools. For these outcomes the odds are based on the probabilities of selecting a given response category or a lower one. A two-tailed statistical test of the hypothesis that the intervention effect was statistically different from zero was performed. The results are reported separately in chapter 3 for mathematics and science teachers.

In-school supports. Across three teacher surveys (conducted in Year 1), teachers were asked to report the number of times they requested in-school support (defined as mentoring or coaching for instruction) in the past month and the number of times they received support. The responses were averaged across surveys; the responses of teachers who did not respond to all surveys were averaged without the missing survey or surveys.

As with the professional development data, these data were highly skewed because of the large number of "no request/received" responses (see details in chapter 3 and associated appendixes).[46] The data were therefore recoded so that average values between 0.0 and 0.5 became 0 and average values equal to or above 0.5 became 1. The intervention effect was estimated using a logit model that accounts for the clustering of teachers within schools. A two-tailed statistical test of the hypothesis that the coefficient for the intervention indicator was statistically different from zero was conducted. The marginal probabilities and their associated standard errors for the AMSTI and control groups are reported, as well as the estimated odds

---

[46] Among AMSTI mathematics teachers, 69 percent did not request and 41 percent did not receive support. Among AMSTI control teachers, 67 percent did not request and 60 percent did not receive support. Among AMSTI science teachers, 57 percent did not request and 36 percent did not receive support. Among control science teachers, 77 percent did not request, and 72 percent did not receive support. Researchers tested ways of capturing the variance in responses that were greater than zero—for example, by creating a 5-level ordinal scale on which all zero responses were put into one category and the remaining responses were ordered and divided into quartiles. The test of whether there was difference between conditions in the distribution of responses across categories did not converge, probably because of the small number of responses greater than zero.

ratio, which compares the odds of a given event occurring for teachers exposed to AMSTI with the odds of the same event occurring for teachers not exposed to AMSTI. For each outcome, the standard error and *p*-value associated with the treatment effect are reported separately for mathematics and science teachers.

Multiplicity adjustments

The Bonferroni procedure (Schochet 2009) was used to control for the familywise error rate. Adjustments were needed because when a "family" of hypotheses tests is considered simultaneously, the combined rate of drawing at least one false positive conclusion is larger than the significance level for any single test considered alone. Without a multiplicity adjustment, the probability of drawing a false positive conclusion concerning at least one effect increases sharply with an increase in the number of contrasts. With the Bonferroni procedure, the significance level for an individual test of effect is set to the value that would be used if there were only a single test (usually .05), divided by the number of tests. In the case of the primary confirmatory analyses of the effect of AMSTI, the unadjusted Type I error rate, .05, was divided by 2—the number of tests performed. Therefore, the significance level for either test was set at .025. The multiplicity adjustment allowed researchers to address whether the intervention had an effect on either mathematics or science. The null hypothesis that the intervention had no effect on either domain is rejected if the estimated effect on either domain was statistically significant at the .025 level.[47]

A separate multiplicity adjustment was carried out for the secondary confirmatory analysis. Two contrasts of interest were noted: the effect of AMSTI on teaching for active learning in mathematics and the effect on teaching for active learning in science. The significance level for either test was set to .025.

Sensitivity analyses

Ten sensitivity analyses were conducted to determine whether the effect on mathematics problem solving was robust to alternative valid approaches to estimating impact:

1. Using the gain score as the outcome variable.
2. Using a model that included as covariates only the pretest and pair indicators and that listwise deleted students without a pretest.

---

[47] The Bonferroni method yields conservative bounds on Type-1 error, and, hence, is a method that has low power (Schochet 2008, p. B-6); however, this was taken into consideration during rounds of technical review of the study proposal, and Bonferroni was deemed appropriate. (Resampling methods are sometimes used instead of the Bonferroni procedure when adjustments are made with correlated test statistics. With these methods, the significance levels for individual tests are often set higher than with the Bonferroni procedure, which means that there is a smaller chance of incorrectly failing to reject the null hypothesis. (That is, the tests are more powerful.) However, with only two contrasts, as is the case here, the additional benefit to using the resampling method is small. Therefore, the Bonferroni adjustment was used.

3. Using a model that included as covariates only the pretest and pair indicators and used the dummy variable approach to impute missing values.
4. Using listwise deletion instead of the dummy variable approach to examine the sensitivity of the effect to alternative strategies for handling missing values of covariates.
5. Using maximum likelihood instead of restricted maximum likelihood to test the sensitivity of findings to the estimation method.
6. Using a model that weighted equally the grade-specific impacts.
7. Using a model that weighted equally the effects specific to subexperiments.
8. Using a model that included both the mathematics problem solving pretest and the reading pretest.
9. Regressing school average outcomes against school averages of the covariates but excluding school averages of the dummy variables used to indicate missing values for the covariates. (Note that the benchmark hierarchical linear model weights schools by the inverse of their variances; therefore, outcomes for larger schools are given greater weight.)[48]
10. Regressing school average outcomes against school averages of the covariates but including school averages of the dummy variables used to indicate missing values for the covariates.

The analyses for science achievement paralleled those for mathematics problem solving, with one exception. Because reading pretests were used as a covariate in the analysis of science outcomes, it was not appropriate to calculate a gain score; no analysis with gain scores as outcomes was therefore performed.

Analyses 4, 5, 7, 9, and 10 were used to examine the sensitivity of the estimates of the effect of AMSTI on active learning instructional strategies to different estimation methods and are described above for impacts on student achievement in mathematics problem solving. To further assess the robustness of the results for these outcomes, an analysis was conducted in which teachers who responded to fewer than 4 items on the 12-item scale were removed from the sample. This analysis was performed to examine whether the results were sensitive to the inclusion of teachers for whom values for active learning had been imputed on the basis of fewer than four item responses.[49]

The number of sensitivity analyses is smaller for teacher outcomes than for student outcomes. There was no equivalent of a pretest measure for teacher outcomes; therefore, checks that relied on modeling a pretest (using a gain score analysis or a model that uses the pretest as

---

[48] The descriptive statistics for distributions of the number of students and teachers in schools in the analytic samples of the Year 1 outcomes are shown in appendix M.
[49] A sensitivity analysis that weights equally the grade-specific impacts was not conducted for the teacher impacts. Grade-specific impacts cannot be computed for teacher outcomes, because a teacher may teach more than one grade, making it impossible to designate the teacher or the teacher’s responses as belonging to a single grade level.

the only covariate, for example) could not be carried over to the set of sensitivity analyses involving teacher effects.

Addressing missing data and nonresponse

Attrition and missing data at the student level. The main model for estimating effects on students included indicators for matched pairs of schools as well as for the following covariates: pretest score, grade level, racial/ethnic minority status, eligibility for free or reduced-price lunch, English proficiency, and gender. Information about matched pairs was complete. Therefore, concerns about missing data focused on missing student posttests and missing values for the covariates.

Where student achievement outcomes (posttests) were missing, observations were listwise deleted and dropped from the analysis. For the sample associated with the mathematics problem solving outcome, posttests were missing for 4.4 percent of the eligible baseline sample of students; no eligible schools were missing all posttests. For the sample associated with the science outcome, posttests were missing for 7.6 percent of the eligible baseline sample of students as well as three eligible schools, a lower than expected rate of attrition. Using case deletion was therefore unlikely to reduce precision by much (and would not compromise the statistical power of the experiment, given that there was much less attrition than expected at the design stage). According to Puma et al. (2009), based on their simulation studies, case deletion "worked as well as, or better than, all of the alternative methods across all of the missing data scenarios" (p. 63). The alternative methods included regression imputation, EM Algorithm with Multiple Imputation, and fully specified regression models with treatment-covariate interactions.

A greater concern was the loss of schools, the unit of randomization. Student achievement outcomes were obtained from the state, and thus receiving student records was straightforward. No schools were lost from the analysis of mathematics problem solving outcomes. Two AMSTI schools and one control school were lost from the analysis of science achievement. The two AMSTI schools were lost because the sample excluded grade 5 and 7 with disabilities. This loss is not considered attrition, because these schools were not members of the eligible sample. The control school was lost when students without a posttest were eliminated. The loss of the control school was considered attrition.

Because this study focuses on intent-to-treat estimates, the analysis included outcomes for students who left the study schools before posttests were administered, provided they remained in the Alabama public school system. The study did not differentiate between students who received the intervention as intended by the program developers and those who received a less complete form of the intervention (or example, if their AMSTI-trained teacher was replaced by a teacher who did not receive AMSTI training.)

Two strategies were used to handle missing values for covariates: the dummy variable method (Puma et al. 2009) and listwise deletion of students with missing values for any covariate. The main analyses used the first strategy. The sensitivity analysis used the second approach.

Attrition and missing data at the teacher level. The analyses of classroom outcomes used teacher-level data. The main model for estimating effects on classroom outcomes included indicator variables for matched pairs of schools as well as the following covariates: teacher degree rank, total years of teaching experience, and total years of teaching experience in the relevant subject. Information about matched pairs was complete. Therefore, concerns about missing data had to do with missing active learning scores and missing values for the covariates.

As schools were the unit of randomization, the greatest concern was loss of schools in the analysis of classroom outcomes. As noted, 1 of the 82 randomized schools (a control school) withdrew from the study the day after randomization. This school was excluded from the survey sample and from the analysis of both secondary confirmatory outcome measures (active learning for mathematics teachers and active learning for science teachers).

For the active learning in mathematics outcome, data from 8 AMSTI teachers and 13 control teachers (21 total teachers) were lost because of missing outcomes. For the active learning in science outcome, data from 9 AMSTI teachers and 17 control teachers (26 total teachers) were lost because of missing outcomes.[50] Teachers for whom values were missing were listwise deleted from the sample. Missing teacher values for the covariates were handled the same way as missing student values.

Exploratory analysis: effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading achievement, teacher content knowledge, and student engagement after one year and variations in effects for student subgroups

This section describes the exploratory analyses involving outcomes after one year. The same models used for the confirmatory impact analyses were used. The models used for the primary confirmatory impact analyses were extended to measure differential impacts of various student subgroups (moderator analyses).[51]

Power analysis for the exploratory analysis of effect on reading performance. The power analysis was informed by the sample-based parameter estimates from the analysis of mathematics problem solving. Given that the analytic models were parallel for both outcomes (reading and mathematics problem solving), including the covariates used, and that the analytic sample sizes were expected to be almost the same, the sample-based values for the student level R2 (0.58), the school-level R2 (0.97), the school sample size (82 schools), and the student sample size (228 students per school) were adopted. A Type-I error rate of 0.05, a Type-II error

---

[50] Dropping these teachers from the analysis did not lead any schools that participated in the study to be excluded from the analysis of impacts on mathematics. However, it did lead to the exclusion of two control schools in the analysis of impacts on science.

[51] The terms *moderator analysis* and *subgroup analysis* are used interchangeably in this report. A moderator defines the subgroup. Therefore, one can look at the difference between subgroups in the program's impact or the moderating effect of subgroup membership on the program's impact. Both imply the same analytic model and the same effect of interest: the interaction between the subgroup indicator variable and the treatment indicator variable.

rate of 0.20, and an unconditional intraclass correlation coefficient of .22 were assumed.[52] Given these specifications, a minimum detectable effect size of 0.055 was computed.[53]

The hierarchical linear model used for reading paralleled the model used for the confirmatory analysis of the impact of AMSTI on student achievement in mathematics problem solving. It is not described again here. The same statistics are reported for both outcomes.

Analysis of effect on teacher content knowledge and student engagement. Data from the teacher surveys administered for this study were used to determine the impact of AMSTI on teacher content knowledge and student engagement. Survey questions were asked separately of mathematics and science teachers, yielding four outcomes: content knowledge of mathematics teachers, content knowledge of science teachers, student engagement in mathematics, and student engagement in science. Teachers rated their content knowledge at their current grade level on a 5-point Likert scale (very low, low, moderate, high, very high), with a sixth option of "not applicable".[54] They rated the average level of student engagement in their classes on a 5-point Likert scale (not engaged, slightly engaged, moderately engaged, almost fully engaged, fully engaged).

The power analysis was partially informed by the sample-based parameter estimates from the secondary confirmatory analyses of impacts on teacher outcomes. The sample sizes were assumed to be similar to those in the analysis of teaching for active learning in science, (78 schools and 5 teachers per school). These numbers were smaller than the achieved sample sizes for the analysis of teaching for active learning in mathematics; the more conservative values for the sample sizes were therefore used. For the other parameters, the values used in the original (not sample-based) power analysis were assumed for impacts on teaching for active learning: an unconditional intraclass correlation coefficient of .20; zero benefit from modeling covariates, including dummy variables to indicate matched pairs; .80 power; and a Type-I error rate of 5 percent. Based on these parameters, a minimum detectable effect size of 0.38 standard deviation was calculated.

Analytic model. A multilevel ordinal logit model that accounts for the clustering of teachers in schools was used to estimate the regression-adjusted average difference between the AMSTI and control groups in the cumulative odds of responding to each category.[55] The linear

---

[52] Hedges and Hedberg (2007) showed that for a heterogeneous sample of schools, the unconditional intraclass correlation for reading outcomes in grades 4–8 is .174–.263. The assumption of an intraclass correlation coefficient value of .22 is therefore reasonable for the reading outcome.
[53] Given the exploratory nature of these analyses, the outcomes were not adjusted for multiple comparisons, following guidance from the Institute of Education Sciences (Schochet 2008).
[54] A response of na cannot be assigned a meaningful numeric value and therefore was coded as missing and dropped from analysis.
[55] For these outcomes the odds are based on the probabilities of selecting a given response category or a lower one. Two-tailed statistical tests of the hypothesis of no difference between conditions in cumulative probability of response were conducted.

component of the model paralleled that used for the secondary confirmatory analysis of the impact of AMSTI on classroom instructional strategies with several exceptions.[56]

From baseline to analytic samples used in Year 1 exploratory analysis: attrition and differential attrition. The number of schools, teachers, and students was tracked throughout the course of the study. The numbers of cases at each stage associated with the Year 1 exploratory analysis (SAT 10 reading, teacher content knowledge, and student engagement) are summarized for each outcome in appendix N, and the attrition rates in AMSTI and control schools for these outcomes are compared.

Baseline and analytic sample counts as well as attrition rates were associated with the SAT 10 reading outcomes for the AMSTI and control groups at the school and student levels (table 2.10).[57] There was no attrition at the school level. At the student level, differential attrition was 0.1 percentage point, and overall attrition was 4.7 percent.

Table 2.10 School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) reading outcome after one year

	Schools			Students		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment, from rosters	41	41	82	12,065	10,492	22,557
Baseline (eligible) sample	41	41	82	10,517	9,109	19,626
Analytic sample	41	41	82	10,019	8,691	18,710
Attrition from baseline (eligible) to analytic sample	0	0	0	498 (4.7)	418 (4.6)	916 (4.7)

Note: Numbers in parentheses are percentages.
Source: Student achievement data from tests administered as part of the state's accountability system.

[56] In some cases, pair and/or school effects were removed where the estimation process did not converge or a random school effect could not be estimated (see chapter 6 for more details).
[57] Data on reading teachers were not collected. Counts and attrition rates are therefore presented only at the school and student levels.

For the sample associated with the teacher content knowledge in mathematics outcome, there was no attrition at the school level (table 2.11). At the teacher level, differential attrition was 2.1 percentage points, and overall attrition was 9.9 percent.

Table 2.11 School and teacher attrition associated with teacher content knowledge in mathematics outcome after one year

	Schools			Teachers		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment	41	41	82	na	na	na
Baseline (eligible) sample	41	40	81	221	205	426
Analytic sample	41	40	81	197	187	384
Attrition from baseline (eligible) to analytic sample	0	0	0	24 (10.9)	18 (8.8)	42 (9.9)

na is not applicable.
Note: Numbers in parentheses are percentages. The baseline was the first point for which information from teachers was available that allowed equivalence tests to be conducted.
Source: Teacher survey data.

For the sample associated with the teacher content knowledge in science outcome, one control group school was lost from the analytic sample (table 2.12). This amounted to differential attrition of 2.5 percentage points and overall attrition of 1.3 percent. At the teacher level, differential attrition was 1.8 percentage points, and overall attrition was 12.4 percent.

Table 2.12 School and teacher attrition associated with teacher content knowledge in science outcome after one year

	Schools			Teachers		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment	41	41	82	na	na	na
Baseline (eligible) sample	40	40	80	203	192	395
Analytic sample	40	39	79	176	170	346
Attrition from baseline (eligible) to analytic sample	0	1 (2.5)	1 (1.3)	27 (13.3)	22 (11.5)	49 (12.4)

na is not applicable.
Note: Numbers in parentheses are percentages. The baseline was the first point for which information from teachers was available that allowed equivalence tests to be conducted.
Source: Teacher survey data.

For the sample associated with the student engagement in mathematics classrooms outcome, there was no attrition at the school level (table 2.13). At the teacher level, differential attrition was 2.6 percentage points, and overall attrition was 9.6 percent.

Table 2.13 School and teacher attrition associated with student engagement in mathematics outcome after one year

	Schools			Teachers		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment	41	41	82	na	na	na
Baseline (eligible) sample	41	40	81	221	205	426
Analytic sample	41	40	81	197	188	385
Attrition from baseline (eligible) to analytic sample	0	0	0	24 (10.9)	17 (8.3)	41 (9.6)

na is not applicable.
Note: Numbers in parentheses are percentages. The baseline was the first point for which information from teachers was available that allowed equivalence tests to be conducted.
Source: Teacher survey data.

For the sample associated with the student engagement in science classrooms outcome, there was no attrition at the school level (table 2.14). At the teacher level, differential attrition was 2.9 percentage points, and overall attrition was 11.9 percent.

Table 2.14 School and teacher attrition associated with student engagement in science outcome after one year

	Schools			Teachers		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Random assignment	41	41	82	na	na	na
Baseline (eligible) sample	40	40	80	203	192	395
Analytic sample	40	40	80	176	172	348
Attrition from baseline (eligible) to analytic sample	0	0	0	27 (13.3)	20 (10.4)	47 (11.9)

na is not applicable.
Note: Numbers in parentheses are percentages. The baseline was the first point for which information from teachers was available that allowed equivalence tests to be conducted.
Source: Teacher survey data.

The analysis examined whether random assignment resulted in statistically equivalent groups at baseline and whether equivalence continued with the analytic sample in the Year 1 exploratory analysis. For the samples associated with student achievement in reading after one year, the equivalence between the AMSTI and control schools was tested on the same student characteristics examined for the confirmatory outcomes.[58] There were no statistically significant differences between the AMSTI and control schools on any of the background characteristics measured for the analytic or baseline samples associated with student achievement in reading after one year.

For samples associated with the exploratory analysis of teacher content knowledge and student engagement after one year, the equivalence between the AMSTI and control schools was tested on the same teacher characteristics used for the confirmatory outcomes. There were no statistically significant differences between the AMSTI and control schools on any of the background characteristics measured for the baseline samples associated with teacher content knowledge or student engagement in mathematics or science after one year.

For the analytic sample associated with teacher content knowledge in mathematics, the AMSTI and control groups differed on one characteristic tested: the distribution of teachers' degree rank ($p = .04$).[59] There were no statistically significant differences between the AMSTI and control schools on any of the background characteristics measured on the analytic sample associated with teacher content knowledge in science.

For the analytic sample associated with student engagement in mathematics, the AMSTI and control groups differed on teachers' degree rank ($p = .04$), and the overall test of whether the covariates as a whole were predictive of membership in an AMSTI or control school was statistically significant ($p = .04$). There were no statistically significant differences between the AMSTI and controls schools on any of the background characteristics measured for the analytic sample associated with student engagement in science. Tables displaying the full results of the equivalence tests of the baseline and analytic samples for the Year 1 exploratory outcomes are in appendix O.

---

[58] Data on reading teachers were not collected. Therefore, equivalence between the AMSTI and control groups was tested only on student-level characteristics.

[59] Teacher degree rank was measured using a 3-level ordinal scale, with teachers indicating how many degrees they had in their teaching content area (none, one, or more than one). For the analytic sample associated with teacher content knowledge in mathematics, 28 AMSTI teachers reported having no degree in the content area, 113 reported having one degree, and 52 reported having more than one degree. Among control group teachers, 39 reported having no degree in the content area, 104 reported having one degree, and 31 reported having more than one degree. The hypothesis of no difference between conditions in the distribution of teacher degree rank was tested using a multilevel ordinal logit model that accounted for the clustering of teachers in schools. (For these outcomes, the odds were based on the probabilities of selecting a given response category or a lower category.) A two-tailed statistical test was conducted for the hypothesis that the intervention effect was statistically different from zero.

Overview of analytic approach used for subgroup (moderator) analyses

All moderator analyses examined the differential impact of AMSTI for subgroups of students. For a given moderator, the full student sample was used, minus students who had a missing value for the moderator.[60] The analytic model was the same as the benchmark model used in the confirmatory analyses but with a cross-level interaction term added to measure the differential impact between subgroups. The details of the model are described below.

Power analysis. The minimum detectable differential effect sizes were computed using sample-based values of the parameters (table 2.15). Because the reading outcomes were not analyzed at the time the power analysis was conducted, sample statistics were not available to estimate the values of minimum detectable differential effect size for this outcome. Sample sizes and parameter values were expected to be similar for reading and mathematics problem solving. The mathematics problem solving samples were therefore used as a guide for the expected power for reading.[61] The statistical power calculations for the moderator analyses are in appendix P.

Table 2.15 Minimum detectable differential effect sizes for moderators for mathematics problem solving, science, and reading outcomes

Moderator	Mathematics problem solving	Science	Reading
Mathematics problem solving pretest	0.14	na	na
Reading pretest	0.16	0.25	0.14
Racial/ethnic minority status	0.10	0.18	0.10
Eligibility for free or reduced-price lunch	0.10	0.18	0.10
Gender	0.09	0.17	0.09

na is not applicable.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

[60] The rationale for not using the dummy variable approach to handling missing values for the moderator is given in a later section.
[61] The mathematics problem solving outcome rather than the science outcome was used as a guide for the minimum detectable differential effect size for reading because both mathematics problem solving and reading outcomes were analyzed for grades 4–8 whereas science outcomes were analyzed only for grades 5 and 7. For moderator variables that took on two values (such as 0 or 1 to indicate subgroups), the minimum detectable differential effect size is the minimum difference in the impact between the two levels of the moderator that is detectable under the constraints of the design, assuming specific levels of Type-I and Type-II error, expressed in standard deviations of the outcome variable. For the purpose of the power analysis, the pretest variables were dichotomized into values below the median ("low") and values at or above the median ("high"). In the actual moderator analyses, three levels were used. Scores were separated into three categories (the first three stanines, the middle three stanines, and the top three stanines) in each grade level. The cutpoints for the stanines were based on the pretest scale scores for the sample. Using three categories instead of two in the analysis should not be detrimental to statistical power and may actually increase it. The results presented in chapter 6 should thus be considered conservative.

Description of the model. As with the confirmatory impact analyses, two-level hierarchical linear regression models were used to estimate differential impacts.[62] Students were modeled at Level 1 and schools at Level 2.

The models used to estimate differential impacts were consistent with the benchmark models used in the confirmatory impact analyses in the following respects:

- To account for variation in the posttest and increase the precision of the impact estimates, the models included the following student-level covariates: pretest score, grade level, racial/ethnic minority status, eligibility for free or reduced-price lunch, English proficiency, and gender.
- To account for the random assignment design, the models included indicator variables that identified the matched pairs of randomized schools. The models also accounted for students being nested in schools.
- To control for potential bias in the impact estimate arising from missing covariate values, the models used a dummy variable method. For each of the covariates included, a dummy variable was created that assumed the value 1 if the value of the variable was missing for a given student and 0 otherwise. The missing values from the original variable were replaced with 0. (As described below, the dummy variable method was not used to handle missing values for the moderator variable.)

The models used to estimate differential impacts had these additional specifications:

- The differential impact was modeled by interacting the moderator variable with the indicator of the treatment effect. (The moderators were coded 0 or 1, except for the pretest, which was divided into three categories.) Each moderator analysis involved a cross-level interaction, because the indicator of the treatment effect was modeled at Level 2 whereas the moderator variable was modeled at Level 1.
- The dummy variable method was not used to handle missing values for the moderator variable. Using this method would have set missing values for the moderator to zero, making the interaction effect of the moderator uninterpretable. Instead, cases that were missing a value for the moderator were dropped from analysis.

---

[62] The model presented here assumes two levels for the moderator. The pretests were categorized into three levels. Analyses of the moderating effects of the pretests used a model similar to the one presented here, except two dummy variables were used to estimate the main effects of the pretest and two interaction terms were used to estimate differences in impact for the three levels of the pretest. The *p*-value associated with the Type III test was used for the effect corresponding to the interaction between the three-level preintervention performance measure and the indicator of treatment status, to assess whether the pretest interacted with treatment.

For each analysis, the reported p-value corresponds to a two-tailed test of the hypothesis that the regression-adjusted difference in impact between the two subgroups was zero (for the analysis of the moderating effect of the pretest, the p-value associated with the Type III test for the fixed effect corresponding to the interaction of the three-level preintervention performance measure and the indicator of treatment status is reported).[63]

The moderating effect of covariate M on the impact of AMSTI on student performance in mathematics problem solving, science, and reading was estimated using a two-level model, with students at Level 1 and schools at Level 2. The model assumed a constant differential effect of AMSTI but included a random school effect:[64]

Level 1 (student level):

$$y_{ij} = \beta_{0j} + \beta_{1j} M_{ij} + \sum_{p=2}^{17} \beta_{pj} COV_{pij} + e_{ij}$$

is the outcome for student i in school j.

- M_{ij} is the student-level moderator variable.
- COV_{pij} is the value of the covariate p. There are 16 covariates other than the moderator. The first eight covariates represent attributes of students. The next eight covariates are dummy variables that correspond to the eight student attribute variables. The dummy variable that corresponds to a given covariate indicates whether the value of that covariate was missing. If the value was missing, the dummy variable took a value of 1; otherwise, it took a value of 0.
- e_{ij} is the random effect associated with student i in school j, conditioning on the other effects in the model.

Level 2 (school level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \overline{X}_j + \gamma_{02} T_j + \sum_{m=3}^{42} \gamma_{0m} I_{(m-3)} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} T_j \qquad p = 2,...,17$$

$$\beta_{pj} = \gamma_{p0}$$

[63] Analyses of differential impacts for subgroups were run one at a time. Each model included the interaction of the indicator of treatment status with one moderator variable only. For example, for mathematics problem solving, the analysis of the effect of subgroup differences on the impact of AMSTI used five separate models, one for each moderator.

[64] The model was used to estimate the differential impact of AMSTI on student performance in mathematics problem solving. Differential impacts on science and reading were estimated using a similar model. Because the analysis of science involved students in grades 5 and 7 only, the model contained only a single dummy variable to indicate grade level. The dummy variable was grade 5; grade 7 was the reference category.

- $\overline{X}_j$ is the school mean of the pretest.
- T_j indicates whether a school was assigned to the intervention or control group (coded 1 if the school was assigned to the intervention and 0 if the school was assigned to the control).
- $I_{(m-3)}$ indicates the matched pair to which a school belonged, taking on a value of 0 or 1. There are 40 indicators for the 41 pairs. The effect γ_{0m} represents the average difference in outcome between pair *m* and the reference pair, controlling for the other effects in the model.
- u_{0j} is the random effect of school *j*, conditioning on the other effects in the model.
- γ_{00} is the covariate-adjusted mean value of the outcome measure for subgroup *M* = 0 across control schools.
- γ_{10} is the covariate-adjusted differential between subgroups *M* = 1 and *M* = 0 in the mean value of the outcome measure across control schools.
- γ_{02} is the mean difference in the covariate-adjusted outcome between treatment and control schools (main effect of treatment) for subgroup *M* = 0.
- γ_{11} is the mean differential between subgroups *M* = 1 and *M* = 0 in the covariate-adjusted main effect of treatment.

Substituting the Level-2 equations into the Level-1 equation yields the following mixed-model formulation:

$$y_{ij} = (\gamma_{00} + \gamma_{01}\overline{X}_j + \gamma_{02}T_j + \sum_{m=3}^{42}\gamma_{0m}I_{(m-3)} + u_{0j}) + (\gamma_{10} + \gamma_{11}T_j)M_{ij} + \sum_{p=2}^{17}\gamma_{p0}COV_{pij} + e_{ij}$$

$$= \gamma_{00} + \gamma_{01}\overline{X}_j + \gamma_{02}T_j + \gamma_{10}M_{ij} + \gamma_{11}T_j * M_{ij} + \sum_{p=2}^{17}\gamma_{p0}COV_{pij} + \sum_{m=3}^{42}\gamma_{0m}I_{(m-3)} + u_{0j} + e_{ij}.$$

For the moderator analyses, the parameter of interest, γ_{11}, indicates the differential in the average impact of AMSTI for subgroup *M* = 1 compared with subgroup *M* = 0. For each of the moderator analyses, a *t*-test was performed to test the null hypothesis that the differential between subgroups in the average impact of AMSTI was zero (that is, $\gamma_{11} = 0$). For ease of interpretation, especially for comparing the size of the differential effect of AMSTI with the size of the average effect, the estimates of the differential effect are presented in terms of standard deviations of the outcome variable for the control group (that is, as effect sizes).[65]

Handling missing data. In the main confirmatory analyses, the dummy variable method was used to handle missing values for the covariates. Where student achievement outcomes (posttests) were missing, listwise deletion was used to drop the observation from the analysis. The same approach was used to handle missing data in the exploratory analyses. The only exception involved the moderator variable. As described, estimates of the effects of the moderator were of substantive interest, so changing missing values of the moderator variable to 0 would have made the moderator effect uninterpretable. Therefore, listwise deletion was used for cases of a missing value for the moderator.

Analysis of average effect on achievement for student subgroups after one year. For each moderator variable, the sample was divided by levels of the moderator, and the regression-adjusted average impact of the program was estimated for each subgroup. The analytic models paralleled those used to estimate average impacts for the full sample (that is, all subgroups combined), except that both the covariates that identified the subgroups whose impact was being estimated and the corresponding dummy variables used to indicate missing values were removed. Approaches to handling missing values for the other covariates and the posttest were the same as those used to estimate average impacts for the subgroups combined (the dummy variable method for missing values of the covariates, listwise deletion for missing posttest scores.) For ease of interpretation, especially so that the size of the subgroup impact of AMSTI can be compared with the size of the average effect, the estimate of the subgroup effect is expressed in standard deviations of the outcome variable for the control group (that is, as effect sizes).[66]

Exploratory analysis: analysis of student achievement in mathematics problem solving and science after two years

The full AMSTI intervention is intended to be delivered over two years. The one-year impact findings presented in chapter 4 thus do not represent the effects of a full course of the intervention. While the original control group schools implemented AMSTI after one year,

---

[65] For each of the three main outcomes, the standard deviation for the control group from the analytic samples from the confirmatory analyses for impacts on mathematics problem solving and science and the corresponding exploratory analysis in reading was used to estimate the average impacts of AMSTI. This step was taken to facilitate comparison of results; the effect size for the moderating effects of the pretests is not reported, because this covariate is expressed in terms of three levels and moderating effect involves two interaction terms.

[66] See previous note.

leaving no untreated control group against which to measure the impact of AMSTI after two years, researchers selected an appropriate method to estimate the two-year effects.

Bell and Bradley (2008, 2009) developed a method to estimate the full effect of an intervention after the control group receives the intervention in a randomized controlled trial. Their method simulates the experimental framework that would have existed had the control group not received the intervention. The method provides an unbiased measure of the effect of the intervention in the second year if the impact of AMSTI during the first year of its implementation is stable over time—that is, if AMSTI affected the original control group schools the same way it affected treatment group schools a year earlier.

The assumption that the first year of exposure to AMSTI had the same effect on control group students in Year 2 as on treatment group students in Year 1 is contingent on a number of factors related to the intervention, the participants, and the context in which the intervention was implemented. Some of these factors can be tested in the data; others cannot. No evidence was found for any of the testable factors to suggest that the Bell-Bradley method yielded biased impact estimates (see below).

This section reviews the theory, assumptions, and analytic methodology of the Bell-Bradley approach and describes how it was used to calculate the effects of two years of AMSTI compared with no AMSTI. Because these analyses build on the random assignment design but do not use conventional randomized controlled trial methods, the findings are considered exploratory rather than confirmatory.

Estimation strategy. The goal of the two-year analysis is to measure the estimated effect on students of two years of AMSTI participation compared with zero years of AMSTI participation. Complicating this goal is the fact that control group schools began implementing the AMSTI program one year after intervention group schools implemented the program. Thus, in the second year, the control group schools do not provide data on student outcomes with zero years of AMSTI participation. The effect of one year of the program will already be incorporated in these students' mathematics and science achievement scores at the end of Year 2.

Bell and Bradley (2008) proposed a method for "recovering" a zero-treatment Year 2 outcome level for control group students in this circumstance. The basic strategy is to subtract from the average control group student outcome in Year 2 the effect that AMSTI had on those students' outcomes that year—or at least a good approximation of that effect. The approximation chosen equals the estimated Year 1 impact of AMSTI on *treatment group* students *a year earlier*. The resulting estimator of two-year effects equals the sum of the treatment-control differences in outcomes at a given grade level in consecutive school years, referred to as the "Year 1 component" and the "Year 2 component" of the overall two-year impact estimate. Appendix Q explains the method in more detail (in mathematical terms) and examines the properties of the estimator it produces.

Researchers concluded that the lack of bias of the Bell-Bradley estimator rests on the equivalence of effects in the AMSTI and control groups in the first year in which each group received the AMSTI intervention. This assumption was subject to partial testing, but its validity cannot be fully confirmed.

Checking the conditions that support the lack of bias of the estimator. Bell and Bradley (2009) laid out the conditions under which the expected impact of the intervention on student outcomes at a given grade level is the same in Year 2 for control group students (the first year they received the AMSTI intervention) as it is in Year 1 for treatment group students (the first year they received the AMSTI intervention). Random assignment ensures that, on average, the treatment and control groups are equivalent at the time of random assignment. However, in order to expect the impacts of AMSTI to be the same in Years 1 and 2, not only the groups but also the intervention must be equivalent, and the passage of time itself must not have changed the way that the impacts occur. Ultimately, what matters is that the impacts experienced by students in AMSTI schools in Year 1 equal the impacts experienced by students in control schools in Year 2. Bell and Bradley translated this requirement into a set of 11 factors that must remain stable over time if the Bell-Bradley estimator is to be unbiased. The research team was able to check the stability of these factors to some degree (table 2.16).[67]

Table 2.16 Factors that must remain stable over time (from Year 1 to Year 2) for the effect of the Alabama Math, Science, and Technology Initiative (AMSTI) to be the same in the Alabama Math, Science, and Technology Initiative and control group schools (in their respective first year of Alabama Math, Science, and Technology Initiative implementation)

Factor	Assessed for AMSTI Study	Type of check
Intervention-related factors		
A. Sponsor's guidelines for required parameters of the intervention's design and implementation	√	Document review
B. District's desire and evolving ability to support the intervention	√	Document review
Factors that could be affected by random assignment to the control group		
C. Composition of schools in the sample that did not implement the intervention when the time came to do so	√	Statistical tests
D. Composition of teachers in implementing schools that did not participate in the intervention or data collection when the time came to do so	√	Statistical tests
E. Effort by schools to implement the intervention	√	Statistical tests
Contextual factors		
F. Characteristics of age cohorts[a]	√	Statistical tests
G. Alternative programs and schools available in the community[a]	√	Statistical tests

[67] Bell and Bradley (2009) provided further logic on why each of these factors is important to the reliability of the method.

Factor	Assessed for AMSTI Study	Type of check
H. Existing school services apart from the intervention[a]	√	Statistical tests
I. Configuration of courses and curricula taught in school and system used by schools to assign course or curricula options to students at a given grade level[a]	√	Statistical tests
J. Characteristics of potential teachers in the community	Not assessed	Not assessed
K. Teacher hiring and course/class assignment practices	Not assessed	Not assessed

a. A single test was performed to test the net effect on the characteristics of students in the AMSTI and control group samples (rows F–I).

The factors are sorted into three groups:

- Factors related to the intervention (factors A–B), such as how AMSTI was carried out.
- Factors that may change because of randomization into the control group (factors C–E), such as how non-AMSTI inputs to students' education development could enhance or inhibit the effect of AMSTI.
- Contextual factors concerning students and the community (factors F–K), such as student characteristics, teacher characteristics, and how students and teachers are brought together for instructional purposes.

These factors may differ at random between the treatment and control group schools without creating bias. The Bell-Bradley estimator assumes that that they do not differ because of systematic changes over time—i.e., changes in the nature of the AMSTI intervention, the individuals who participate, or the context in which AMSTI functions between the first year of implementation in AMSTI schools and the first year of implementation (a year later) in control schools.

Nine of the 11 factors could be tested for equivalence between treatment group schools in Year 1 and control group schools in Year 2. Shifts over time in unmeasured or unobservable factors could have occurred that bias the two-year findings. Researchers checked those that they could with the available data.

Both of the intervention-related factors (A and B) could be checked to some extent through document review. Researchers had information on guidelines for the intervention (A) from ALSDE and on how those guidelines changed between the two consecutive years of implementation. One indicator of district support (B) is the science kit rotation schedule, as district staff could facilitate or inhibit the implementation of AMSTI through policies or procedures for rotating AMSTI materials among participating teachers.

Several other factors were checked for equivalence at the time of initial AMSTI implementation. First in this set were factors concerning the potential for random assignment into the control group to affect schools in ways that influence the effectiveness of AMSTI once it is implemented (see factors C–E in table 2.16). The first of these factors—the composition of nonimplementing schools (C)—was measured in the AMSTI evaluation data by looking at the share of schools that implemented the program at their designated time points and any notable characteristics of nonimplementing schools in the treatment and control groups. The share of teachers agreeing to complete surveys in support of the research (D) could also be checked and compared over consecutive years of implementation. Once engaged in AMSTI, teacher receipt of professional development training on the curriculum, use of teaching support, time spent on mathematics and science instruction, and use of AMSTI materials provide proxies of the effort expended by schools to implement the intervention (E).

As Bell and Bradley (2009) explain, contextual factors F–I matter only in their net influence on the types of students who receive the intervention in Year 1 in the treatment group schools and in Year 2 in the control group schools. These two sets of students could differ for a number of reasons. If they differ by chance, because of random assignment of schools, there is no bias. However, trend factors over time may create systematic differences that do give rise to bias. One potential trend element arises from the fact that the two groups of students come from two different age cohorts (factor F), because they reach the same grade level one year apart. Birth cohorts in the population can differ in ways that could influence the effects of the intervention. A second potential source of bias is that in any birth cohort, not all students from a given birth cohort will participate in AMSTI when implemented in public schools (thus potentially altering the mix of students for whom impacts are estimated); depending on the alternative schools available in the community (factor G), some students may attend private schools. This factor could change between Year 1 and Year 2 of the study, so that schooling alternatives differ between treatment group schools when they implement AMSTI and control group schools when they do so a year later. Existing school services apart from the intervention (H) and various school curricular and scheduling practices (I) could also change with time, moderating the effects of the AMSTI intervention. Researchers were unable to identify a source of data on characteristics of children in a birth cohort for a geographic area as small as a school district (factor F). They were also unable to catalog all nonpublic schools that might draw students away from AMSTI participation in their local public schools (factor G). Data used for the current evaluation report profiles the student population served by AMSTI in each set of schools, treatment versus control, reflecting the net result of differences in factors F–I.

The availability, hiring, and assigning of teachers to AMSTI may also differ between treatment group schools in Year 1 and control group schools in Year 2, if there is an underlying trend. Profiling the pool of potential teachers available for hire at a school (factor J) exceeded the study's data collection capabilities, as did examination of individual schools' teacher hiring and course assignment practices (factor K).

All of the available indicators were used to determine whether there were clear violations of the assumption of stability from one year to the next that undergirds the Bell-Bradley estimator. This analysis is summarized with the findings on the two-year estimated effects of AMSTI in chapter 5.

From baseline to analytic samples: attrition and differential attrition in the Year 2 sample. The estimation method used to measure two-year effects required data from two consecutive calendar years. Appendix R provides details on the development of the analytic samples for the two-year analysis of mathematics problem solving and science.[68] This section examines attrition in those samples to gauge the potential for attrition bias in the analytic samples used to generate the two-year estimates.

As noted above, the Bell-Bradley estimator equals the sum of the treatment-control differences in outcomes at a given grade level in consecutive school years; these two differences are referred to as the “Year 1 component” and the “Year 2 component” of the estimator. To obtain these components, two related samples are needed: one contributing to the estimation of the Year 1 component of the estimator, the second contributing to the estimation of the Year 2 component. The baseline sample for each component included all students without disabilities in the appropriate grades (grades 4–8 for mathematics and grades 5 and 7 for science). Students, teachers, and schools remained in the analytic sample if a student did not transfer from a school in Subexperiment 1 to a school in Subexperiment 2; identifiers were available for the student, teacher, and school; and the student posttest was available.

Baseline and analytic sample counts and attrition rates were associated with the two student-level outcomes (SAT 10 mathematics problem solving and SAT 10 science) (tables 2.17 and 2.18). Total attrition and attrition for the AMSTI and control groups are displayed at the school, teacher, and student levels. (See appendix R for the number of schools, teachers, and students lost at each stage in the process of identifying the analytic sample from the baseline sample.)

Table 2.17 School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome after two years

	Schools			Teachers			Students		
Item	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**	**AMSTI**	**Control**	**Total**
Sample contributing to Year 1 component of the estimated two-year effect									
Random assignment, from rosters	41	41	82	251	234	485	12,198	10,514	22,712
Baseline (eligible) sample	41	41	82	246	229	475	10,517	9,109	19,626
Analytic sample	41	41	82	243	229	472	9,520	8,474	17,994
Attrition from baseline (eligible) to analytic sample	0	0	0	3 (1.2)	0	3 (0.6)	997 (9.5)	635 (7.0)	1,632 (8.3)
Sample contributing to Year 2 component of the estimated two-year effect									

[68] The sample sizes in tables 2.17 and 2.18 differ from the sample sizes in tables 2.5 and 2.6 because students who skipped or repeated a grade in Year 2 were excluded from the sample used to estimate the two-year effect. See appendix R for additional information.

Item	Schools			Teachers			Students		
	AMSTI	Control	Total	AMSTI	Control	Total	AMSTI	Control	Total
Random assignment, from rosters	41	41	82	217	170	387	11,574	10,081	21,655
Baseline (eligible) sample	41	41	82	213	165	378	10,139	8,612	18,751
Analytic sample	40	41	81	208	164	372	9,386	8,144	17,530
Attrition from baseline (eligible) to analytic sample	1		1	5	1	6	753	468	1,221
	(2.4)	0	(1.2)	(2.4)	(0.6)	(1.6)	(7.4)	(5.4)	(6.5)

Note: Numbers in parentheses are percentages.
Source: Student achievement data from tests administered as part of the state's accountability system.

The maximum overall attrition rate associated with the SAT 10 science outcome was 10.9 percent for students contributing to the estimation of the Year 1 component of the estimated two-year effect (table 2.18). No other overall attrition rate exceeded 10 percent. No differential attrition rate exceeded 5 percentage points.

Table 2.18. School, teacher, and student attrition associated with Stanford Achievement Test Tenth Edition (SAT 10) science outcome after two years

Item	Schools			Teachers			Students		
	AMSTI	Control	Total	AMSTI	Control	Total	AMSTI	Control	Total
Sample contributing to Year 1 component of the estimated two-year effect									
Random assignment, from rosters	41	41	82	233	213	446	12,065	10,492	22,557
Baseline (eligible) sample	39	41	80	103	95	198	4,480	3,688	8,168
Analytic sample	39	40	79	101	90	191	3,914	3,364	7,278
Attrition from baseline (eligible) to analytic sample		1	1	2	5	7	566	324	890
	0	(2.4)	(1.3)	(1.9)	(5.3)	(3.5)	(12.6)	(8.8)	(10.9)
Sample contributing to Year 2 component of the estimated two-year effect									
Random assignment, from rosters	41	41	82	206	156	362	11,574	10,081	21,655
Baseline (eligible) sample	39	41	80	91	61	152	4,253	3,415	7,668
Analytic sample	38	40	78	85	60	145	3,843	3,161	7,004
Attrition from baseline (eligible) to analytic sample	1	1	2	6	1	7	410	254	664
	(2.6)	(2.4)	(2.5)	(6.6)	(1.6)	(4.6)	(9.6)	(7.4)	(8.7)

Note: Numbers in parentheses are percentages.
Source: Student achievement data from tests administered as part of the state's accountability system.

Sample equivalence at baseline and in analytic Year 2 sample. Testing 10 characteristics for the mathematics sample and 7 characteristics for the science sample revealed no statistically significant differences between AMSTI and control schools in either the baseline sample or the analytic sample used in the two-year analyses. Moreover, for both mathematics and science, joint tests of significance for all covariates taken together found no overall difference. (The equivalence between AMSTI and control groups on student demographic characteristics for the samples used in the estimation of two-year effects of AMSTI is demonstrated in appendix S.)

Estimation. The model used to estimate the two-year impact followed the same specifications as the one used to estimate the one-year impact, with the exception that an additional dummy variable was introduced to indicate whether the outcome was measured at the end of Year 1 or at the end of Year 2. This variable was interacted with the indicator of treatment status. The two-year impact estimate was based on the coefficient for this interaction term and the one corresponding to the indicator of treatment status. Because science test scores were available only for students in grades 5 and 7, the outcomes in consecutive years at a given grade level were from different students, and therefore it was not necessary to figure in dependencies due to repeated measures for individuals. The same was not true of math problem solving, where the standard errors were adjusted post hoc because of repeated measures for students in certain grades. The full specification of the model and how it was estimated are provided in appendix T.

3: Analyses of Alabama Math, Science, and Technology Initiative (AMSTI) implementation

This chapter first briefly describes each of the three main components of the intervention in AMSTI schools according to the AMSTI theory of action (professional development; program materials, technology, and other resources; and in-school supports; see figure 1.1 in chapter 1). It then presents the results of implementation analyses for each component by: assessing the extent to which AMSTI components were implemented in Year 1, using descriptive statistics, and identifying differences between AMSTI and control conditions in the presence of each AMSTI component and similar components in control schools during Year 1, using inferential statistics. These analyses provide context for assessing and understanding the measured effects on student and teacher outcomes presented in chapter 4. Finally, the chapter draws some general conclusions about the two major aims of the analyses.

There are four methodological limitations to these analyses:

- Objective benchmarks for AMSTI implementation do not exist. Therefore, the analyses aim simply to describe program implementation for each program component.
- The descriptive implementation analyses are based only on data collected for Subexperiment 2; how well the results generalize to Subexperiment 1 is unknown.
- Implementation data rely on self-reports of relevant constructs. Self-reported data are subject to a variety of potential reporter biases. Therefore, caution should be exercised in the interpretation of these results, as the reliability and validity of these self-reported data is currently unknown.
- The analyses reflect implementation of AMSTI components for the first year of a two-year intervention.[69] All three components are intended to be implemented over two years, so full implementation could not be expected to have occurred in the first year of the intervention.

Alabama Math, Science, and Technology Initiative (AMSTI) implementation components

Professional development

AMSTI summer institutes are the main source of professional development training for teachers for increasing hands-on, inquiry-based learning in the classroom. AMSTI teachers are required to attend these trainings, which provide five days of training per subject (five days for mathematics and five days for science for a total of 10 days) for primary grade teachers and 10 days of training per subject for middle school grade teachers. Summer institute trainings are built

---

[69] Additional details on the rationale for assessing Year 1 implementation data only are included in chapter 2.

around a standardized curriculum provided by the Alabama State Department of Education (ALSDE), the developers of AMSTI.

Training is delivered by master teachers certified as AMSTI trainers who have successfully completed workshops provided by ALSDE. AMSTI trainers follow a curriculum that covers grade- and subject-specific topics. Trainers are expected to follow the AMSTI instructional methods when delivering the training. Lesson demonstration by AMSTI trainers is a key part of the workshops. Hands-on activities, small-group discussion, and skills practice are emphasized; lecture and whole-group discussion are used less often.

Program materials, technology, and other resources

In addition to professional development training, the AMSTI Committee recommended that AMSTI teachers have ready access to the materials and supplies necessary to implement inquiry-based, hands-on activities critical to AMSTI instruction. These materials can include lab supplies, thermometers, digital cameras, and various test kits. AMSTI personnel deliver the materials to schools as kits. In each region, the kits are rotated from school to school in three-month to semester-long blocks. The rotation and delivery of kits is coordinated by the AMSTI sites.[70]

In-school supports

Faculty members from two regional universities are available at the summer institutes and throughout the school year to support AMSTI teachers and schools. Following the summer institutes, full-time AMSTI mathematics and science specialists from the AMSTI site in each region provide AMSTI teachers with on-site mentoring in their classrooms, to help newly trained teachers become comfortable with AMSTI pedagogy. AMSTI teachers are also encouraged to collaborate with other teachers implementing AMSTI, to create an in-school support network.

Alabama Math, Science, and Technology Initiative (AMSTI) implementation results

Data collected from professional development training logs, teacher interviews, and principal interviews from Subexperiment 2 used for descriptive analyses, as well as teacher surveys from Subexperiment 1 and 2 used for inferential analyses, were collected and analyzed to assess AMSTI implementation in Year 1. These data sources, data collection procedures, sampling methods, and data analysis methods are described in detail in chapter 2; they are reviewed briefly here to provide context. Results are presented on the extent to which each component was implemented in Year 1 and the differences between the AMSTI and control conditions in each component during Year 1.

[70] Additional details on AMSTI site regions are in chapter 1.

Extent to which the Alabama Math, Science, and Technology Initiative (AMSTI) was implemented in Year 1

Professional development. Two measures were used to assess the extent to which the professional development component of AMSTI was implemented: coverage of particular training topics and the use of particular instructional methods. Training topics and modes of instructional delivery comprise the major content of the AMSTI development trainings. Consequently, the degree to which trainers covered the topics in their training manuals provided an estimate of teacher exposure to the lessons they were expected to teach as part of AMSTI. The degree to which trainers modeled the specific instructional methods during the training provided an estimate of the degree of teacher exposure to the repertoire of inquiry-based instructional methods they were expected to use in their classrooms as part of AMSTI.

Coverage of training topics. Trainers were asked to complete daily training logs in which they rated the extent to which they covered each topic in the AMSTI manual using a 5-point Likert scale (0 = no coverage, 1 = limited extent, 2 = moderate extent, 3 = large extent, and 4 = full extent). The number of training topics varied by grade and subject: 7 topics for grade 5 mathematics, 13 for grade 5 science, 15 for grade 7 mathematics, and 13 for grade 7 science. The number of days of training and the number of trainers also varied by grade level. For grade 5 mathematics and science, there were 5 days of training per subject (5 for mathematics and 5 for science for a total of 10 days) and 4 trainers. For grade 7 mathematics and science, there were 10 days (two weeks of training) and 2 trainers.

AMSTI trainers are instructed to cover the topics outlined in their training manual. However, the degree to which topics are covered is not prescribed. Without such guidance, it is impossible to assess whether topics were covered “adequately” during trainings. In the absence of any preestablished criteria for adequate coverage, researchers assessed the average number of days a topic was covered to at least a “moderate” extent in each of the four grade/subject levels. This assessment provided an estimate of the level of exposure (as indexed in average number of days) AMSTI teachers had to each topic during training. “Moderate” was chosen as the cut-off for topic coverage because a trainer covering a topic to a “limited” extent or less may not have provided substantial coverage of (and therefore exposure to) a topic during training.[71]

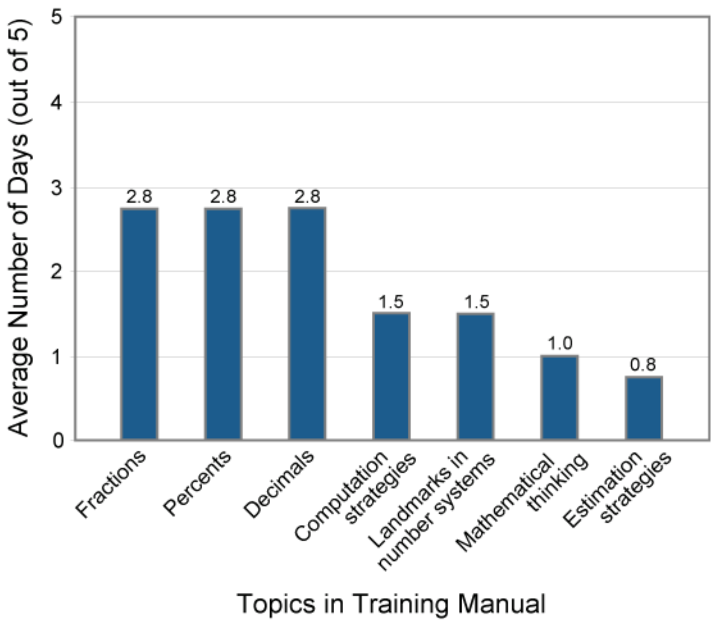
Researchers tallied the total number of days each trainer reported covering a specific topic to at least a moderate extent, summed the total number of days all trainers within a grade/subject level covered a topic to at least a moderate extent, and divided the sum by the total number of trainers within that grade/subject level. For example, for the topic *fractions* in grade 5 mathematics, the number of days each of the four trainers covered fractions to at least a moderate extent was summed (3 days + 3 days + 3 days + 2 days = 11 days) and divided by the number of trainers (4), yielding an average of 2.75 days of training that fractions were covered to at least a

[71] The average number of days a topic is covered to a moderate extent or more may underestimate the coverage of topics warranting more in-depth coverage on a single day of training and overestimate the coverage of topics warranting less in-depth coverage on a single day of training.

moderate extent for grade 5 mathematics. A list of full topic names, the average number of days each was covered to at least a moderate extent, and corresponding standard errors are in appendix U.

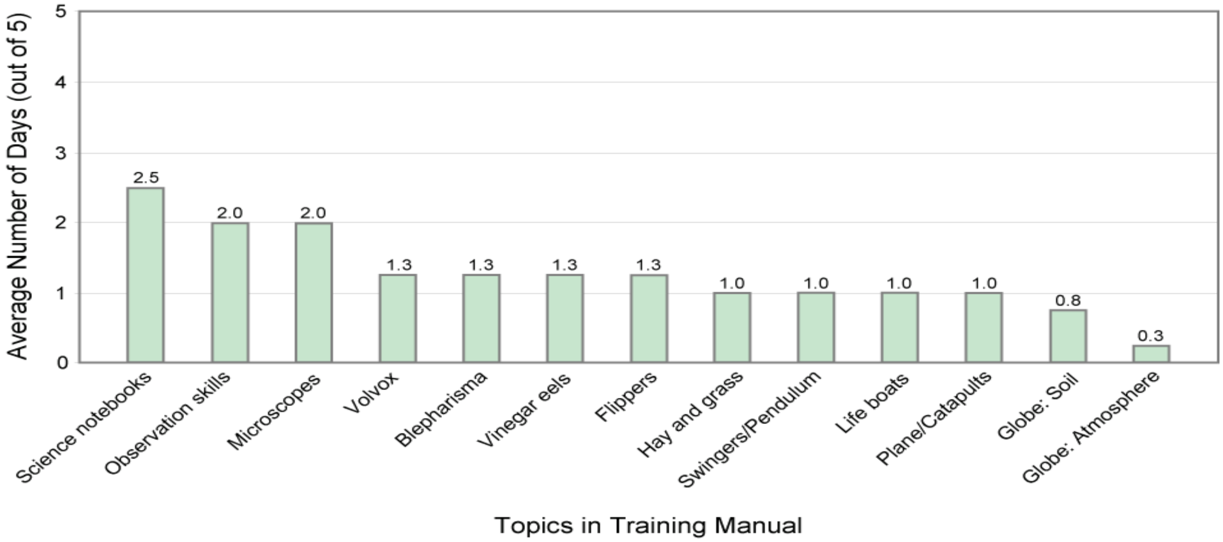
Six out of 7 topics (86 percent) were covered to a moderate extent or more on at least one day of training for grade 5 mathematics (figure 3.1). Eleven out of 13 topics (85 percent) were covered to a moderate extent or more on at least one day of training for grade 5 science (figure 3.2).

Figure 3.1 Average number of days Alabama Math, Science, and Technology Initiative (AMSTI) trainers covered topics in training manual to a moderate extent or more in grade 5 mathematics



Note: Each of four trainers had 5 logs each, for a total of 20 logs.
Source: Professional development training logs.

Figure 3.2 Average number of days the Alabama Math, Science, and Technology Initiative (AMSTI) trainers covered topics in training manual to a moderate extent or more in grade 5 science



Note: Each of four trainers had 5 logs each, for a total of 20 logs.
Source: Professional development training logs.

Because there were only two trainers per subject for grade 7, findings by subject could not be published without risk to participant confidentiality. Instead, the results were aggregated for mathematics and science topics. There were 28 topics to be covered over the course of training for grade 7 mathematics and science: 15 for mathematics and 13 for science. On average, grade 7 AMSTI trainers covered 96 percent of those topics (27 out of 28) to a moderate extent or more on at least one day of training.

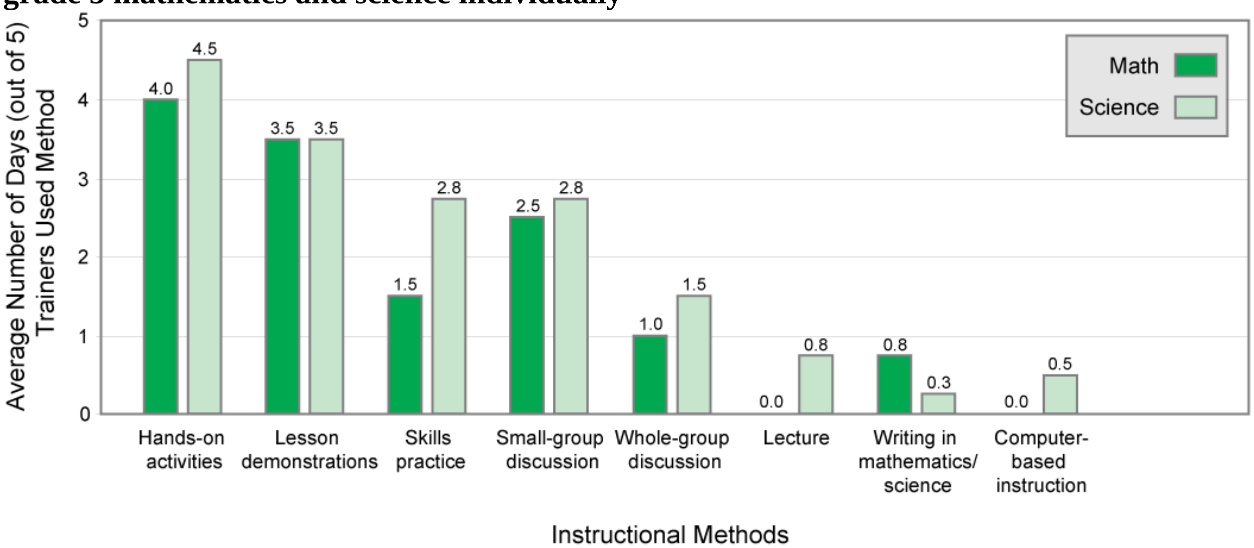
Extent of instructional methods use. All trainers were asked to report daily in their training logs the percentage of time they spent using each of the following instructional methods: hands-on activities, lesson demonstrations, skills practice, small-group discussion, whole-group discussion, lecture, writing in math/science, and computer-based instruction. Trainers reported the time spent on each method on a 5-point Likert scale (0 = no time, 1 = 1–25 percent of the time, 2 = 26–50 percent, 3 = 51–75 percent, 4 = 76–100). For each instructional method, researchers calculated the average number of days across all trainers within a grade/subject level that a method was used more than 25 percent of the time during summer professional development trainings.

AMSTI has no benchmarks for the proportion of training time to be devoted to particular instructional methods. The AMSTI theory of action does specify that trainers should use all of these instructional methods concurrently over the course of training; that hands-on activities, small-group discussion, and skills practice should be emphasized; and that lecture and whole group discussion be used less often. Without more specific guidance, it could not be determined whether instructional methods were "adequately" used over the course of trainings. Rather, analyses aim to describe, based on empirically-determined cutpoints, the relative use of instructional methods over the course of trainings. Researchers used the average number of days a method was used more than 25 percent of the time as an index of the level of exposure of AMSTI teachers to each instructional method during training. A 25 percent cutpoint was chosen because it provided enough variability in trainer use of instructional methods to distinguish the degree of use of hands-on activities, small-group discussion, and skills practice compared with lecture and whole-group discussion. Using a lower or higher cutpoint yielded almost no variability in the average number of days methods were used (1–25 percent on almost every day or 51–100percent on almost every day), which is consistent with the expectation that methods should be used concurrently.

The average number of days trainers used each method more than 25 percent of the time was calculated by tallying the number of days each trainer reported spending at least that amount of time using the instructional method, summing the number of days across all trainers within a grade/subject level, and dividing the sum by the number of trainers within that level. Numerical values and standard errors for these averages are in appendix U.

In grade 5 mathematics and science training, hands-on activities were used more than 25 percent of the time—on average, 4.0 of 5 days for mathematics and 4.5 of 5 days for science (figure 3.3). Lesson demonstrations were used more than 25 percent of the time—on average, 3.5 of 5 days for both mathematics and science. At the other end of the spectrum, for both mathematics and science training, lectures, writing in mathematics and science, and computer-based instruction were used more than 25 percent of the time on less than one of the five training days.

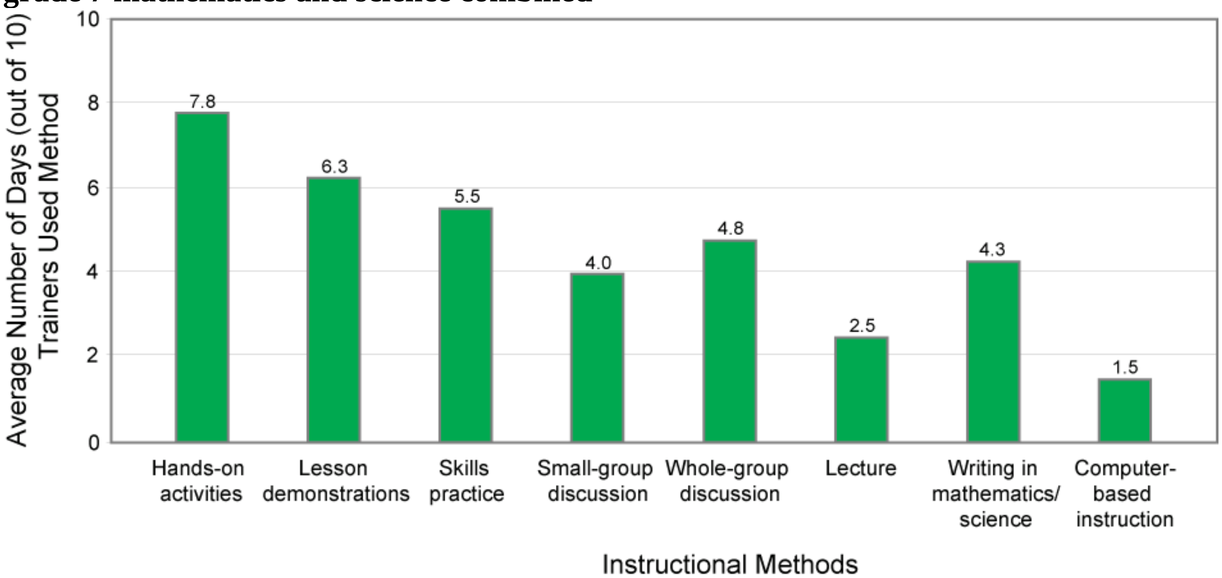
Figure 3.3 Average number of days Alabama Math, Science, and Technology Initiative (AMSTI) trainers used various instructional methods more than 25 percent of the time in grade 5 mathematics and science individually



Note: Each of four trainers had 10 logs each (5 for each subject), for a total of 40 logs.
Source: Professional development training logs.

Because there were only two trainers per subject for grade 7, the findings could not be reported by subject. Doing so might risk participant confidentiality. Therefore, the average number of days both grade 7 mathematics and science trainers used a specific instructional method more than 25 percent of the time are reported (figure 3.4). The results show that on average, hands-on activities were used more than 25 percent of the time, (7.8 out of 10 training days), lesson demonstrations more than 25 percent of the time (6.3 out of 10 training days), and skills practice more than 25 percent of the time (5.5 days out of 10 training days). Other instructional methods were used more than 25 percent of the time on less than half of the instructional days.

Figure 3.4 Average number of days Alabama Math, Science, and Technology Initiative (AMSTI) trainers used various instructional methods more than 25 percent of the time in grade 7 mathematics and science combined

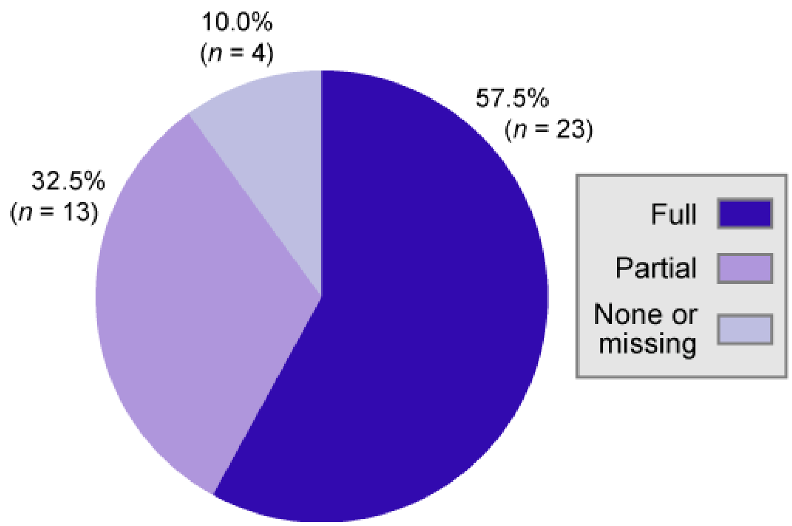


Note: Each of four trainers had 10 logs each (5 for each subject), for a total of 40 logs.
Source: Professional development training logs.

*Program materials, technology, and other resources***.** Teacher reports of the availability of AMSTI materials during the school year were assessed in order to characterize the extent of implementation for the program materials, technology, and other resources. During interviews, teachers were asked, “To what extent have you been able to access AMSTI materials?” Teacher responses were categorized into three levels of implementation (full, partial, and none) based on content analysis. (For details, see chapter 2.)

More than half of teachers (58 percent) reported full access to materials or that materials were readily available during the school year (figure 3.5). Another third (32 percent) stated that they had partial or limited access. Ten percent stated that they had no access or provided a response related to their personal use of materials rather than access to them (coded as "missing").[72]

Figure 3.5 Reported access to Alabama Math, Science, and Technology Initiative (AMSTI) materials by Alabama Math, Science, and Technology Initiative (AMSTI) teachers



Source: Teacher interview data.

In-school supports. Interviews with principals were used to assess the availability of follow-up support during the school year. As part of a question about the support received from the AMSTI sites, principals from Subexperiment 2 AMSTI schools were asked, "Has support been provided when teachers need it?" All 17 principals who responded to the question reported that the follow-up supports had been provided when needed.[73]

Differences between Alabama Math, Science, and Technology Initiative (AMSTI) and control schools during Year 1

Data from the teacher surveys (in Year 1 of Subexperiment 1 and Subexperiment 2) were used to estimate differences between AMSTI and control schools for each program component. Logistic regression was used to estimate differences for the professional development and in-school supports components; ordinal regression was used to estimate differences for the program materials, technology, and other resources components. (Details on estimation methods and procedures are in chapter 2.) Parameter estimates on the probability scale and of specific model tests are in appendix V.

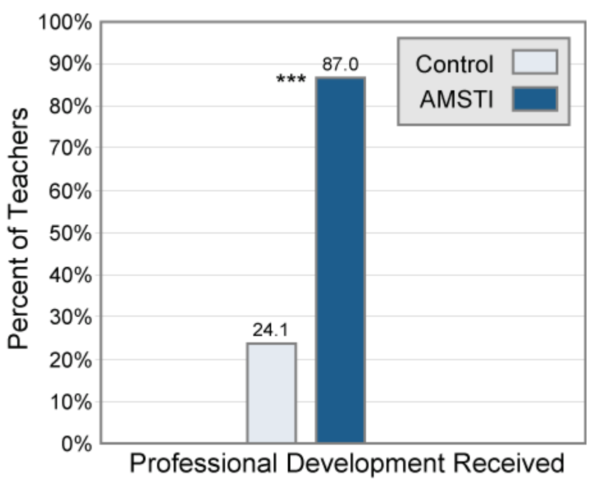
---

[72] Although the "no access" and "missing" categories are conceptually distinct, the small number of responses in these categories necessitated aggregating them to protect participant confidentiality.
[73] Four of the 21 principals surveyed (19 percent) did not respond to this question. These data were coded as "missing." Missing responses may indicate that the principal did not know the answer to the question or that he or she provided a response that did not answer the question.

Professional development. The first teacher survey, administered in January of Year 1, asked both AMSTI and control teachers to report the number of hours of professional development (both AMSTI and non-AMSTI professional development) they received the previous summer.[74] As shown in the descriptive statistics in appendix W, this variable was highly skewed because of a large number of "zero" responses. The main source of variation in this variable came from whether teachers received professional development rather than from how much professional development they received. Therefore, the variable was dichotomized into teachers receiving no professional development and teachers receiving at least some professional development during the summer prior to the implementation of AMSTI in classrooms.

AMSTI mathematics teachers were more likely to have received professional development than were control teachers: 87 percent of AMSTI mathematics teachers (n = 114) and 24 percent of control mathematics teachers (n = 27) reported receiving professional development during the same summer (figure 3.6). This difference was statistically significant ($p < .01$). AMSTI mathematics teachers who reported receiving professional development during the summer reported an average of 50.6 hours. Control mathematics teachers who reported receiving professional development reported an average of 14.7 hours.

Figure 3.6 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control mathematics teachers receiving summer professional development



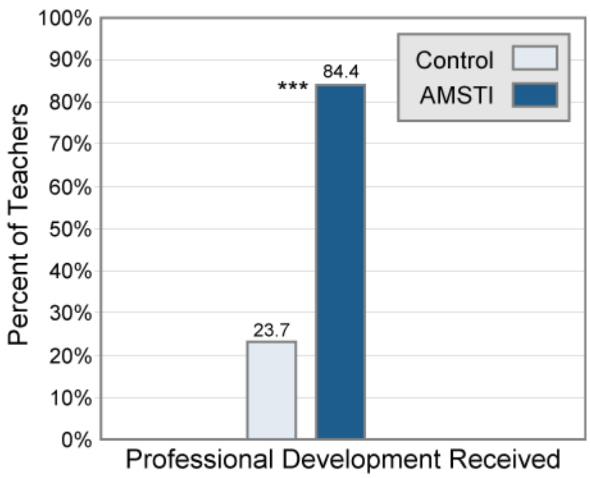
*** Significant at $p < .01$ based on a test of the difference between AMSTI and control teachers in the log-odds of receiving summer professional development.
Note: n = 68 schools; 243 teachers.
Source: Teacher survey data.

---

[74] Response rates were lower for the questions about summer professional development experiences than they were for other questions. Respondents were not required to answer all questions. Researchers hypothesize that teachers did not answer these questions when they wished to indicate that they had not received professional development. There is no way to confirm this hypothesis.

AMSTI science teachers were more likely to receive professional development than were control teachers: 84 percent of AMSTI science teachers ($n = 97$) and 24 percent of control group science teachers ($n = 22$) reported receiving professional development in the summer before AMSTI classroom implementation (figure 3.7). This difference was statistically significant ($p < .01$). AMSTI science teachers who reported receiving professional development during the summer reported an average of 65.5 hours. Control science teachers who reported receiving professional development during the summer reported an average of 32.6 hours.

Figure 3.7 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control science teachers receiving summer professional development

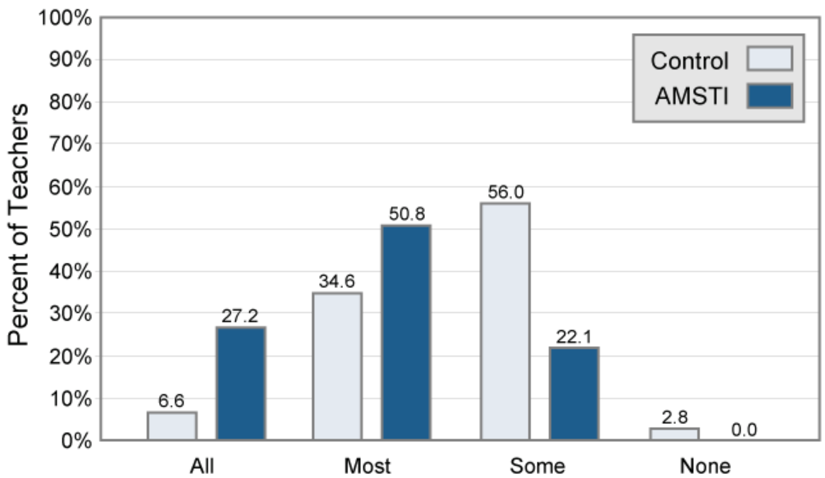


*** Significant at $p < .01$ based on a test of the difference between AMSTI and control teachers in the log-odds of receiving summer professional development.
Note: n = 62 schools; 208 teachers.
Source: Teacher survey data.

Program materials, technology, and other resources. In the third teacher survey (administered in March of Year 1), teachers were asked how well equipped their classrooms were with the materials and equipment needed for mathematics and science instruction. Teachers used a 4-point Likert scale ranging from "I have all of the materials/manipulatives that I need" to "I do not have any of the materials/manipulatives that I need." Because their training emphasized hands-on, inquiry-based teaching methods, AMSTI teachers may have had higher expectations about the materials needed. Consequently, for a given level of availability of materials, it is possible that AMSTI teachers may have been less likely than control teachers to report that they had the materials and manipulatives needed.

AMSTI mathematics teachers were more likely to report having greater access to materials and manipulatives than were control mathematics teachers: 78 percent of AMSTI mathematics teachers ($n = 152$) and 41 percent of control mathematics teachers ($n = 75$) reported having most or all of the mathematics manipulatives they needed (figure 3.8). This difference was statistically significant ($p < .01$).[75]

Figure 3.8 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control mathematics teachers reporting access to materials and manipulatives

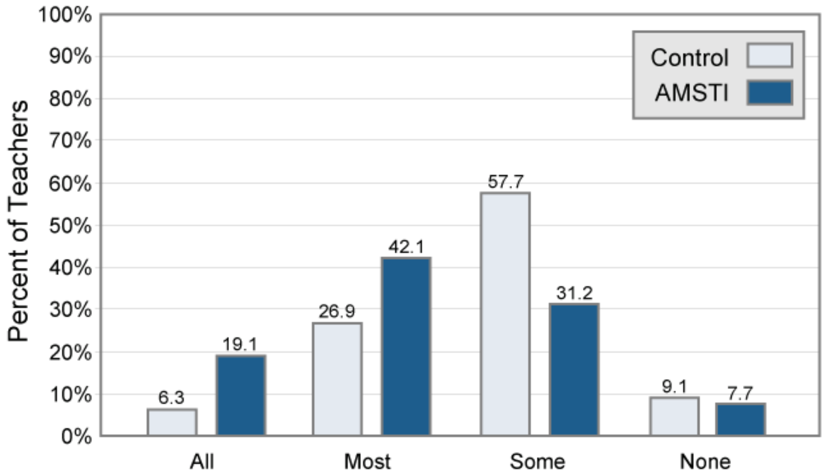


Note: n = 78 schools; 372 teachers. A single test was conducted to determine whether there was a significant difference between AMSTI and control schools in the reported level of access to materials and manipulatives. An ordinal logit model was conducting accounting for clustering of teachers within schools. The difference between the treatment and control groups was statistically significant ($p < .01$).
Source: Teacher survey data.

AMSTI science teachers were more likely than were control science teachers to report having greater access to materials and manipulatives: 61 percent of AMSTI science teachers (n = 112) and 33 percent of control science teachers (n = 58) reported having most or all of the science manipulatives they needed (figure 3.9). This difference was statistically significant ($p <$ .01).

[75] An ordinal logit model was conducted accounting for clustering of teachers within schools; the model tested for significant treatment-control differences in the reported level of access to materials and manipulatives.

Figure 3.9 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control science teachers reporting access to materials and manipulatives



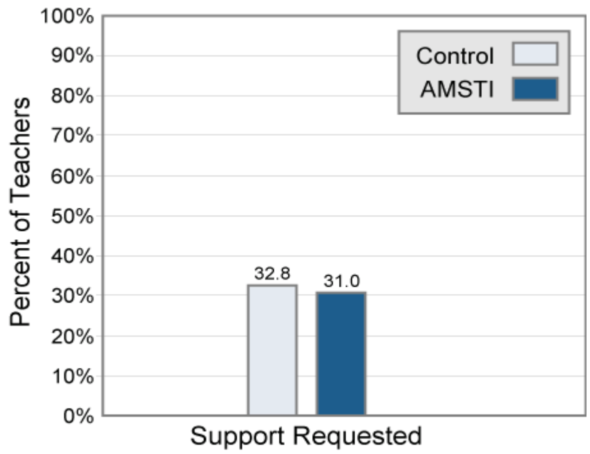
Note: n = 78 schools; 358 teachers. A single test was conducted to determine whether there was a significant difference between AMSTI and control schools in the reported level of access to materials and manipulatives. An ordinal logit model was conducted accounting for clustering of teachers within schools. The difference between the treatment and control groups was statistically significant ($p < .01$).
Source: Teacher survey data.

In-school supports. In three of the teacher surveys (those administered in February through April), teachers were asked to report separately the number of times they requested and received support within the past month. (Support was defined as mentoring or coaching for instruction.) Two variables were constructed by averaging the responses to questions about the number of times teachers requested and received support across survey occasions, yielding an average teacher value.[76] These variables were highly skewed, because of a large number of "no request" responses (see appendix W). The variables were therefore dichotomized based on whether teachers requested/received any support for instruction.

[76] For both outcomes, if a teacher responded on fewer than three occasions, responses were averaged over the smaller number of responses.

In mathematics there was no statistically significant difference between the AMSTI and control groups (p = .62): 31 percent of AMSTI mathematics teachers (n = 65) and 33 percent of control group mathematics teachers (n = 62) reported requesting support during this period (figure 3.10).

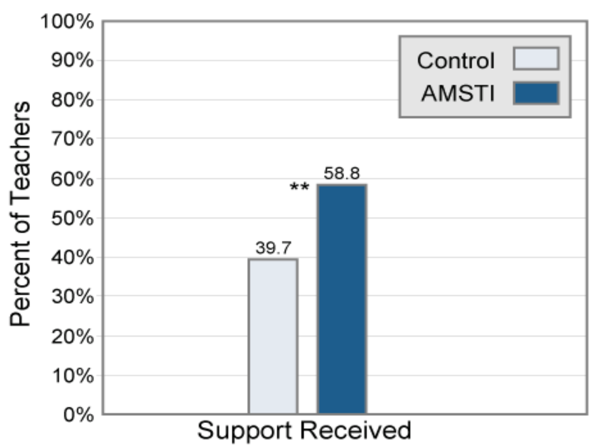
Figure 3.10 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control mathematics teachers requesting support for instruction



Note: n = 80 schools; 399 teachers. No significant difference based on a test of the differences between AMSTI and control teachers in the log-odds of requesting support for instruction.
Source: Student achievement data from tests administered as part of the state's accountability system.

AMSTI mathematics teachers were more likely to receive support than were control teachers: 59 percent of AMSTI mathematics teachers (n = 124) and 40 percent of control group mathematics teachers (n = 75) reported receiving support (figure 3.11). This difference was statistically significant ($p < .05$).

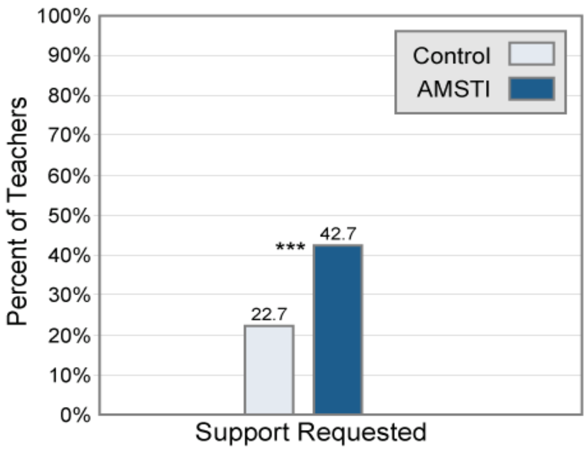
Figure 3.11 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control mathematics teachers receiving support for instruction



** Significant at $p < .05$ based on a test of the differences between AMSTI and control teachers in the log-odds of receiving support for instruction.
Note: n = 81 schools; 400 teachers.
Source: Teacher survey data.

AMSTI science teachers were more likely to request support than were control group science teachers: 43 percent of AMSTI science teachers (n = 82) and 23 percent of control group science teachers (n = 40) requested support for instruction (figure 3.12). This difference was statistically significant ($p < .01$).

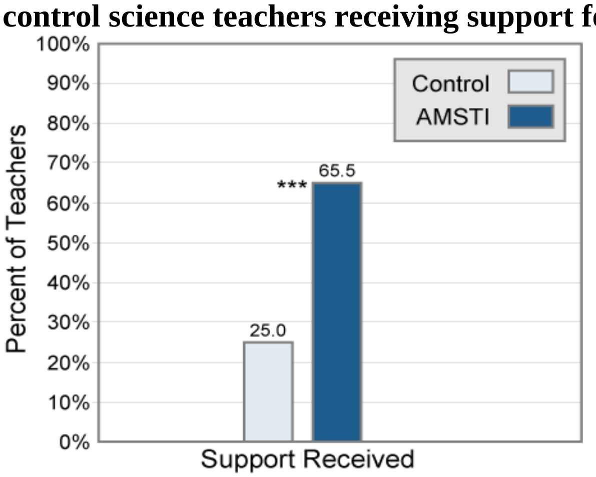
Figure 3.12 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control science teachers requesting support for instruction



*** Significant at $p < .01$ based on a test of the differences between AMSTI and control teachers in the log-odds of requesting support for instruction.
Note: n = 80 schools; 368 teachers.
Source: Teacher survey data.

AMSTI science teachers were more likely to receive support than were control group science teachers: 65 percent of AMSTI science teachers (n = 122) and 25 percent of control science teachers (n = 49) reported receiving support (figure 3.13). This difference was statistically significant ($p < .01$).

Figure 3.13 Percent of Alabama Math, Science, and Technology Initiative (AMSTI) and control science teachers receiving support for instruction



*** Significant at $p < .01$.based on a test of the differences between AMSTI and control teachers in the log-odds of receiving support for instruction.
Note: n = 80 schools; 367 teachers.
Source: Teacher survey data.

Conclusions

Extent to which the Alabama Math, Science, and Technology Initiative (AMSTI) was implemented in Year 1

Descriptive data on the professional development component of AMSTI in Year 1 indicate that, on average, trainers covered 85–96 percent of topics outlined in their training manuals to at least a moderate extent on at least one day of training. Grade 5 mathematics trainers covered 86 percent of topics, and grade 5 science trainers covered 85 percent of topics. Grade 7 mathematics and science trainers combined covered 96 percent of topics.

Both grade 5 and grade 7 trainers used the inquiry-based instructional methods of hands-on activities and lesson demonstration more than 25 percent of the time on more than half of training days. On average, trainers used hands-on activities more than 25 percent of the time on 4.0 out of 5 available days for mathematics, 4.5 out of 5 available days for science for grade 5, and 7.8 out of 10 available days for grade 7 mathematics and science combined. On average, trainers used lesson demonstrations more than 25 percent of the time on 3.5 out of 5 available training days for grade 5 and 6.3 out of 10 available training days for grade 7.

Trainers used the less inquiry-based methods of lecture, writing, and computer-based instruction more than 25 percent of the time on a minority of days. The average number of days (out of 5 available training days) grade 5 mathematics trainers used these instructional methods 25 percent of the time or more was zero for lecture and computer-based instruction and 0.8 for writing. Grade 5 science trainers spent 0.8 days on lecture, 0.5 on computer-based instruction, and 0.3 on writing. The average number of days (out of 10 available) that grade 7 trainers used these instructional methods 25 percent of the time or more was 4.3 for writing, 2.5 for lecture, and 1.5 for computer-based instruction.

About 58 percent of AMSTI teachers reported having full access to AMSTI materials, 32 percent reported having partial or limited access, and 10 percent reported having no access or had missing data. All of the principals responding to the interview question reported that follow-up support provided to teachers had been provided when needed.

Differences between Alabama Math, Science, and Technology Initiative (AMSTI) and control conditions during Year 1

AMSTI mathematics and science teachers were more likely than their control counterparts to receive professional development in the summer before Year 1: among mathematics teachers, 87 percent of AMSTI teachers ($n = 114$) and 24 percent of control teachers ($n = 27$) reported receiving professional development the summer before Year 1. The difference was statistically significant at $p < .01$. Among science teachers, 84 percent of AMSTI teachers ($n = 97$) and 24 percent of control teachers ($n = 22$) reported receiving professional development the summer before Year 1. This difference was statistically significant at $p < .01$.

AMSTI teachers reported greater access to materials and manipulatives than did control teachers during Year 1. Among mathematics teachers, AMSTI teachers were more likely than control teachers to report greater access to materials and manipulatives ($p < .01$): 78 percent of

AMSTI teachers (*n* = 152) and 41 percent of control teachers (*n* = 75) reported having access to all of most of the mathematics manipulatives. Among science teachers, AMSTI teachers were also more likely than control teachers to report greater access to materials and manipulatives ($p < .01$): 33 percent of control teachers (*n* = 58) and 61 percent of AMSTI teachers (*n* = 112) reported having all or most of the science manipulatives they needed.

Among mathematics teachers, AMSTI teachers were more likely than control teachers to receive support in Year 1 ($p < .05$): 59 percent of AMSTI teachers (*n* = 124) and 40 percent of control teachers (*n* = 75) reported receiving support. AMSTI teachers were not more likely than control teachers to have requested support for instruction ($p = .62$) in Year 1: 31 percent of AMSTI teachers (*n* = 65) and 33 percent of control group teachers (*n* = 62) reported requesting support during this period. Among science teachers, AMSTI teachers were more likely than control teachers to both request and receive instructional support in Year 1 (both significant at $p < .01$): 43 percent of AMSTI teachers (*n* = 82) and 23 percent of control teachers (*n* = 40) requested support for instruction; 65 percent of AMSTI teachers (*n* = 122) and 25 percent of control teachers (*n* = 49) reported receiving support.

4: Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on mathematics problem solving achievement, science achievement, and classroom instructional strategies after one year

The primary goal of the AMSTI intervention is to improve student achievement in mathematics and science. As posited by the AMSTI theory of action (see figure 1.1 in chapter 1), the mediating link between the intervention and student achievement is improving mathematics and science classroom practices through hands-on activities, inquiry-based activities, and practices promoting higher-order thinking skills.

This chapter presents confirmatory evidence on whether AMSTI had an effect on student performance in mathematics and science after one year. It also examines whether AMSTI had an effect on classroom practices after one year, as measured by the amount of active learning instruction in mathematics and science classrooms. (Results for the exploratory analysis of the two-year effects of AMSTI on student performance in mathematics and science are in chapter 5.) This chapter addresses the following primary and secondary confirmatory research questions:

- ***Primary confirmatory research question: effects on student achievement after one year***
  - What is the effect of AMSTI on:
    a. student achievement in mathematics problem solving *after one year*?
    b. student achievement in science *after one year*?

- ***Secondary confirmatory research question: effects on classroom practice after one year***
  - What is the effect of AMSTI on:
    a. the use of active learning instructional strategies by mathematics teachers *after one year*?
    b. the use of active learning instructional strategies by science teachers *after one year*?

Effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement after one year

As detailed below, AMSTI had a positive and statistically significant effect on student achievement in mathematics for students in grades 4–8 after one year of implementation, as measured by the SAT 10 mathematics problem solving assessment. The main effect across grade levels was 2.06 scale score points. The adjusted effect size for this outcome was 0.05 standard deviation, equivalent to a difference of 2 percentile points for the average control group student if the student had received AMSTI. Ten sensitivity analyses were conducted. All of the effect estimates from these tests were positive and within 1.08 points of one another on the scale of the outcome measure; all but one of these estimates were statistically significant.

The estimated effect of AMSTI on achievement of grades 5 and 7 students in science after one year, as measured by the SAT 10 science assessment, was not statistically significant. Nine sensitivity analyses were conducted, eight consistent with the finding of no effect.

As detailed below, AMSTI had a positive and statistically significant effect on the minutes of active learning strategies teachers reported using in mathematics and science classrooms (adjusted effect sizes of 0.47 standard deviation for mathematics and 0.32 standard deviation for science). This result is consistent with the AMSTI theory of action, which emphasizes such strategies as a way to improve achievement in mathematics and science.

Effect on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving

Descriptive results. Summary statistics for the analytic sample were obtained, and these are provided below (table 4.1).

Table 4.1 Sample statistics for analytic sample of grade 4–8 students used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year

Characteristic	AMSTI schools	Control schools
Number of schools	41	41
Number of teachers	244	229
Number of students	10,022	8,691
Unadjusted posttest mean score on SAT 10 test of mathematics problem solving	658.3	659.7
Standard deviation of posttest scores	43.5	42.4

Source: Student achievement data from tests administered as part of the state's accountability system.

Estimated effect. AMSTI had a statistically significant effect on student achievement in mathematics (table 4.2[77]).[78]The regression-adjusted estimate of the effect of AMSTI on end-of-year mathematics problem solving performance was 2.06 scale score units (standard error = 0.66, p = .004).[79] This estimate represented a difference of 0.05 standard deviation in favor of AMSTI schools, equivalent to a gain of 2 percentile points for the average control group student if the student had received AMSTI. The average effect of AMSTI can be translated into an estimated

[77] In the main body of the report, for the primary and secondary confirmatory analyses, we round the *p*-value to the third decimal place so that it can be compared to the level of statistical significance, which also is specified to the third decimal place (alpha=.025).
[78] A comparison of assumed statistical power population parameters with corresponding actual sample statistics is presented in appendix X, so that readers can evaluate the statistical power of the design. See appendixes Y–AB for the relevant effect estimates.
[79] The unadjusted posttest mean is lower for AMSTI schools than for control schools, whereas the main result presented in table 4.2 shows AMSTI schools scoring higher on average than control schools. This reversal is likely caused by regression adjustments in the statistical model used to estimate the effect, as detailed in chapter 2.

28 days of additional student progress beyond students receiving conventional mathematics instruction.[80]

Table 4.2 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year

AMSTI mean	Control group mean	Estimated effect	*p*-value	95 percent confidence interval	Effect size
661.8	659.7	2.1			
(43.5)	(42.4)	(0.7)	.004	[0.7, 3.4]	0.05

Note: Numbers in parentheses are standard deviations for the means and standard error for the estimated effect. Sample sizes are in table 4.1. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean. The *p*-value corresponds to the significance test for the effect of AMSTI in the regression model. The approach to reporting the effect size and *p*-value falls within the range of options considered acceptable in rigorous evaluations by the Institute of Education Sciences (Garet et al. 2010). The effect size was computed by dividing the regression-adjusted effect estimate by the standard deviation of the posttest scores for the control group. Between-grade differences in the posttest were factored out of the standard deviation in the denominator of the effect size.
Source: Student achievement data from tests administered as part of the state's accountability system.

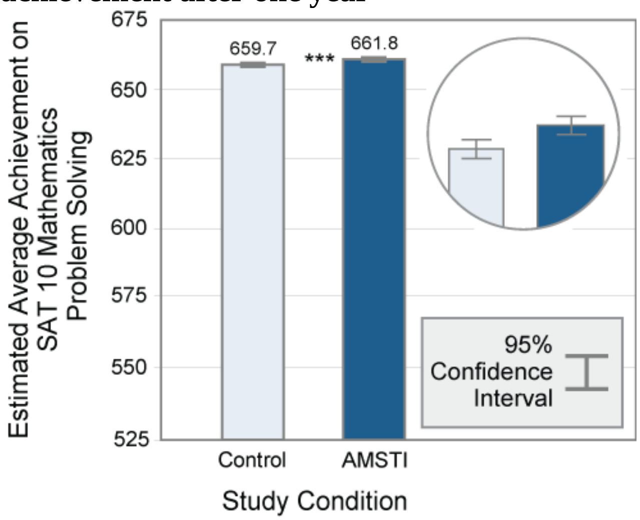
Although the study was powered to achieve a minimum detectable effect size of 0.20 standard deviation for estimating the impact for a single outcome, the estimated minimum detectable effect size, given the sample-based parameter estimates and figuring in the adjustment for multiple comparisons, was 0.063. This differed from the planned minimum detectable effect size for two reasons. First, estimates of the key power parameters differed from the values assumed in designing the study. The assumption was that covariates would account for only 64 percent of the between-school variance in the posttest. In fact, the observed sample statistics indicated that the covariates accounted for 97 percent of this variation. In addition, attrition was expected to reduce the sample of 82 schools to 66 by the end of the trial. This level of attrition did not occur. (A full comparison of the assumed parameter values used for the power analysis and the corresponding estimated values is in appendix X.) Second, the Bonferroni correction was adopted for multiple outcomes, which reduced the effective alpha level for the analysis.

It is possible to provide a visual representation of the results in table 4.2 (figure 4.1). The bar graph represents average performance on the SAT 10 mathematics problem solving assessment in AMSTI and control schools.[81] Below average performance on the SAT 10

[80] The Bonferroni adjustment was applied separately in the analysis of effects on students and the analysis of effects on classroom practices. In each case there were two comparisons, and the significance level was adjusted to .025.
[81] The goal of presenting the bar graphs that display the results of impact analyses is to communicate two pieces of information: (1) the point estimates for performance in each condition based on the benchmark regression model, and (2) the level of uncertainty associated with the impact estimate. To accommodate both pieces of information, the confidence intervals are constructed so that they overlap only in the event that there is lack of statistical significance at the .05 level. The combined length of both intervals is the length of the 95% confidence interval for the impact estimate. The difference in length between the bars corresponds to the regression-adjusted impact estimate from the benchmark model.

mathematics problem solving assessment in AMSTI and control schools is presented (see figure 4.1).

Figure 4.1 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year



*** Significant at $p < .01$.
Note: $n = 82$ schools; 18,713 students
The control mean is the unadjusted mean. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean.
Source: Student achievement data from tests administered as part of the state's accountability system.

Sensitivity analyses. Ten additional analyses were conducted to examine the sensitivity of the estimate of the effect of AMSTI on mathematics problem solving when the data are examined in different ways. These variations can affect both the estimated size of the effect and the *p*-value. The detailed results of these analyses are in appendix AC.

All of the sensitivity analyses yielded effect estimates that were positive and within 1.08 points of one another on the scale of the outcome measure. Nine yielded effect estimates that were statistically significant at the .025 level, which is consistent with the main finding. The exception was the sensitivity analysis in which the pretest and pairs were modeled as the only covariates and cases without a pretest were listwise deleted. In this analysis, the *p*-value for the impact estimate was .087.

Effect on Stanford Achievement Test Tenth Edition (SAT 10) science

Descriptive results. Summary statistics for the analytic sample were obtained, and these are provided below (table 4.3).

Table 4.3 Sample statistics for analytic sample of grade 5 and 7 students used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year

Characteristic	AMSTI schools	Control schools
Number of schools	39	40
Number of teachers	102	90
Number of students	4,082	3,446
Unadjusted posttest means[a]	653.5	654.6
Standard deviation of posttest scores	33.1	31.3

a. The posttest is the scale score for the SAT 10 science test.
Source: Student achievement data from tests administered as part of the state's accountability system.

*Estimated effect***.** The effect of AMSTI on end-of-year science performance was not statistically significant (table 4.4). The regression-adjusted estimate of the effect was 1.55 scale score units (standard error = 0.90, p = .092).[82]

Table 4.4 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year

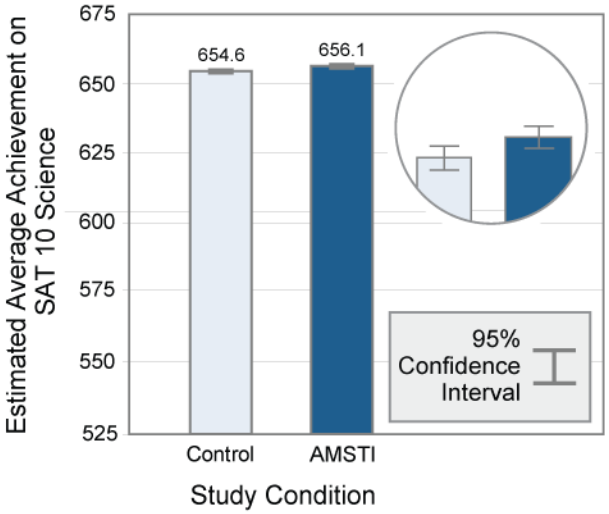
AMSTI mean	Control mean	Estimated effect	*p*-value	95 percent confidence interval	Effect size
656.1	654.6	1.6			
(33.1)	(31.3)	(0.9)	.092	[–0.3, 3.4]	0.05

Note: Numbers in parentheses are standard deviations for the means and standard error for the estimated effect. Sample sizes are in table 4.3. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean. The *p*-value corresponds to the significance test for the effect of AMSTI in the regression model. The approach to reporting the effect size and *p*-value falls within the range of options considered acceptable in rigorous evaluations by the Institute of Education Sciences (Garet et al 2010). The effect size was computed by dividing the regression-adjusted effect estimate by the standard deviation of the posttest scores for the control group. Between-grade differences in the posttest were factored out of the standard deviation in the denominator of the effect size.
Source: Student achievement data from tests administered as part of the state's accountability system.

[82]The unadjusted posttest mean is lower for AMSTI schools than for control schools, whereas the main result presented in table 4.4 shows AMSTI schools scoring higher on average than control schools. This reversal is likely caused by regression adjustments in the statistical model used to estimate the effect, as detailed in chapter 2.

Figure 4.2 provides a visual representation of the results in table 4.4.

Figure 4.2 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year



Note: n = 79 schools; 7,528 students. The control mean is the unadjusted mean. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean.
Source: Student achievement data from tests administered as part of the state's accountability system.

Sensitivity analyses. Several additional analyses were conducted to examine the sensitivity of the estimate of the effect of AMSTI on science achievement. All results except one were consistent with the main finding of no significant effect. The exception was a statistically significant result in the model that weighted schools equally by regressing average school posttests against average school values of the covariates and including among the independent variables the school proportions of students with missing values for each covariate. (When this model was run without the covariates that indicate the proportions of students with missing values, the estimate of the effect of AMSTI was not statistically significant.) Details of these analyses are in appendix AD.

Effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on classroom practices after one year

The secondary goal of the experiment was to estimate the effect of AMSTI on classroom practices by analyzing teachers' self-reported levels of use of active learning instructional strategies in mathematics and active learning instructional strategies in science. Analysis of teacher survey responses resulted in a measure of minutes of active learning strategies averaged over several two-week periods. The Alabama State Department of Education (ALSDE) provides guidelines to schools on the suggested number of minutes of instruction by subject per day. For grades 4–6, the recommended allocation is 60 minutes a day for mathematics and 45 minutes a day for science. Where grades 7 and 8 are housed with other elementary grades, schools may choose the time requirements listed for grades 4–6 or those listed for grades 7–12. For grades 7–12, ALSDE does not specify a daily recommended number of minutes for mathematics or science. Rather, a minimum of 140 clock hours of instruction is required for one unit of credit, and the Alabama High School Graduation Requirements include four credits each for mathematics and for science.[83]

Effect on active learning instructional strategies in mathematics classrooms

Descriptive results. Summary statistics for the analytic sample were obtained, and these are provided below (table 4.5).

Table 4.5 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms after one year

Characteristic	AMSTI schools	Control schools
Number of schools	41	40
Number of teachers	213	192
Mean minutes of active learning strategies for mathematics over two-week period[a]	192.6	129.1
Standard deviation	147.5	105.3

a. As explained in chapter 2, on four surveys administered from January to April, teachers reported the number of active learning strategies used in their mathematics classroom within a given two-week period. The mean was computed by averaging teachers' responses over the four surveys (or the number for which an outcome was available). This number represents the minutes of active learning strategies averaged over several 10-day periods.
Source: Teacher survey data.

Estimated effect. The effect of AMSTI on active learning strategies for mathematics instruction was statistically significant (table 4.6). The regression-adjusted effect estimate was 49.83 (standard error = 11.49, $p < .001$), which represented a difference of 0.47 standard deviation in favor of AMSTI schools.

It is possible to provide a visual representation of the results in table 4.6 (figure 4.3).

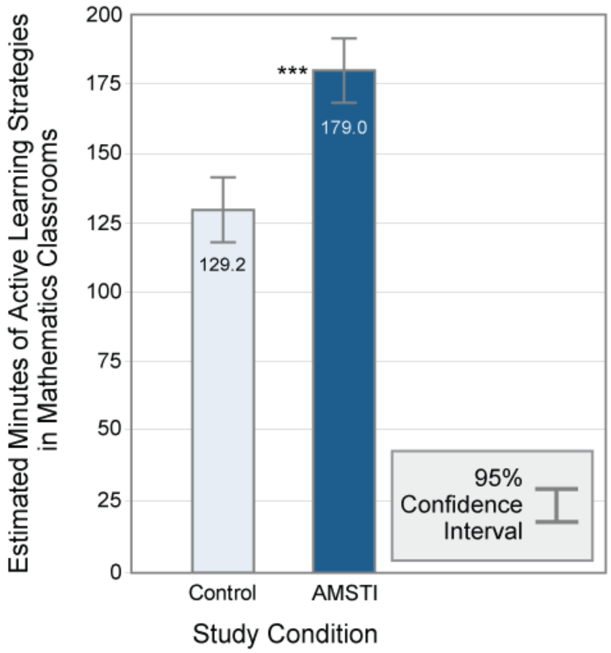
Table 4.6 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms after one year

AMSTI mean	Control mean	Estimated effect	*p*-value	95 percent confidence interval	Effect size
179.00	129.2	49.8			
(147.5)	(105.3)	(11.5)	< .001	[26.6, 73.1]	0.47

Note: Numbers in parentheses are standard deviations for the means and standard error for the estimated effect. Sample sizes are in table 4.5. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean. The *p*-value corresponds to the significance test for the effect of the regression model. The approach to reporting the effect size and *p*-value falls within the range of options considered acceptable in rigorous evaluations by the Institute of Education Sciences (Garet et al. 2010). The effect size was computed by dividing the regression-adjusted effect estimate by the standard deviation of the posttest scores for the control group.
Source: Teacher survey data**.**

[83] Retrieved doc May 14, 2010 from https://docs.alsde.edu/documents/54/07sciapp

Figure 4.3 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms after one year



*** Significant at $p < .01$.
Note: $n = 81$ schools; 405 teachers. The control mean is the unadjusted mean. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean.
Source: Student achievement data from tests administered as part of the state's accountability system.

Sensitivity analyses. Several analyses were conducted to examine the sensitivity of the estimate of the effect of AMSTI on active learning instructional strategies in mathematics classrooms. All results were consistent with the main findings of a statistically significant effect on active learning instructional strategies in mathematics. The details of these analyses are in appendix AE.[84]

[84] The estimation routines converged for all but one of the analyses, yielding orthogonal variance components estimates at the school and student levels. The exception was the analysis that used full maximum likelihood estimation instead of restricted maximum likelihood estimation, which did not produce an estimate for the variance component at the school level. For that analysis. the estimated G matrix was not positive definite.

Effect on active learning instructional strategies in science classrooms

Descriptive results. Summary statistics of the analytic sample were obtained, and these are provided below (table 4.7).

Table 4.7 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms after one year

Characteristic	AMSTI schools	Control schools
Number of schools	40	38
Number of teachers	194	175
Mean minutes of active learning strategies for science over two-week period[a]	181.9	127.0
Standard deviation	134.8	125.8

a. As explained in chapter 2, on four surveys administered from January to April, teachers reported the number of active learning strategies used in their science classroom within a given two-week period. The mean was computed by averaging teachers' responses over the four surveys (or the number for which an outcome was available). This number is the minutes of active learning strategies averaged over several two-week periods.
Source: Student achievement data from tests administered as part of the state's accountability system.

Estimated effect. The effect of AMSTI on active learning strategies in science instruction was statistically significant (table 4.8). The regression-adjusted effect estimate was 40.07 (standard error = 11.77, p = .002), representing a difference of 0.32 standard deviation in favor of AMSTI schools.

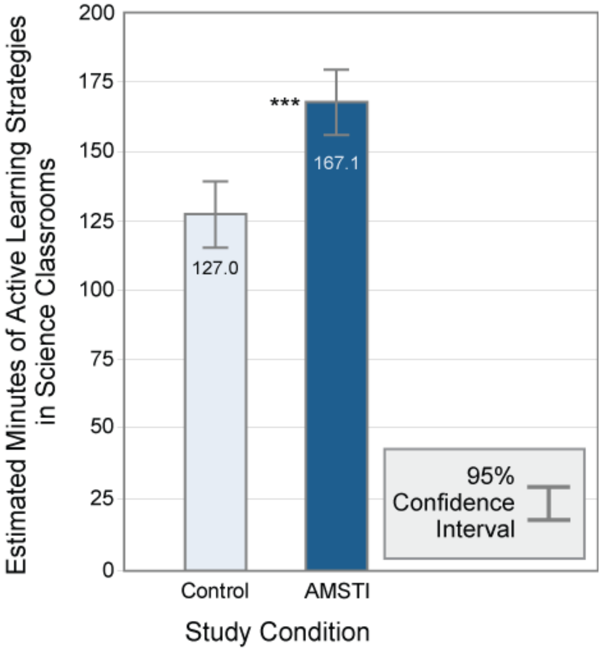
Table 4.8 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms after one year

AMSTI mean	Control mean	Estimated effect	*p*-value	95 percent confidence interval	Effect size
167.1	127.0	40.1			
(134.8)	(125.8)	(11.8)	.002	[16.2, 63.9]	0.32

Note: Numbers in parentheses are standard deviations for mean values and standard error for the estimated effect. Sample sizes are in table 4.7. The control mean is the unadjusted mean. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean. The *p*-value corresponds to the significance test for the effect from the regression model. The approach to reporting the effect size and *p*-value falls within the range of options that have been considered acceptable in rigorous evaluations by the Institute of Education Sciences (Garet et al. 2010). The effect size is computed by dividing the regression-adjusted effect estimate by the standard deviation of the posttest scores for the control group.
Source: Teacher survey data.

It is possible to provide a visual representation of the results in table 4.8 (figure 4.4).

Figure 4.4 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms after one year



*** Significant at $p < .01$.
Note: n = 78 schools; 369 teachers. The control mean is the unadjusted mean. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean.
Source: Teacher survey data.

Sensitivity analyses. Several analyses were conducted to examine the sensitivity of the effect estimate of AMSTI on active learning instructional strategies in science. All results from these analyses were consistent with the main findings of a statistically significant effect. The details of these analyses are in appendix AF.[85]

---

[85] The estimation routines converged for all but one of the analyses, yielding orthogonal variance components estimates at the school and student levels. The exception was the analysis that used full maximum likelihood estimation instead of restricted maximum likelihood estimation, which did not produce an estimate for the variance component at the school level. For that analysis, the estimated G matrix was not positive definite.

5: Effect on mathematics problem solving and science achievement after two years

AMSTI was developed as a two-year intervention. Thus, the full effect is expected to be felt only after two years. The AMSTI evaluation used random assignment to assess the effect of AMSTI on student achievement and classroom practices after one year. In Year 2, the AMSTI and no AMSTI conditions could no longer be compared using standard randomized controlled trial methods, because teachers in the initial control group schools began using AMSTI in their classrooms in the second year of the experiment. An exploratory method was therefore necessary to estimate the effects of the intervention on student achievement in mathematics problem solving and student achievement in science over the full two years of implementation.

The method used was developed by Bell and Bradley (2008, 2009). Although the method uses the experimental structure of the data, it requires an added assumption not necessary in conventional experimental analyses, making the results exploratory rather than confirmatory. It assumes that the impact of the initial year of AMSTI remains constant over time—that the program affects students in the control group schools in Year 2 the same way it did students in the intervention group schools a year earlier.

Table 2.16 in chapter 2 provides a list of conditions that, if met, result in this assumption of equal effects in the first year of implementation being true and, hence, the Bell-Bradley estimator yielding unbiased estimates of two-year effects. The study team was able to assess to some extent whether some of these conditions hold using data collected for the evaluation (see appendix AG). For 61 of the 66 indicators tested, the evidence suggests the tested conditions hold. While the intervention and control samples are comparable on most of the tested factors, unmeasured differences may still exist. If so, estimated two-year effects may be biased if on net the assumption of equal effects in the first year of AMSTI exposure is not met.

This chapter presents an assessment of AMSTI's possible two-year effects on student achievement. This evidence is used to address the exploratory research question below.

- ***Exploratory research question: effects on student achievement after two years***
  - What is the effect of AMSTI on:
    a. student achievement in mathematics problem solving *after two years*?
    b. student achievement in science *after two years*?

Given the exploratory nature of the analysis, the outcomes were not adjusted for multiple comparisons (Schochet 2008). Due to the uncertainty introduced by the assumptions required by the analysis of the results presented in this chapter, the findings are considered exploratory and thus only meant to *suggest* that a two-year effect may be present for both mathematics and science and warrant further research on the longer-term effect of AMSTI.

Effect on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement

Summary statistics for the analytic sample were obtained, and these are provided below (table 5.1). (See chapter 2 and appendices Q and T for details on construction of the estimator.) Students for whom mathematics outcome data were missing for the year in question were excluded from the sample for that year.

Table 5.1 Summary statistics on analytic sample used to determine estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after two years

	Sample for first year of outcome data		**Sample for second year of outcome data**	
Characteristic	**AMSTI**	**Control**	**AMSTI**	**Control**
Number of schools	41	41	40	41
Number of teachers	243	229	208	164
Number of students	9,520	8,474	9,386	8,144
Unadjusted posttest mean scale score for SAT 10 mathematics problem solving[a]	658.5	660.3	659.3	660.7
Standard deviation	43.8	42.5	43.5	43.6

a. As explained in chapter 2 and in detail in appendix R, the sample size indicated in this table as contributing to the first year of outcome data is different from the sample size used in the one-year confirmatory analysis, because students who skipped or repeated a grade in Year 2 were excluded from the sample used to estimate the two-year effect.
Source: Student achievement data from tests administered as part of the state's accountability system.

The estimated effect of two years of AMSTI on mathematics achievement was statistically significant (table 5.2). The regression-adjusted estimate of this effect was 3.74 scale score units (adjusted standard error = 1.66, p = .030).[86] This estimate represents a difference of 0.10 standard deviation, equivalent to a 4 percentage point gain over the achievement level of the average non-AMSTI student. This estimate of the average effect of AMSTI after two years translates to an estimated 50 days of additional student progress at the rate of advancement experienced by students receiving conventional mathematics instruction.

---

[86] The standard error for the mathematics problem solving outcome was adjusted to reflect the nesting of scores within students, which was not modeled. The post hoc adjustment is described in appendix AH. The original preadjustment robust standard error was 1.63.

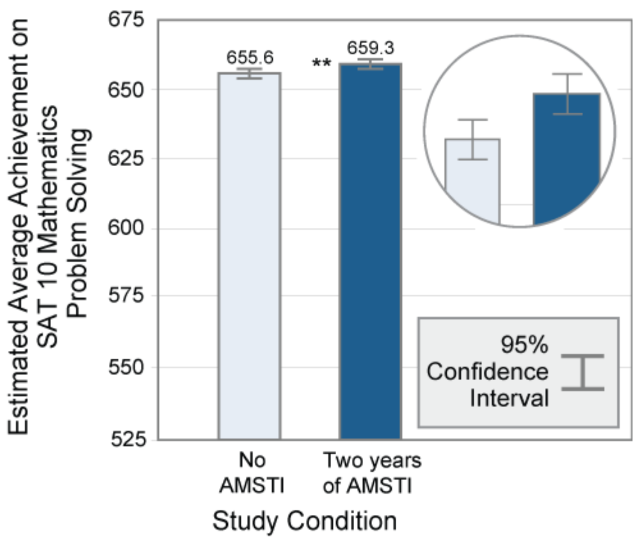
It is possible to provide a visual representation of the results in table 5.2 (figure 5.1).

Table 5.2 Estimated effect of two years of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after two years

Mean outcome after two years of AMSTI	Mean outcome after two years without AMSTI	Estimated effect of AMSTI	*p*-value	95 percent confidence interval	Effect size
659.3	655.6	3.7 (1.7)	.03	[0.4 – 7.1]	0.10

Note: Number in parentheses is standard error. Sample sizes are in table 5.1. The mean outcome after two years of AMSTI was the unadjusted mean outcome for the students in the AMSTI schools in Year 2. The mean outcome after two years without AMSTI was obtained by subtracting the regression-adjusted estimated two-year effect of AMSTI from the mean outcome after two years of AMSTI. The effect size was computed by dividing the regression-adjusted estimated two-year effect by the pooled standard deviation of the Year 2 achievement score for the groups of students. Between-grade differences in the Year 2 achievement scores are factored out of the standard deviation in the denominator of the effect size. The *p*-value corresponds to the significance test for the estimated two-year effect of AMSTI from the regression model. See appendix AI for parameter estimates for the models used to generate the relevant effect estimates.
Source: Student achievement data from tests administered as part of the state's accountability system.

Figure 5.1 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after two years



** Significant at $p < .05$.
Note: The Bell-Bradley estimate uses a sample for the first year of outcome data (n = 82 schools; 17,994 students) and a sample for the second year of outcome data (n = 81schools; 17,530 students).
Source: Student achievement data from tests administered as part of the state's accountability system.

Effect on Stanford Achievement Test Tenth Edition (SAT 10) science achievement

The samples for science were smaller than those for mathematics, because the SAT 10 science test is mandatory only for grades 5 and 7. Students for whom science outcome data were missing for the year in question were excluded from the sample for that year (table 5.3).

Table 5.3 Sample statistics for analytic sample used to determine estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after two years

	Sample for first year of outcome data		Sample for second year of outcome data	
Characteristic	**AMSTI**	**Control**	**AMSTI**	**Control**
Number of schools	39	40	38	40
Number of teachers	101	90	78	60
Number of students	3,914	3,364	3,843	3,161
Unadjusted posttest mean[a]	654.3	655.0	654.7	653.4
Standard deviation of posttest scores	33.1	31.3	33.2	30.6

a. The posttest is the scale score for the SAT 10 test of science. As explained in chapter 2 and in detail in appendix R, the sample size listed in this table as contributing to the first year of outcome data is different from the sample size used in the one-year confirmatory analysis, because students who skipped or repeated a grade in Year 2 were excluded from the sample used to estimate the two-year effect.

Source: Student achievement data from tests administered as part of the state's accountability system.

The estimated effect of two years of AMSTI on science achievement for students in grades 5 and 7 was statistically significant (table 5.4). The regression-adjusted estimate of the two-year effect was 4.00 scale score units (standard error = 1.90, $p = .038$), representing a difference of 0.13 standard deviation, equivalent to a 5 percentage point gain over the achievement level of the average non-AMSTI student. Because the SAT 10 science test is required only for grades 5 and 7 in Alabama, the science pretest scores needed to translate the two-year effect of AMSTI on science achievement into an equivalent number of conventional instructional days were not available.

Table 5.4 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after two years

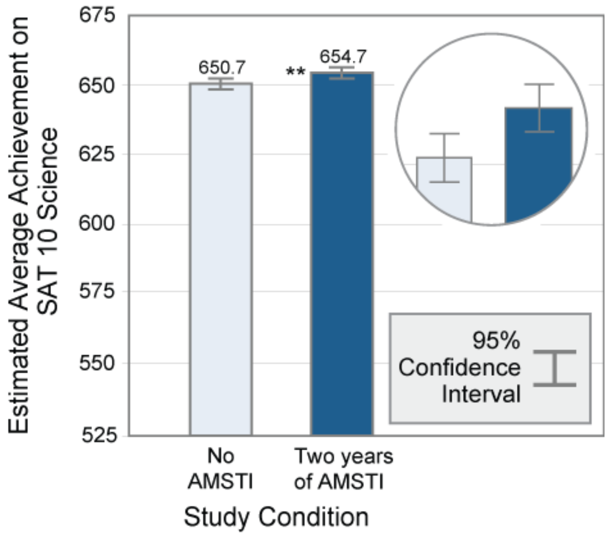
Mean outcome with two years of AMSTI	Mean outcome after two years without AMSTI	Estimated effect of AMSTI	*p*-value	95 percent confidence interval	Effect size
654.7	650.7	4.0 (1.9)	.04	[0.2 – 7.8]	0.13

Note: Number in parentheses is standard error. Sample sizes are in table 5.3. The mean outcome with two years of AMSTI is the unadjusted mean outcome for the students in the AMSTI schools in Year 2. The mean outcome after two years without AMSTI is obtained by subtracting the regression-adjusted estimated two-year effect of AMSTI from the mean outcome after two years of AMSTI. The effect size is computed by dividing the regression-adjusted estimated two-year effect by the pooled standard deviation of the Year 2 achievement score for the two groups of students. Between-grade differences in the Year 2 achievement scores are factored out of the standard deviation in the denominator of the effect size. The *p*-value corresponds to the significance test for the estimated two-year effect of AMSTI from the regression model. See appendix AI for parameter estimates for the models used to generate the relevant effect estimates.

Source: Student achievement data from tests administered as part of the state's accountability system.

It is possible to provide a visual representation of the results in table 5.4 (figure 5.2).

Figure 5.2 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after two years



** Significant at $p < .05$.
Note: The Bell-Bradley estimate uses a sample for the first year of outcome data (n = 79 schools; 7,278 students) and a sample for the second year of outcome data (n = 78 schools; 7,004 students).
Source: Student achievement data from tests administered as part of the state's accountability system.

The analysis of the two-year impact of AMSTI on student achievement is exploratory. Readers should exercise caution in interpreting the results. For instance, we remind the reader that with exploratory analyses multiplicity adjustments are not performed. As a consequence, a less strict criterion is used with exploratory analyses to decide whether a particular result achieves statistical significance, with the drawback that this increases the probability of finding a spuriously significant impact. Two-year effects on both mathematics problem solving (p = .030) and science (p = .038) reach statistical significance under the less strict criterion (alpha = .05). Under the more strict criterion used with the primary confirmatory analyses (alpha = .025), these results would not have been considered statistically significant.

6: Effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading achievement, teacher content knowledge, student engagement, and variations in effects for student subgroups after one year

This chapter presents evidence on additional exploratory questions that are important to understanding the full effects of AMSTI and identifying ways of improving the program. Specifically, it explores whether participation in AMSTI had an effect on reading achievement, as quasi-experimental evaluations have found (Miron and Maxwell 2007). It also presents evidence on whether AMSTI had a positive effect on two intermediate outcomes, teacher content knowledge and student engagement, as hypothesized in the AMSTI theory of action. In addition, the chapter also provides evidence on variations in the effects of AMSTI on achievement for subgroups of students.

As exploratory findings, these results are not meant to support firm conclusions, but rather to suggest hypotheses for further research and provide insights that could help improve program services. Given the exploratory nature of these analyses, the outcomes were not adjusted for multiple comparisons (Schochet 2008).

This chapter addresses the following exploratory research questions:

- ***Effects on student achievement in reading after one year***
  - What is the effect of AMSTI on student achievement in reading *after one year*?

- ***Effects on teacher content knowledge and student engagement after one year***
  - What is the effect of AMSTI on:
    a. mathematics teachers' reported levels of content knowledge *after one year*?
    b. science teachers' reported levels of content knowledge *after one year*?
  - What is the effect of AMSTI on:
    a. mathematics teachers' reported levels of student engagement *after one year*?
    b. science teachers' reported levels of student engagement *after one year*?

- ***Variations in effects on student achievement for specific subgroups of students after one year***
  - Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by student pretest score? What is the effect of AMSTI on these outcomes after one year for students with pretest scores that fall in the low, middle, and high ranges?
  - Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by low-income status, proxied by enrollment in the free or reduced-price lunch program? What is the effect of AMSTI on these outcomes after one year for students enrolled in the free or reduced-price lunch program and students who are not enrolled?
  - Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by racial/ethnic minority status? What is the effect of AMSTI on these outcomes after one year for racial/ethnic minorities and for White students?

- Do the one-year effects of AMSTI on student achievement in (a) mathematics problem solving, (b) science, and (c) reading vary by gender? What is the effect of AMSTI on these outcomes after one year for boys and for girls?

Effects on student achievement in reading

Summary statistics for the analytic sample were obtained, and these are provided below (table 6.1).

Table 6.1 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) reading achievement after one year

Characteristic	AMSTI schools	Control schools
Number of schools	41	41
Number of teachers	229	210
Number of students	10,019	8,691
Unadjusted posttest means[a]	663.4	663.2
Standard deviation of posttest scores	38.2	36.8

a. The posttest is the scale score for the SAT 10 test of reading.
Source: Student achievement data from tests administered as part of the state's accountability system.

AMSTI had a statistically significant effect on student achievement in reading (table 6.2). The regression-adjusted estimate of its effect on end-of-year reading performance was 2.34 scale score units (standard error = 0.47, $p < .001$), a difference of 0.06 standard deviation in favor of AMSTI schools. This effect is equivalent to a gain of 2 percentile points for the average control group student if the student had received AMSTI. It can be translated to an estimated 40 days of additional student progress over students receiving conventional reading instruction.[87]

[87] The section reporting the effect of AMSTI on mathematics problem solving in chapter 4 discusses plausible reasons for why a statistically significant result was obtained, given a smaller measured effect size than the study was powered to detect. A similar argument applies here. The sample associated with the SAT 10 reading outcome was almost exactly the same as for the sample associated with the SAT 10 mathematics problem solving outcome after one year.

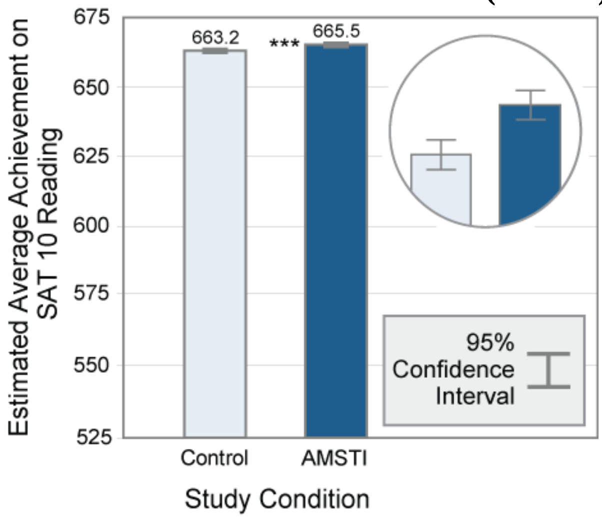
Table 6.2 Estimated effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) reading after one year

AMSTI mean	Control group mean	Estimated effect	*p*-value	95 percent confidence interval	Effect size
665.5	663.2	2.3			
(38.2)	(36.8)	(0.5)	< .01	[1.4, 3.3]	0.06

Note: Numbers in parentheses are standard deviations for means and standard error for the estimated effect. Sample sizes are in table 6.1. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean. The *p*-value corresponds to the significance test for the effect of AMSTI in the regression model. The approach to reporting the effect size and *p*-value falls within the range of options considered acceptable in rigorous evaluations by the Institute of Education Sciences (Garet et al. 2010). See appendix AJ for parameter estimates for the models used to generate the relevant effect estimates. The effect size was computed by dividing the regression-adjusted effect estimate by the standard deviation of the posttest scores for the control group. Between-grade differences in the posttest were factored out of the standard deviation in the denominator of the effect size.
Source: Student achievement data from tests administered as part of the state's accountability system.

It is possible to provide a visual representation of the results in table 6.2 (figure 6.1).

Figure 6.1 Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) reading achievement after one year



*** Significant at $p < .01$.
Note: n = 82 schools; 18,710 students
The control mean is the unadjusted mean. The AMSTI mean was obtained by adding the regression-adjusted estimate of the average one-year effect of AMSTI to the unadjusted control mean.
Source: Student achievement data from tests administered as part of the state's accountability system.

Effects on teacher content knowledge and student engagement

Teachers rated their content knowledge at their current grade level on a 5-point Likert scale (very low, low, moderate, high, very high), with a sixth option of "not applicable."[88] Teachers rated the average level of student engagement in their classes on a 5-point Likert scale (not engaged, slightly engaged, moderately engaged, almost fully engaged, fully engaged).

Effect on teacher-reported content knowledge in mathematics

Summary statistics for the analytic sample were obtained, and these are provided below (table 6.3).

Table 6.3 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher content knowledge in mathematics after one year

Characteristic	AMSTI	Control
Number of schools	41	40
Number of teachers	197	187

Source: Teacher survey data.

AMSTI mathematics teachers were not more likely to report higher levels of content knowledge than were control mathematics teachers:[89] 92 percent of AMSTI mathematics teachers (n = 81) and 94 percent of control mathematics teachers (n = 175) reported high or very high levels of content knowledge (table 6.4). The difference in the distribution of responses was not statistically significant, with a regression-adjusted estimate of the average difference between AMSTI and control groups in the cumulative odds of response of 1.21 (p = .408).

Table 6.4 Teacher counts in each response category for teacher content knowledge in mathematics

Response categories	AMSTI teachers	Control teachers	Total teachers
Very low, Low, Moderate[a]	16	12	28
	(8)	(6)	(7)
High	63	77	140
	(32)	(41)	(36)
Very high	118	98	216
	(60)	(52)	(56)

Note: Numbers in parentheses are percentages. Numbers may not sum to 100 percent because of rounding.
a. Response categories were collapsed when fewer than three teachers responded in a specific category.
Source: Teacher survey data.

---

[88] A response of na cannot be assigned a meaningful numeric value and therefore was coded as missing and dropped from analysis.
[89] The model with dummy variables for pairs did not converge; therefore, this result is from a model where pair effects were excluded.

Effect on teacher-reported content knowledge in science

Summary statistics for the analytic sample were obtained, and these are provided below (table 6.5).

Table 6.5 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher content knowledge in science after one year

Characteristic	AMSTI	Control
Number of schools	40	39
Number of teachers	176	170

Source: Teacher survey data.

AMSTI science teachers were not more likely to report higher levels of content knowledge than were control science teachers:[90] 74 percent of AMSTI science teachers (n = 130) and 69 percent of control science teachers (n = 118) reported high or very high levels of content knowledge (table 6.6). The difference in the distribution of responses was not statistically significant, with a regression-adjusted estimate of the average difference between AMSTI and control groups in the cumulative odds of response of 1.41 (p = .129).[91]

Table 6.6 Teacher counts in each response category for teacher content knowledge in science

Response categories	AMSTI teachers	Control teachers	Total teachers
Very low, Low, Moderate[a]	46	52	98
	(26)	(31)	(28)
High	77	71	148
	(44)	(42)	(43)
Very high	53	47	100
	(30)	(28)	(29)

Note: Numbers in parentheses are percentages. Numbers may not sum to 100 percent because of rounding.
a. Response categories were collapsed when fewer than three teachers responded in a specific category.
Source: Teacher survey data.

[91] Estimation of the school-level random effect led to the variance component reaching a boundary constraint. The variance component for schools was assigned a value of 0, with no p-value. The effect was therefore excluded. Without the school-level random effect, the estimate of impact is nonsignificant. Had it been possible to model a school-level random effect, incorporating the dependencies among observations within schools would have reduced the precision of the effect estimate, suggesting that the result would have remained nonsignificant if the effects of clustering had been modeled.

Effect on teacher-reported student engagement in mathematics classrooms

Summary statistics for the analytic sample were obtained, and these are provided below (table 6.7).

Table 6.7 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on student engagement in mathematics after one year

Characteristic	AMSTI	Control
Number of schools	41	40
Number of teachers	197	188

Source: Teacher survey data.

AMSTI mathematics teachers were more likely to report higher levels of student engagement than control mathematics teachers: 70 percent of AMSTI mathematics teachers (n = 137) and 59 percent of control mathematics teachers (n = 110) reported students were almost fully or fully engaged (table 6.8). The difference in the distribution of responses was statistically significant, with a regression-adjusted estimate of the average difference between AMSTI and control groups in the cumulative odds of response of 1.76 (p = .024).[92] (None of the teachers reported that students were not engaged.)

Table 6.8 Teacher counts in each response category for student engagement in mathematics

Response categories	AMSTI teachers	Control teachers	Total teachers
Not engaged	0	0	0
	(0)	(0)	(0)
Slightly engaged	10	16	26
	(5)	(9)	(7)
Moderately engaged	50	62	112
	(25)	(33)	(29)
Almost fully engaged	95	91	186
	(48)	(48)	(48)
Fully engaged	42	19	61
	(21)	(10)	(16)

Note: Numbers in parentheses are percentages. Numbers may not sum to 100 percent because of rounding.
Source: Teacher survey data.

[92] The model with dummy variables for pairs did not converge; therefore, this result is from a model where pair effects were excluded.

Effect on teacher-reported student engagement in science classrooms

Summary statistics for the analytic sample were obtained, and these are provided below (table 6.9).

Table 6.9 Sample statistics for analytic sample used to determine effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher student engagement in science after one year

Characteristic	AMSTI	Control
Number of schools	40	40
Number of teachers	176	172

Source: Teacher survey data.

AMSTI science teachers were more likely to report higher levels of student engagement than control science teachers: 77 percent of AMSTI science teachers (n = 135) and 56 percent of control science teachers (n = 96) reported students were almost fully or fully engaged (table 6.10). The difference in the distribution of responses was statistically significant, with a regression-adjusted estimate of the average difference between AMSTI and control groups in the cumulative odds of responding to each category of 3.32 (p = .003).

Table 6.10 Teacher counts in each response category for student engagement in science

Response categories	AMSTI teachers	Control teachers	Total teachers
Not engaged, Slightly engaged	8	11	19
	(5)	(6)	(6)
Moderately engaged	33	65	98
	(19)	(38)	(26)
Almost fully engaged	89	76	165
	(51)	(44)	(47)
Fully engaged	46	20	66
	(26)	(12)	(19)

Note: Numbers in parentheses are percentages. Numbers may not sum to 100 percent because of rounding.
a. Response categories were collapsed when fewer than three teachers responded in a specific category.
Source: Teacher survey data.

Subgroup differences in average effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement after one year

To further explore potential effects of AMSTI, the study examined whether AMSTI affected some subgroups of students and not others. To do this, the main impact model was expanded to include a subgroup dummy variable and a term interacting treatment status with the subgroup dummy.[93] This step yielded the coefficient for the interaction term and the associated statistical significance. The impact for each subgroup was also estimated. These analyses were conducted for each student outcome (mathematics problem solving, science, and reading).

Subgroups were formed based on characteristics of students before random assignment, including racial/ethnic minority status, eligibility for free or reduced-price lunch, gender, and pretest level (table 6.11). As explained in appendix B, there were three primary reasons for selecting these subgroups. First, the No Child Left Behind Act of 2001 requires states to disaggregate student achievement data for these subgroups and to intervene to close achievement gaps. Second, AMSTI aims to improve achievement for all students. Determining whether AMSTI has differential effects for subgroups of students reveals whether AMSTI is meeting its goal of providing equitable opportunities to learn. Third, the results of these analyses can inform AMSTI developers on where to focus program improvements to strengthen the intervention for groups receiving less benefit.

The moderating effects of racial/ethnic minority status, eligibility for free or reduced-priced lunch, gender, and pretest level were examined individually. No statistically significant subgroup differences in the one-year effect of AMSTI on student achievement in mathematics problem solving or science were found. (No test of differential effects achieved a *p*-value lower than .05.) However, there was a significant differential effect of AMSTI on student performance in reading depending on racial/ethnic minority status, with the effect of AMSTI favoring Whites.[94] The effect of AMSTI on reading achievement was not statistically significant for minority students ($p = .294$) but statistically significant and positive for White students ($p < .001$).[95] Besides this, there were no statistically significant differential effects on student achievement in reading for the categories considered.

---

[93] For the analysis of the moderating effect of the pretest, students were categorized into three levels depending on their pretest scores. The expanded model used two dummy variables to indicate membership in the three categories along with two interactions between these dummy variables and the indicator of treatment status.

[94] The racial/ethnic minority status for this study was coded as 0 for White students and 1 for minority students. For the analytic sample associated with the SAT 10 reading outcome (after one year), 57% were White, and 43% were minorities (39% Black, 4% Hispanic, Asian, or Native Americans/Alaskans).

[95] See appendixes AM–AO for parameter estimates for the models used to generate the relevant effect estimates.

Table 6.11 Average effects and variation in effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on student achievement for subgroups of students after one year

		Mathematics problem solving			Science			Reading		
Covariate		**Estimated effect**	***p*-value**	**Effect size[a]**	**Estimated effect**	***p*-value**	**Effect Size[a]**	**Estimated effect**	***p*-value**	**Effect size[a]**
	Average effect	2.1 (0.7)	.004	0.05	1.6 (0.9)	.092	0.05	2.3 (0.5)	< .01	0.06
Racial/ethnic minority status	Differential effect (added effect for minority students)	–1.1 (1.2)	.37	–0.03	–0.9 (1.3)	.49	–0.03	–3.0 (0.9)	< .01	–0.08
	Effect for minority students	0.9 (0.8)	.27	0.02	0.8 (0.9)	.37	0.03	0.7 (0.6)	.29	0.02
	Effect for White students	3.9 (0.9)	< .01	0.09	2.1 (1.1)	.07	0.07	3.1 (0.5)	< .01	0.08
Free or reduced-priced lunch status	Differential effect (added effect for students enrolled in the free or reduced-price lunch program)	–1.2 (1.8)	.26	–0.03	–2.1 (1.1)	.06	–0.07	–0.9 (0.8)	.26	–0.02
	Effect for students enrolled in the free or reduced-priced lunch program	1.5 (0.7)	.04	0.04	0.8 (0.9)	.42	0.02	1.8 (0.6)	.04	0.05
	Effect for students not enrolled in the free or reduced-priced lunch program	3.4 (0.9)	< .01	0.08	2.2 (1.1)	.06	0.07	2.3 (0.4)	< .01	0.06
Gender	Differential effect (added effect for boys)	–0.1 (0.8)	.91	< –0.01	0.0 (1.2)	.97	< 0.01	0.0 (0.7)	.99	< 0.01
	Effect for boys	2.2 (0.8)	< .01	0.05	1.8 (1.0)	.08	0.06	2.4 (0.5)	< .01	0.07
	Effect for girls	2.0 (0.7)	< .01	0.05	1.2 (0.9)	.17	0.04	2.2 (0.6)	< .01	0.06
SAT 10 reading pretest[b]	Differential effect	†	.83	†	†	.08	†	†	.57	†
	Effect for low pretest group (stanines 1–3)	0.3 (0.7)	.71	0.01	–1.8 (1.4)	.20	–0.09	0.6 (0.7)	.41	0.02
	Effect for middle pretest group (stanines 4–6)	1.3 (0.9)	.16	0.03	1.7 (1.0)	.11	0.05	1.2 (0.7)	.10	0.03

Covariate		Mathematics problem solving Estimated effect	*p*-value	Effect size[a]	Science Estimated effect	*p*-value	Effect Size[a]	Reading Estimated effect	*p*-value	Effect size[a]
	Effect for high pretest group (stanines 7–9)	0.9 (1.6)	.58	0.02	–0.9 (1.7)	.61	–0.03	0.3 (0.9)	.78	0.01
SAT 10 mathematics problem-solving pretest[b]	Differential effect	†	.26[c]	†	na	na	na	na	na	na
	Effect for low pretest group (stanines 1–3)	0.4 (0.8)	.63	0.01	na	na	na	na	na	na
	Effect for middle pretest group (stanines 4–6)	1.6 (0.7)	.04	0.04	na	na	na	na	na	na
	Effect for high pretest group (stanines 7–9)	1.0 (1.5)	.50	0.02	na	na	na	na	na	na

na is not applicable.

† With more than two subgroups for this moderator we do not present estimates (or effect sizes) of the impact for the reference subgroup or the additional impact associated with belonging to a given subgroup in relation to the reference subgroup. These are provided in table AL1 in appendix AL.

Note: Number in parentheses is standard error. For a given moderator, the estimate of differential impact was based on the full sample (i.e., all subgroups combined) and using a model with an interaction between the moderating variable and the indicator of treatment status. Subgroup impacts were estimated using mutually exclusive subsamples corresponding to each level of the moderator and using a model like the benchmark model for obtaining the regression-adjusted estimate of average impact. Because the differential and subgroup impacts are estimated using different samples and models, we do not observe certain properties in the results that one might expect had we obtained all estimates using just the model that includes the interaction between the moderator and treatment status. For example, the difference in the point estimates for the subgroup impacts does not correspond exactly to the estimate of differential impact based on the model with the interaction.

Also, in some cases the average impact based on the whole analytic sample lies outside the range of the subgroup impacts. The whole analytic sample is not identical to the union of the subsamples used for a given moderator and corresponding subgroup analyses, because the latter set excludes students with a missing value for the moderator variable. This and the use of regression adjustments in estimating the average effects result in the average impact based on the whole analytic sample lying outside the range of the subgroup impacts in some cases.

a. For each of the three main outcomes, mathematics problem solving, science and reading, we use the estimated standard deviation for the control group from the analytic samples used to estimate the average impacts of AMSTI (i.e., from the confirmatory analyses for impacts on mathematics and science, and the corresponding exploratory analysis in reading) as the denominator in the standardized effect size estimate. We do this in order that all estimates for a given scale are expressed in terms of the same standard deviation units to facilitate comparison of results.

b. We divided the pretest into three categories low, for scores that belong to stanines 1-3 in their respective grade levels; middle, for scores that belong to stanines 4-6 in their respective grade levels; high for scores that belong to stanines 7-9 in their respective grade levels. The cutpoints for the stanines were based on the pretest scale scores for the sample. As explained in appendix B, the study's Technical Working Group advisors recommended that we examine whether the effect of AMSTI on student achievement in mathematics varies depending on students' pretest scores on the SAT 10 reading exam. In the absence of a science pretest, we examined whether the effect of AMSTI on student achievement in science varies depending on students' pretest scores on the SAT 10 reading exam. Therefore, we include analyses of the moderating effects of the reading pretest on the impacts of AMSTI on mathematics problem solving and science.

c. The *p*-value is for the type-3 test of fixed effects for the interaction effect.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

7: Summary of findings and study limitations

After a brief recap of the study design, this chapter summarizes the confirmatory and exploratory findings on the effect of AMSTI on the achievement of upper-elementary and middle school students and on classroom practices hypothesized to improve students' achievement. It also summarizes the effect of AMSTI on teacher content knowledge and student engagement and variations in effects on student achievement by specific subgroups after one year. The chapter concludes by identifying the study's strengths and limitations.

Study design

This study is the first randomized controlled trial testing the effectiveness of AMSTI in improving mathematics problem solving and science achievement in upper-elementary and middle schools. AMSTI is an initiative specific to Alabama and was developed and supported through state resources.

In the cluster randomized trial, schools were randomized within matched pairs in which one school was randomly assigned to participate in AMSTI starting the first year and the second school was assigned to a control group the first year and to participate in AMSTI the second year. In all, 82 schools, 780 teachers, and 30,000 students participated in the study.[96] The study's internal validity is based on a randomization procedure and is strengthened by the low rate (less than 5 percent) of attrition at all levels over the follow-up period.

The statistically unbiased estimates of the effect of AMSTI were generated under authentic conditions for this program as implemented under ordinary conditions in volunteer schools in Alabama. The study did not alter implementation specifically for the experiment but followed schools as they participated in the standard initiative.

Summary of confirmatory findings

Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on mathematics problem solving and science achievement after one year

An important finding is the positive and statistically significant effect of AMSTI on mathematics achievement as measured by the SAT 10 mathematics problem solving assessment administered by the state to students in grades 4–8. After one year in the program, student mathematics scores were higher than those of a control group that did not receive AMSTI by 0.05 standard deviation, equivalent to 2 percentile points. (If the 50th percentile control student had been placed in an AMSTI school, the student would have scored in the 52nd percentile.)

---

[96] These numbers represent the approximate number of unique teachers and students in the 82 study schools in Years 1 and 2 of Subexperiment 1 and Subexperiment 2. For precise numbers of teachers and students included in each analysis, see chapter 2.

Nine of the 10 sensitivity analyses yielded effect estimates that were statistically significant at the .025 level, consistent with the main finding.

The effect is smaller than expected. Whether the statistically significant effect is important for education is open to interpretation. It might, however, be useful to convert the effect into the more policy-relevant metric of additional student progress measured in days of instruction. In these terms, the average effect of AMSTI can be translated into an estimated 28 days of additional student progress over students receiving conventional mathematics instruction. This value was obtained by dividing the estimate of the effect by the mean pretest to posttest difference on the SAT 10 mathematics problem solving assessment for the control group and assuming a 180-day school year.

The estimated effect of AMSTI on science achievement measured after one year was not statistically significant. Based on the SAT 10 science test administered by the state to students in grades 5 and 7, no difference between AMSTI and control schools could be discerned after one year.

Effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on classroom instructional practices after one year

Changes in classroom instructional strategies, especially an emphasis on more active-learning strategies, are important to the AMSTI theory of action. Therefore, a secondary investigation of classroom practices was conducted, based on data from survey responses from teachers. For both mathematics and science, statistically significant differences were found between AMSTI and control teachers in the average reported time spent using the strategies. The effect of AMSTI on these instructional strategies was 0.47 standard deviation in mathematics and 0.32 standard deviation in science.

Summary of exploratory findings

Effect on achievement in mathematics problem solving and science after two years

Estimating the two-year effect is complicated by the fact that the control group received AMSTI in the second year. To estimate this effect, a method that took advantage of the experimental structure of the data but required additional assumptions was used. Due to the uncertainty introduced by the assumptions required by the analysis, the findings are considered exploratory and thus only meant to *suggest* that a two-year effect may be present for both mathematics and science and warrant further research on the longer-term effect of AMSTI.

Two years of AMSTI appeared to have a positive and statistically significant effect on achievement in mathematics problem solving, compared to no AMSTI. The two-year effect estimate represents a difference of 0.10 standard deviation, equivalent to a gain of 4 percentile points for the average control school student if the student had received AMSTI for two years. This estimate can be translated into an estimated 50 days of additional student progress at the rate of advancement experienced by students receiving conventional mathematics instruction.

Two years of AMSTI appeared to have a positive and statistically significant effect on achievement in science. The effect estimate represents a difference of 0.13 standard deviation and is equivalent to a gain of 5 percentile points for the average control school student.[97] Because the SAT 10 science test is required in Alabama only for grades 5 and 7, the science pretest scores needed to calculate the translation of the two-year effect of AMSTI on science achievement into the number of conventional instructional days were not available.

Effect on reading achievement after one year

AMSTI appeared to have a positive and statistically significant effect on reading achievement as measured by the SAT 10 test of reading administered by the state to students in grades 4–8. Reading scores of AMSTI students exceeded those of an equivalent control group that did not receive AMSTI by 0.06 standard deviation. This improvement is equivalent to 2 percentile points and can be translated into an estimated 40 days of additional student progress over students receiving conventional reading instruction.

Effect on teacher-reported content knowledge and student engagement after one year

AMSTI did not appear to have a statistically significant effect on teacher-reported content knowledge in mathematics or science after one year. AMSTI did have a positive and statistically significant effect on student engagement after one year, measured on a 5-point scale ranging from "not engaged" to "fully engaged." AMSTI mathematics and science teachers were more likely than control teachers to rate their students as achieving higher levels of engagement; the regression-adjusted estimates of the average difference between AMSTI and control groups in the cumulative odds of response were 1.76 ($p = .024$) and 3.32 ($p = .003$) for mathematics and science, respectively.

Effect on different subgroups after one year

AMSTI did not appear to have statistically significant differential effects on student achievement in mathematics problem solving or science based on racial/ethnic minority status, enrollment in the free or reduced-price lunch program, gender, or pretest level. In reading, AMSTI did appear to have a statistically significant differential effect for minority and White students. This difference in estimated impact of 3.04 scale score points ($p < .001$) between the two groups can be translated into the metric of days of student progress, where progress is

[97] The analysis of the two-year impact of AMSTI on student achievement is exploratory. Readers should exercise caution in interpreting the results. For instance, we remind the reader that with exploratory analyses multiplicity adjustments are not performed. As a consequence, a less strict criterion is used with exploratory analyses for deciding whether a particular result achieves statistical significance, with the drawback that it increases the probability of finding a spurious impact. For the two-year impact on mathematics problem solving ($p = .030$) and science ($p = .038$) the results reach statistical significance under the less strict criterion (alpha = .05). Under the more strict criterion used with the primary confirmatory analyses (alpha = .025) these results would not have been considered statistically significant.

measured by the average gain in test scores over the course of the school year by the control group using conventional reading instruction. In this metric, White students in AMSTI made an estimated 52 more days of progress than minority students in AMSTI. The effect of AMSTI on reading achievement for minority students was not statistically significant (p = .294); for White students, AMSTI had a positive and statistically significant effect on reading achievement (p < .001).

Study limitations

Although this study employed a rigorous design, there are limitations to the generalizability of its findings, for four main reasons:

- The results apply only to schools that volunteered to participate after a commitment by the principal and staff. The results would not necessarily hold if, for example, AMSTI were adopted by the state as a required instructional program for all schools.
- The effects of AMSTI were contrasted with the conventional program of instruction in Alabama schools. Implementation in other states would face a different counterfactual.
- Although AMSTI uses active-learning strategies and has much in common with other mathematics and science programs influenced by the same principles, its implementation and support systems have many unique characteristics. The results would not necessarily apply to similar programs.
- The long-term effects of AMSTI are not established by this study. A randomized experiment inherently has to begin with schools that are new to the intervention. The study of longer-term effects would require continued tracking of the sample of schools, or, alternatively, an observational study of schools that joined the program many years before the beginning of this experiment in fall 2006.

Appendix A. Explanation of primary and secondary confirmatory outcome measures

For the primary confirmatory outcome of student achievement, the state of Alabama provided six academic measures. Because AMSTI is aimed primarily at mathematics and science, it made sense to examine results in those areas. Because only one measure of science achievement (the SAT 10) was specified, that measure was included in the confirmatory analysis.

There were three measures of mathematics achievement. The SAT 10 problem-solving scale was considered to be more closely aligned to the AMSTI theory of action than the other two measures. In contrast to the SAT 10 procedures scale or the Alabama Reading and Mathematics Test, which have broader scope, problem solving is more likely to be affected by an initiative that emphasizes hands-on, inquiry-based activities relevant to real-life experience and supports students in developing higher-order thinking skills. In addition, the AMSTI coordinator confirmed that the problem solving scale was more closely aligned with the AMSTI model than the procedures scale or the state mathematics test.

Although it was believed that the program would also have an effect on reading, this outcome was considered a potentially beneficial side effect rather than a main goal. For this reason, the two measures of reading from the state's six academic measures were not included among the confirmatory outcomes.

For the secondary confirmatory analysis, a single outcome variable was constructed based on the instructional practices emphasized by AMSTI. Of the classroom practices measured in the surveys conducted for this study, three were most closely aligned to the approach favored by AMSTI: hands-on activities, inquiry-based activities, and practices promoting higher-order thinking skills. Absent instruction that focused on activity- and inquiry-based approaches, in theory the effect of AMSTI on students would not be realized. Therefore, it was critical to examine whether AMSTI affected this specific intermediate outcome. This analysis is sufficiently important that it was considered confirmatory. A composite variable, the active learning scale, was constructed to analyze this outcome (see chapter 2 for details.)

Appendix B. Explanation of exploratory research questions

This appendix discusses the exploratory research questions, providing a rationale for each.

Effects of two years of exposure to the Alabama Math, Science, and Technology Initiative (AMSTI)

The effect of two years of exposure to AMSTI cannot be estimated without additional assumptions, because the control group entered the AMSTI program after a year in the study and was thus no longer a pure control group in the second year. Two years after random assignment, control schools had participated in AMSTI for one year and intervention schools had participated in AMSTI for two years. Given this problem, the effect of AMSTI after two years was examined as part of the exploratory analysis (see chapter 2 for a description of the methodology and chapter 5 for results).

Effect on student achievement in reading

For SAT 10 reading, the study investigated whether participation in AMSTI mathematics and science instruction had an effect (favorable or unfavorable) on reading achievement after one year. This question was examined for three reasons.

First, according to the developers, “AMSTI had purposefully incorporated various reading and writing practices into its modules” (AMSTI n.d.) based on the mathematics standards from the National Council of Teachers of Mathematics and the science standards promoted by the International Society for Technology in Education. These standards were incorporated into curriculum, assessment, and professional development. They include learning practices such as incorporating reading strategies, connecting to literature, and using writing to represent thought processes or justify conclusions. The AMSTI writing requirements include the keeping of science notebooks and math journals (T. Beers, AMSTI math coordinator, e-mail July 16, 2010).

Second, results from previous quasi-experimental evaluations of AMSTI found effects on reading and writing that researchers “attributed” to AMSTI (Miron and Maxwell 2007). According to documentation from the director of AMSTI referencing these external evaluation results, these evaluations concluded that reading and writing scores “were found to be considerably higher in AMSTI schools as compared with scores from a control group of non-AMSTI schools with similar demographics. Statistical significance was found in many cases. Such findings confirm that AMSTI has successfully included strategies for addressing reading and writing, as the students learn math and science using hands-on activities” (Ricks 2008, cover letter). In the most recent evaluation report, based on 2007 standardized test data from the 195 schools that adopted AMSTI since 2002 and the 576 schools from the same school systems that served as controls, Miron and Maxwell (2007) reported statistically significant findings in reading and writing that varied by grade level and content area. For the SAT 10, advantages for AMSTI schools in reading reached statistical significance for grades 4 and 5 but not grades 6–8.[98] For the Alabama Direct Assessment of Writing (ADAW), given in study grades 5 and 7, none of the findings reached statistical significance. The quasi-experimental research on which this conclusion is based leaves room for alternative interpretations, such as selection bias in the process by which some schools joined AMSTI and others entered the control group. Replicating the findings reported in these evaluations using a stronger design is therefore important.[99]

Third, the study's technical working group advisors recommended that research correlating reading achievement with mathematics and science outcomes be reviewed. Data from 2005 American College Testing (ACT) Program suggest that "the clearest differentiator in reading between students who are college ready and students who are not is the ability to comprehend *complex* texts" (American College Testing 2006, p. 2). Based on these results, ACT suggests that policymakers and educators "strengthen reading instruction in *all* high school courses by incorporating complex reading materials into course content." Courses in mathematics and science can challenge students to read and understand complex texts, providing opportunity for students to improve their foundational reading skills and strategies. AMSTI in particular provides opportunities for students to read and understand complex texts. It is therefore of great interest to understand whether these opportunities translate into improvement in reading achievement, thereby potentially facilitating college readiness.

Effect on teacher content knowledge

The intermediate effect of AMSTI on teacher content knowledge was measured after one year. The primary focus of AMSTI professional development is to change teacher content knowledge (University of Alabama). A body of research (Garet, Porter, Desimone, Birman, and Yoon 2001; Mullens, Murnane, and Willett 1996; Shulman 1987) supports the notion that teacher content knowledge has a positive effect on changes in classroom practices among teachers of mathematics and science. Researchers have also found that mathematics and science teachers' content knowledge is related to improvements in student academic achievement (Goldhaber and Brewer 1996; Hill, Rowan, and Ball 2005; Kennedy 2008; Mullens, Murnane, and Willett 1996). Therefore, on the fourth teacher survey, both AMSTI and control teachers were asked to rate their content knowledge for teaching mathematics or science at the grade level

---

[98] Findings showing the positive impact of AMSTI on student achievement reached statistical significance on the SAT 10 for grade 5 mathematics and science as well. No findings for grades 6–8 reached statistical significance.

[99] As with many quasi-experimental studies, the results may have been subject to substantial bias, particularly because schools participating in AMSTI have to apply to the program and may already be more motivated to improve their mathematics and science programs than schools that do not apply. In addition, none of the evaluations examined whether there was baseline equivalence on possible confounders, including pretest measures, or adjusted for potential bias caused by nonequivalence. These weaknesses raise concerns about the internal validity of study findings, specifically whether the observed increase in AMSTI students' achievement can be explained only by their participation in the program and not by other plausible explanations, such as prior achievement.

they currently taught.[100] Teachers responded on a fixed Likert scale (very low, low, moderate, high, very high), with a sixth option of "not applicable."

Effect on student engagement

The study also examined the effect of AMSTI on student engagement after one year. AMSTI's emphasis on student active learning in the classroom is designed to motivate and engage students in learning (AMSTI Committee 2000). Moreover, "several studies have demonstrated a positive correlation between behavioral engagement and achievement-related outcomes" (Fredricks, Blumenfeld, and Paris 2004, p. 70). In addition, Singh, Granville, and Dika (2002), using structural equation models on observational data, found that student engagement is a predictor of mathematics and science achievement. The fourth teacher survey asked both AMSTI and control teachers to rate the average level of student engagement in their classes during the school year.[101]

Effects on subgroups of students

Pretest

The study examined whether the effects of AMSTI varied with students' pretest scores. For the SAT 10 mathematics problem solving outcome, the study examined whether students with higher mathematics pretest scores experienced larger or smaller AMSTI effects on the mathematics outcome. The study's technical working group advisors recommended that the study also examine whether the effect of AMSTI on student achievement in mathematics varied depending on students' pretest scores on the SAT 10 reading exam. The study therefore investigated this potential moderating effect as well.

For the SAT 10 science outcome, researchers examined the moderating effect of SAT 10 reading pretest scores on the impact of AMSTI. Because the SAT 10 science test is required only for students in grades 5 and 7 in Alabama, none of the students with science outcome measures had science pretest scores from the previous year. In the absence of a science pretest results, the SAT 10 reading pretest was used as a pretest covariate in the model for the primary confirmatory analysis of the effect of AMSTI on student performance in science; the moderating effects of this covariate were also examined.

For the SAT 10 reading outcome, the study measured the moderating effect of the SAT 10 reading pretest on the effect of AMSTI.

---

[100] Survey questions were asked separately for mathematics and science teachers.

[101] Survey questions were asked separately for mathematics and science teachers. Survey instructions indicated that students would be considered fully engaged if they not only paid full attention but also participated fully and completed all assignments. Teachers responded on a five-point Likert scale (not engaged, slightly engaged, moderately engaged, almost fully engaged, fully engaged).

Free or reduced-price lunch status, racial/ethnic minority status, and gender

The study examined potential differential effects of AMSTI for students with different characteristics. It compared students based on racial/ethnic status,[102] socioeconomic status, and gender. Although evidence on subgroup differences for interventions that are similar to AMSTI is limited, there are at least three reasons to examine effects for these subgroups. First, No Child Left Behind requires that states disaggregate student achievement data for these subgroups and that interventions be adopted to close achievement gaps. Below are data on the achievement gaps in Alabama, based on student gender, ethnicity, and free or reduced-price lunch (FRL) status, as measured by the SAT 10 assessment during the 2005/06 school year (table B1). Second, AMSTI aims to improve student achievement for all students. By examining if AMSTI has differential effects on different subgroups of students, the study checked whether AMSTI was meeting its goal of providing equitable opportunities to learn. Third, the results from these analyses will inform the developers about where to focus program improvements to strengthen the intervention for the groups receiving less benefit.

Table B1 Achievement gaps on the Stanford Achievement Test Tenth Edition (SAT 10) in Alabama, 2005/06

Group	Percent tested[a]	Percentile[b]	Percent in group[c]
Grade 8 reading			
Black	96.42	32	36.44
White	98.02	58	59.24
Free lunch	96.11	33	42.29
Reduced-price lunch	97.73	44	9.13
Boys	97.00	44	51.18
Girls	97.68	52	48.82
Grade 8 mathematics			
Black	96.23	35	36.37
White	97.93	59	59.19
Free lunch	96.16	36	42.32
Reduced-price lunch	97.75	46	9.13
Boys	96.97	49	51.17
Girls	97.69	53	48.83
Grade 7 science			
Black	95.75	39	37.14
White	97.06	63	58.54
Free lunch	95.09	40	44.34
Reduced-price lunch	97.29	53	9.20
Boys	95.93	53	51.71
Girls	97.04	55	48.29

a. Percentage of students enrolled in each group that took this test.
b. Relative standing of group (national average is 50).
c. Percent of tested students group represents.
Source: ALSDE Accountability Reporting System (http://www.alsde.edu/Accountability/Accountability.asp).

---

[102] The majority of students in the study were either Black (39 percent) or White (57 percent) Percentages are from the analytic sample associated with the SAT 10 mathematics problem solving outcome after one year.

Appendix C. Selection and random assignment of schools

Selection of schools for randomization was conducted by AMSTI staff and the study's researchers, in order to achieve the buy-in of AMSTI staff and to capitalize on their thorough understanding of the characteristics of the regions and schools involved. The process was also designed to meet the resource constraints of ALSDE and to meet the design specifications called for by the study's statistical power analysis.

Subexperiments

The number of AMSTI schools that could be supported in a given year was based on the program's operating budget. It was not known in advance, because it was dependent on the number of mathematics and science teachers in each selected school. Because the annual AMSTI budget could not support introducing all of the schools called for by the power analysis in a single school year, school participation had to be staggered, requiring two subexperiments.

Limiting the pool of schools for selection

The process of selecting schools for randomization for each subexperiment took place at a meeting of the AMSTI staff and directors of the regional sites, to which the research team was invited. The primary purpose of the meeting was to select the schools to be admitted to the program from among those that had applied. ALSDE's process made use of spreadsheets designed by the AMSTI staff that calculated the incremental cost to the program of adding each school (depending on size and grade level) as it was selected to participate in the study. Having agreed to participate in the study, the AMSTI staff gave the researchers some leeway in selecting which schools would participate, allowing for a purposive selection of schools from the pool of applicants. However, the process of selecting had to take place within the time constraints of the meeting. The selection could not take place in advance, because, although the list of applicants was available, the AMSTI staff still had to eliminate certain schools from consideration, either because they were deemed ineligible or because AMSTI staff had already promised the school that it could participate in AMSTI that year.

Selecting and assigning schools

The approach to selection and assignment was based on a process that took into consideration a budget that allowed a limited number of schools to be selected, the goal of purposively selecting a sample of schools that was representative of the regions involved, and the goal of selecting pairs of similar schools in order potentially to raise the precision of the effect estimate by conducting randomization within pairs.

The randomization scheme was constrained to be blocked by region, as it was the AMSTI procedure that each region should have approximately equal numbers of AMSTI schools. The AMSTI staff also needed to ensure that there was a reasonable proportion of elementary and middle schools, as well as large and small schools. The research team decided as a matter of design to use the available student achievement data (primarily percentage proficient in mathematics) and demographic data (primarily percentage racial/ethnic minority and percentage enrolled in the National School Lunch Program)[103] to identify pairs of schools that were similar in these and other characteristics (such as school size) and ensure that the average achievement and demographic characteristics were similar to the average of the region generally. Paired randomization was used to increase the precision of the effect estimate and because it had the benefit of conveying to the AMSTI staff the rationale for randomization, especially the idea of assigning to the two conditions schools in each pair that were matched in various respects. The research team also found that paired randomization using a coin toss once the pairs were identified was readily understood by educators, improving cooperation in the experiment. The goal of purposefully matching the sample to the regional characteristics was to support the validity of the sample as representing the region.

The sample by region is described below (table C1). It also shows the number of schools in the initial pool of applications from each region and the number selected for the study.

Table C1 Primary matching characteristics of sites, by region

		Number of applicant schools eligible to receive AMSTI and participate in the study	Percentage of students that were		
			Math proficient	**Racial/ethnic minority**	**Enrolled in school lunch program**
Subexperiment 1					
Region 1	Region	25	66	36	60
	Selected	14	67	37	65
Region 2	Region	26	67	53	62
	Selected	14	71	56	64
Region 3	Region	23	72	32	49
	Selected	12	71	32	41
Subexperiment 2					
Region 4	Region	31	70	22	49
	Selected	20	74	21	53
Region 5	Region	39	57	74	67
	Selected	22	55	70	77

[103] Researchers used public sources to determine the achievement levels and demographics of all applicants and came to the meeting with a spreadsheet to record the selections and results of the randomization. They explained the randomization process to the meeting participants, as well as the ground rules, the most important of which was that once an assignment was made, it could not be changed.

The real-time and interactive nature of pair selection

The selection of pairs was itself an interactive process constrained by the criterion stated above, but it was neither formally algorithmic nor fully controlled by the research team. Because the pairing decisions were made in real time during the meeting, it was not possible to apply a formal algorithm to form the pairs.

Applicants were displayed on a spreadsheet projected for the participants to view. As pairs were proposed and agreed on, an Alabama state official tossed a coin and the assignment of the school to the AMSTI or control condition was recorded. Although informal, the process was deliberate and followed a sequence of considerations, beginning with similarity of grade configuration (related to the split between elementary and middle schools) and then considering mathematics scores, percent of minority students, and (when possible) percent of students enrolled in the National School Lunch Program. AMSTI regional directors also provided input on the appropriateness of the pairings, based in many cases on their local knowledge of similarities that went beyond those captured in the formal criteria; in some cases, they corrected or updated the data.

As pairs were selected, a spreadsheet was used to perform a running calculation to confirm that the demographics of the combined pairs in each region were similar to those of the region's schools. In some cases, this consideration was the deciding factor in choosing between two matched pairs.

Appendix D. Statistical power analysis

At the planning stage of the randomized trial, a statistical power analysis was conducted to determine the number of schools required to detect an effect of AMSTI on student performance equal to 0.20 standard deviation of the distribution of posttest scores. This level is consistent with that set in other evaluations sponsored by the National Center for Education Evaluation and Regional Assistance (Garet et al. 2010).[104] In addition to a minimum detectable effect size of 0.20, the statistical power analysis was based on the following assumptions:

- Two-level hierarchical design, with students at Level 1 and schools randomized at Level 2.
- Statistical power of .80.
- Statistical significance level of .05 for a two-tailed test.
- Effect of AMSTI modeled as a fixed effect at Level 2.
- Two hundred and eighty students per school (8 teachers per school times 35 students per teacher).[105]
- Pretest/posttest correlation between school-level scores of .80 (R^2 of .64).[106]

---

[104] Because of lack of reliable information on how much additional precision was obtained from using a matched-pairs strategy, the sample size calculation did not depend on the use of that strategy—that is, in choosing the sample size, researchers assumed no benefits of pairing. If the pairing strategy accounted for 50 percent of the variation in the outcome after modeling the pretest, the minimum detectable effect size would be reduced to 0.15. Because some benefit was expected of pairing, the best estimate is that the experiment was powered to detect a minimum detectable effect size of 0.15–0.20. The effect on precision of modeling additional covariates was not figured in. The decision to include covariates was made after the experiment had started, as part of the strategy to handle missing values (the dummy variable approach to handling missing values is discussed in the section on data analysis). We assumed that modeling covariates would probably further increase precision. Because the benefits of using a matched-pairs design and modeling additional covariates were not figured into the power analysis, the results of the analysis can be considered conservative.

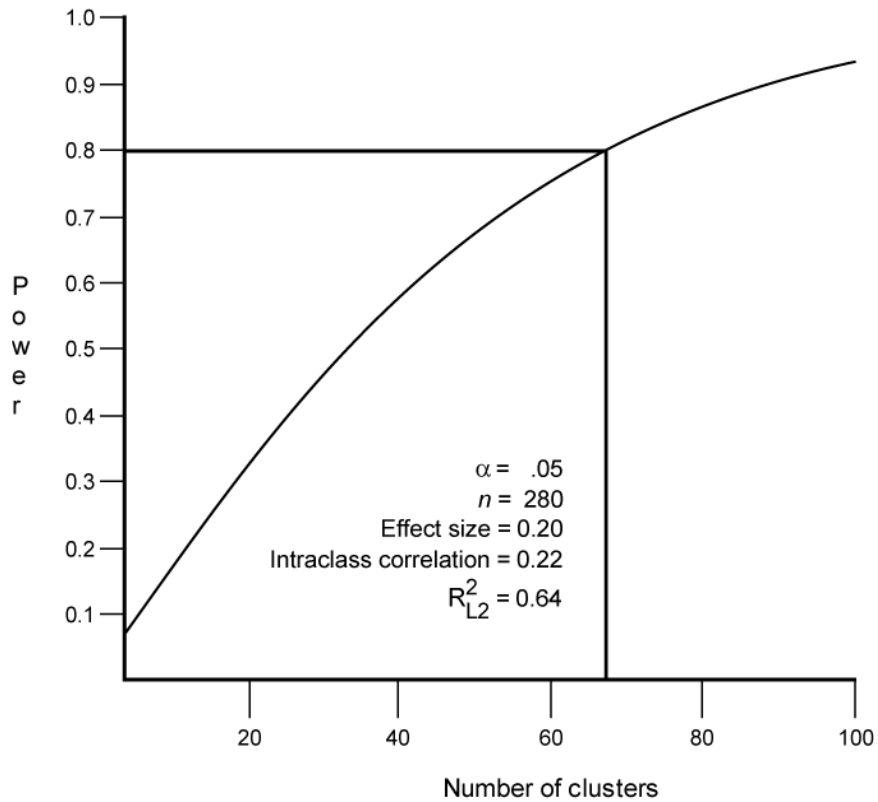
[105] This figure is the estimated number of students per school in grades 4–8 for whom a mathematics posttest was available. The analysis of science outcomes included students from grades 5 and 7 only. Reducing the student sample to 112 (that is, two-fifths of 280) had a small effect on the minimum detectable effect size, because schools were the unit of randomization. The minimum detectable effect size for the science outcome was 0.22 as a result of this difference.

[106] One can think of two components of R^2, the school-level R^2 (R_c^2) and the student-level R^2 (R_i^2). Use of Optimal Design Software allowed a nonzero value to be modeled for the first parameter but not the second. In selecting a value for R^2 in the analysis of student outcomes, researchers considered only the effect of modeling the school-level pretest. Reliable information was not available for determining either the proportion of student-level variance accounted for by modeling the pretest at the student level or the proportion of variance at either the school or student level explained by covariates other than the pretest (racial/ethnic minority status, free or reduced-price lunch status, English learner status, gender, grade, and pair indicators). Therefore, as a result of modeling these additional covariates, R_c^2 was expected to be greater than assumed (0.64) and R_i^2 to be greater than zero. For teacher outcomes, both R^2components

- Unconditional intraclass correlation coefficient of .22. Hedges and Hedberg (2007) showed that for a heterogeneous sample of schools, the unconditional intraclass correlation for reading outcomes in grades 4–8 ranges fell between .174 and .263. For mathematics outcomes, the intraclass correlation coefficient fell between .185 and .264.
- A school-level attrition rate of 20 percent.
- A minimum detectable effect size of 0.20 standard deviation units.

Optimal Design Software (Raudenbush, Spybrook, Liu, and Congdon 2006) was used to perform the power analysis. Below, the statistical power as a function of the number of schools is displayed (figure D1).

Figure D1 Record of technical settings and result of power analysis



---

were assumed to be zero. However, the covariates modeled—degree rank, years of teaching experience, years of teaching relevant subject and pair indicators—were expected to account for some of the variance in the outcome at the teacher and school levels. Appendix X compares assumed and observed values of the R^2 parameters.

Sixty-six schools were required to detect an effect size of 0.20 with .80 power for student-level outcomes. Assuming 20 percent attrition, the required sample was 82 schools. The benefits to precision of using a matched-pairs design or of modeling covariates were not figured in; the results of the analysis can therefore be considered conservative.

For the power analysis addressing the effect of AMSTI on classroom practices, the minimum detectable effect size was calculated assuming 82 schools (66 schools with attrition figured in) and 8 mathematics/science teachers per school. This analysis assumed an intraclass correlation coefficient of .20, a two-level design, statistical power of .80, and a Type I error rate of 5 percent. The benefits to precision of using a matched-pairs design or of modeling covariates were not figured in; the results of the analysis can therefore be considered conservative.

A preintervention measure of the outcome variable was not available; it was therefore not possible to include a “pretest” in the analysis. Eighty-two schools allowed an effect size of 0.34 for teacher outcomes to be detected with .80 power. Assuming 20 percent attrition of schools increases the minimum detectable effect size to 0.38. Although this effect size is larger than the one designed to detect the effect on students, it is appropriate for examining effects on classroom instructional practices, because the AMSTI theory of action stipulates that the effects of AMSTI on student outcomes are mediated through an initial, more immediate effect on classroom practices. Greater effects are therefore expected on classroom practices than on student performance.

Appendix E. Data collection procedures and timeline

This appendix describes the data collection procedures and the timeline of the study. As previously mentioned, Year 1 signifies the first year of AMSTI implementation for the schools randomized to the AMSTI group in each subexperiment (2006/07 for Subexperiment 1, 2007/08 for Subexperiment 2). Year 2 signifies the second year of AMSTI implementation for the schools randomized to the AMSTI group, and first year of AMSTI implementation for the schools randomized to the control group, in each subexperiment (2007/08 for the Subexperiment 1, 2008/09 for Subexperiment 2).

Table E1 Completion dates for data collection procedures

Task	Subexperiment 1	Subexperiment 2
Researchers collected written permission from districts for student rosters	February 2006	May 2007
Researchers received class rosters, identified teacher sample, and populated database with district, school, teacher, and student data	December 2006–February 2007 for Year 1	November 2007–January 2008 for Year 1
	November 2007–January 2008 for Year 2	November 2008–January 2009 for Year 2
Researchers conducted teacher surveys (January–April) and received results (January–May)	January 2007–May 2007 for Year 1	January 2008–May 2008 for Year 1
	January 2008–May 2008 for Year 2	January 2009–May 2009 for Year 2
Researchers identified students of selected mathematics and science teachers and recorded student names and state-assigned identification numbers	February 2007 for Year 1	May 2008 for Year 1
	February 2008 for Year 2	December 2008 for Year 2
Alabama State Department of Education verified student records and sent student achievement data and additional demographics to researchers	January 2008 for Year 1	January 2009 for Year 1
	January 2009 for Year 2	January 2010 for Year 2
Researchers populated database with student achievement data and additional demographics	January 2008 for Year 1	January 2009 for Year 1
	January 2009 for Year 2	January 2010 for Year 2
Researchers deleted student name data and replaced with export identification numbers	January 2008 for Year 1	January 2009 for Year 1
	January 2009 for Year 2	January 2010 for Year 2
Researchers populated database with teacher survey data	June 2009 for Year 1	June 2009 for Year 1
	June 2009 for Year 2	January 2010 for Year 2
Shared identified data files with researchers for analysis	January 2008 (student-level analysis) for Year 1	February 2009 (student-level analysis) for Year 1
	June 2009 (classroom-level analysis) for Year 1	June 2009 (classroom-level analysis) for Year 1
	November 2008 (student-level analysis) for Year 2	February 2010 (student-level analysis) for Year 2

Appendix F. Description of program implementation data collected but not used in report

This appendix reports the data obtained from measures that were collected as part of the overall study, but that were not used as part of this specific study or report. These measures include: classroom observations, professional development teacher surveys, professional development observations, and principal surveys.

Classroom observations

Classroom observations were collected in order to obtain descriptive information on mathematics or science instruction in teachers' classrooms, document students' participation in instructional activities, and determine the extent to which teachers' instructional strategies were related to the AMSTI model and materials. Classroom observation data were originally collected to inform additional research questions about program implementation that are not addressed in this report.

Observations were conducted at AMSTI and control schools.[107] Forty teachers in 21 AMSTI schools and 41 teachers in 20 control schools were observed in Year 1 of Subexperiment 2 (the 2007/08 school year).[108] For each classroom observation, two researchers took notes during the observation period, after which both completed the classroom observation protocol. The observers discussed their ratings and resolved any differences in order to complete one protocol with consensus ratings. Interrater reliability before consensus ratings ranged from 76.2 percent for classroom context to 83.6 percent for learning objectives emphasized. Subsequently, one researcher entered the consensus ratings online for data analysis. Researchers were trained by the Academy for Educational Development (AED) on the observational protocol during a two-day training in the summer prior to implementation. During this training, researchers watched videos of mathematics and science instruction, individually rated teachers and classrooms on relevant constructs, and then discussed ratings in order to reach consensus.

The classroom observation form was adapted from the Authentic Instructional Practices Classroom Observation form (Borman et al. 2000) and the Reformed Teaching Observation Protocol (Piburn et al. 2000). The protocol was highly prescriptive, prompting researchers to provide ratings on three main constructs or scales of interest: learning objectives emphasized, use of "authentic instruction" principles,[109] and classroom context.

---

[107] Teachers were selected for classroom observation and teacher interviews through a joint selection process. See section in chapter 2 on teacher interviews for description of the selection process.
[108] As previously explained, observation data were collected only for Subexperiment 2 because researchers did not receive approval from the Office of Management and Budget to administer this implementation measure in Subexperiment 1.
[109] Authentic instruction, as defined by Newmann and Wehlage (1993), refers to a teacher's use of innovative and challenging instruction that provides students with the opportunity to use higher-order thinking skills, explore topics in depth, connect lesson material to their own lives, engage in substantive

For learning objectives emphasized, observers rated classrooms on the degree of emphasis the teacher placed on six learning objectives—knowledge, comprehension, application, analysis, synthesis, and evaluation—using a three-point Likert scale (not present at all, present to some extent, and present to a great extent).

For use of authentic instruction principles, observers rated classrooms on the degree to which teachers used each of the six instructional principles, using a Likert scale reflecting each construct:

- *Coherence of the material*: 3-point scale (presenting material in superficial fragments, covering some overarching concepts, covering overarching concepts in depth).
- *Connection with students' out-of-school experiences*: 3-point scale (no clear connections, incidental connections, students are encouraged to see the connection).
- *Connection to other content disciplines*: 3-point scale (no clear connections, superficial connections, conceptual-level connections).
- *Substantive conversation*: 4-point scale (no substantive conversation, occasional teacher probes, frequent teacher probes, students encouraged to converse among themselves).
- *Teacher support for students*: 3-point scale (not good, mixed, positive).
- *Student engagement*: 4-point scale (students inattentive, students occasionally on-task, students on-task most of the time, all but one or two students deeply engaged).

Observers rated teacher use of technology, student use of technology, teacher use of manipulatives, and student use of manipulatives on a 3-point Likert scale (none used, used procedurally, used to enhance understanding). They rated the extent to which different student grouping patterns (individuals, pairs, small groups, whole class) were used in classrooms on a 5-point Likert scale (none of the time, 1–25 percent of the time, 26–50 percent of the time, 51–75 percent of the time, 76 percent or more of the time). They rated teacher comfort level and teacher knowledge of subject matter on a 4-point Likert scale (not at all, a little, fairly, very). They also recorded descriptive information about the class, such as the setting and seating arrangements, grade level, and estimated ethnicity and gender composition of the class.

Classroom observations were conducted in both control and AMSTI classrooms during Years 1 and 2 of the study. Data from classroom observations are not included in this report because they do not inform the specific aims or research questions specified for it.

---

conversation, have social support in the classroom that supports high levels of achievement, and be engaged in the lesson.

Professional development teacher surveys

Surveys were administered to teachers at the beginning and end of summer institute trainings in order to gauge how well trainings increased their knowledge and skills in content areas as well as to gain feedback on the trainings. The surveys were originally collected to inform additional research questions about program implementation that are not addressed in this report.

Teachers rated their knowledge and skills in their grade and subject area at the beginning of the summer professional development trainings and at the conclusion of the training. Because each grade/subject covered distinct content during the training, eight separate surveys were developed, to enable training recipients to rate their knowledge and skills in subject areas specific to their grade/subject and Year 1 or Year 2 training status. The surveys administered at the end of the training also contained questions about teachers' backgrounds, the extent to which teachers considered themselves able to implement AMSTI overall and their perceived preparation regarding their ability to implement specific aspects of the program, anticipated challenges to implementation, the need for follow-up support and the type of support needed, and feedback about specific aspects of the training, including quality and length of the training, aspects of the training teachers enjoyed, and suggestions for improvement.

Trainers distributed professional development teacher surveys at the beginning and end of the training to all attendees who were grade 5 and 7 mathematics and science teachers. All surveys were anonymous, in order to encourage candid responses. Trainers collected the surveys, sealed them in envelopes, and gave them to the training site coordinators, who mailed them to AED.

Professional development surveys were administered in Years 1 and 2 of the study. Data from professional development teacher surveys were not included in this report because they do not inform the specific aims or research questions specified for it.

Professional development observations

Professional development observations were conducted in order to obtain information on the summer institute training environment and participant engagement. The observations were originally collected to inform additional research questions about program implementation that are not addressed in this report.

During the training sessions, research staff used a protocol to guide observation of the session. The protocol prompted observers to report, in narrative or open-ended form, the topics covered during that training; the type and extent of use by the trainer of instructional methods (for example, lecture, small group discussion, skills practice); instructional tools used by the trainer; and the degree to which participants appeared engaged in the training session. Observers also identified external factors that could have influenced the training sessions (for example, classroom conditions or a disgruntled participant). Each trainer was observed at least once over the course of training.

Professional development observations were conducted in Years 1 and 2 of the study. Data from professional development observations were not included in this report because they do not inform the specific aims or research questions specified for this report.

Principal surveys

Web-based surveys were administered to principals at the beginning of each school year, in order to gather data on baseline conditions for exploratory analyses to determine whether school conditions moderated the effect of AMSTI on student performance as well as to inform additional research questions about program implementation that are not addressed in this report. The survey instrument was developed by Empirical Education researchers, based on the information needs of ALSDE. The items were adapted from items used in previous Empirical Education studies. The domains included the following:

- Professional development (types, frequency, effect on learning).
- Instructional time, student assessment, technology (availability, support, comfort).
- Teacher background, equipment and materials (availability, use, satisfaction).
- Instructional strategies (inquiry, hands-on, higher-order thinking skills).
- Teacher planning and collaboration.
- Student engagement.
- Meeting an existing need.
- Other school initiatives.
- Community partnerships.

Principals were emailed a consent form, which researchers reviewed with them over the telephone. Once a principal faxed back a signed consent, researchers emailed the principal a survey invitation. Nonrespondents received emails and telephone calls to ensure acceptable response rates.

Surveys were administered in Years 1 and 2 to principals in AMSTI and control schools. Data from one question on the principal surveys administered in Year 1 was used to assess how many schools had prior exposure to Leadership Academy for Math, Science, and Technology (described in chapter 1). Additional data from the principal surveys were not included in this report because they do not inform the specific aims or research questions specified for it.

OMB Number: 1850–0831
Expiration Date: 07/31/2010

Appendix G. Alabama Math, Science, and Technology Initiative (AMSTI) teacher survey #3

The collection of information in this study is authorized by Public Law 107-279 Education Sciences Reform Act of 2002, Title I, Part C, Sec. 151(b) and Sec. 153(a). Participation is voluntary. You may skip questions you do not wish to answer; however, we hope that you will answer as many questions as you can. Your responses are protected from disclosure by federal statute (PL 107-279 Title I, Part C, Sec. 183). All responses that relate to or describe identifiable characteristics of individuals may be used only for statistical purposes and may not be disclosed, or used, in identifiable form for any other purpose, unless otherwise compelled by law. Data will be combined to produce statistical reports. No individual data that links your name, school name, address, telephone number, or identification number with your responses will be included in the statistical reports.

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 1850–0831 (expiration date: 07/31/2010). The time required to complete this information collection is estimated to average 20 minutes, including the time to review instructions, search existing data resources, gather the data needed, and complete the information collection. If you have any comments concerning the accuracy of the time estimate or suggestions for improving this form, please contact: the Department of Education 50 North Ripley Street PO Box 302101 Montgomery, AL 36104. If you have comments or concerns regarding the status of your individual submission, e-mail directly to: XXX at XXX@empiricaleducation.com or call toll free 1-888-486-XXXX ext. XXX.[110]

[110] Researchers' contact information has been removed.

You may want your lesson planner in front of you to answer some of the questions.

Identification:

1. Please enter your first and last name here ____

2. **During the past two weeks,** what curricular and other print materials did you use to teach *mathematics and/or science?* **Mark all that apply.**

__AMSTI supplied: (Please list)

____

____

__A+ Learning Computer Program

__Accelerated Math

__Alabama Course of Study

__Alabama Science in Motion

__Carolina Biological

__CPO Science

__Edutest

__Glencoe

__Harcourt Brace

__Holt Science

__Houghton Mifflin

__Integrated Science

__Lightspan

__Macmillan

__Math for Today

__McGraw-Hill

__Saxon Math

__Scholastic

__Science World

__Scott Foresman Science

__SRA Intervention Math

__Other: (Please list)

____

____

Math

3a. Do you teach mathematics during the current (2008–2009) school year?
__Yes
__No (Go to question 15a)

3b. Do you teach mathematics to students who are not assigned to you on your school's official computerized class roster?

Examples:
-swapping students based on test scores or other factors
-team teaching where you and another teacher teach both your own students and that teacher's students
-supporting another teacher to teach the students in that teacher's classroom
-other

__Yes, please specify ___________________________
__No, I only teach math to students in my own class(es) (Go to question 3f)

3c. Please name the teachers whose students you teach math, or whose students you partner in teaching math, or whom you support in the classroom for math.

Please indicate if you teach ALL the students assigned to this teacher or a smaller group of their students.
____

3d. If you swap math students based on test scores, which test do you use to make that determination? ____

3e. If you swap math students based on test scores, what is the score range of the students you teach?
____

3f. Have you taught the same groups of math students since at least October of this 2008–2009 school year?

__Yes
__No; please explain why not: _____________________________________

3g. Is there anything else you would like us to know about your math classes?
____

Math Instructional Strategies

The following questions are attempting to understand the number of hours that students receive of each type of instruction. Each question asks you to reflect upon the last two weeks (ten full days) of instruction.

4a. **Think back on your last two weeks (10 full days) of instruction**: approximately how many minutes did your students spend doing math in your class? Please enter the total number of minutes. *Be sure to consider all activities, including discussion, lecture, reading, watching video, hands-on activities, worksheets, and activities that integrate math with other subjects.*

Minutes of math instruction ______

4b. The number in question 4a represents my minutes of instruction

__Daily

__Weekly

__For two weeks

4c. How many math classes (that is, different groups of students) do you teach?

__1 (Go to question 4e)

__2 (Go to question 4d)

__3 (Go to question 4d)

__4 (Go to question 4d)

__5 (Go to question 4d)

__6 (Go to question 4d)

__7 (Go to question 4d)

__8 (Go to question 4d)

__Other, please specify __________________(Go to question 4d)

4d. Is the number in question 4a the sum of the minutes for all math classes or the average minutes per class?

__Sum

__Average

__Other, please specify ________________

4e. For the remainder of the math instruction section of this survey, please continue to calculate your responses in the same manner as you did for question 4a.

__OK

4f. Is there anything else you would like us to know about the number of minutes of math instruction that you reported? ____

5. Consider the following description of Inquiry-Based Instruction in which students do ***all*** of the following activities as part of the learning process:

- Make observations
- Pose questions
- Examine books and other sources of information to see what is already known
- Plan investigations
- Review what is already known in light of experimental evidence
- Use tools to *gather*, *analyze*, and *interpret data*
- Propose answers, explanations, and predictions
- Communicate the results

During the past two weeks, approximately how many minutes did students participate in Inquiry-Based Instruction in your math class?
Minutes of Inquiry-Based math instruction ____________

6. **During the past two weeks,** approximately how many minutes did students participate in hands-on math activities (involving active participation; applied, as opposed to theoretical)? Please enter the total number of minutes.
Minutes of hands-on math instruction ____________

7. During the past two weeks, how many minutes were your students engaged in math activities that required higher-order thinking skills (that is, where students advance from skills such as *focusing* and *information gathering* to skills such as *integrating* and *evaluating*)? Please enter the total number of minutes.
Minutes of higher-order thinking skills in math __________

8. During the past two weeks, about how much time did you teach using AMSTI supplied math print materials? Please enter the total number of minutes. If you do not teach AMSTI, please enter "0."
Minutes using AMSTI supplied math print materials _________

9. During the past two weeks, what type of math assessments did you use in your classroom? Please check all that apply.
__Informal assessments, such as questioning and observation, to gauge student learning
__Formative paper and pencil assessments (that is, assessments that occur regularly throughout the year in order to inform instruction)
__Performance-based assessments (that is, assessing students based on their application of knowledge, skills, and work habits through the performance of tasks that are meaningful and engaging to students)
__Standardized assessments
__Other, please describe______________
__I did not administer any math assessments

Math Professional Development

For AMSTI: Please include any professional development you have received as part of the AMSTI program or in any way connected with AMSTI.

For non-AMSTI: Please include all non-AMSTI professional development you have received.

10a. The following questions refer to math Professional Development (PD) activities in which you have participated during the past month. If you have not participated in math professional development activities, please enter "0" for number of hours.

During the past month, how much professional development have you received for your math program? Please do not include support or collaboration meetings. Please enter the total hours of training in each box.

AMSTI mathematics ___________
Non-AMSTI mathematics ___________

10b. To what extent have the math professional development activities increased the following?
1 = Not at all or very little, 2 = To some extent, 3 = A great deal, NA = Not applicable

[Note: On the Web-based survey, these types of questions appear as matrices]

_____Your ability to incorporate technology into your teaching

_____Your ability to use new teaching methods

_____Your ability to teach basic skills and facts

_____Your classroom management strategies

_____Your ability to teach critical thinking skills to your students

_____Your students' academic achievement

_____The way you assess student work

The next questions are about asking for and receiving support. If you did not ask for or receive support, please enter "0" for total times.

11a. During the past month, how many times did you try contacting someone for support (for example, for mentoring or coaching) with math instruction?

AMSTI mathematics total times___________
Non-AMSTI mathematics total times ___________

11b. During the past month, how many times did someone actually provide support (for example, for mentoring or coaching) with math instruction?

AMSTI mathematics total times__________
Non-AMSTI mathematics total times__________

11c. To what extent have the math support activities listed in question 11b increased the following?

1 = Not at all or very little, 2 = To some extent, 3 = A great deal, NA = Not applicable

_____Your ability to incorporate technology into your teaching

_____Your ability to use new teaching methods

_____Your ability to teach basic skills and facts

_____Your classroom management strategies

_____Your ability to teach critical thinking skills to your students

_____Your students' academic achievement

_____The way you assess student work

12a. During the past month, how frequently have you had collaboration meetings with other teachers (for example, for planning lessons) for math?

1 = Never, 2 = Once or twice, 3 = At least weekly, 4 = Daily, NA = Not applicable

AMSTI mathematics__________
Non-AMSTI mathematics__________

12b. To what extent have the math collaboration activities listed in question 12a increased the following?

1 = Not at all or very little, 2 = To some extent, 3 = A great deal, NA = Not applicable

_____Your ability to incorporate technology into your teaching

_____Your ability to use new teaching methods

_____Your ability to teach basic skills and facts

_____Your classroom management strategies

_____Your ability to teach critical thinking skills to your students

_____Your students' academic achievement

_____The way you assess student work

13. During the past two weeks, how many hours (both paid and unpaid time) did you spend planning your math lessons? Please enter the total number of hours.

Math __________

Math Materials

14a. How well is your classroom equipped with the types of math manipulatives you need?

__I have all the types that I need

__I have most the types that I need

__I have some of the types that I need

__I don’t have any manipulatives

14b. How well is your classroom supplied with quantities of math manipulatives?

__I have enough manipulatives for all of my students

__I have enough manipulatives for most of my students

__I have enough manipulatives for some of my students

__I don’t have any manipulatives

Science

15a. Do you currently teach science?

__Yes
__No (Go to question 27)

15b. Do you teach science to students who are not assigned to you on your school’s official computerized class roster?

Examples:
-swapping students based on test scores or other factors
-team teaching where you and another teacher teach both your own students and that teacher’s students
-supporting another teacher to teach the students in that teacher’s classroom
-other

__Yes, please specify ___________________________
__No, I only teach science to students in my own class(es) (Go to question 15f)

15c. Please name the teachers whose students you teach science, or whose students you partner in teaching science, or whom you support in the classroom for science.

Please indicate if you teach ALL the students assigned to this teacher or a smaller group of their students. ________________________________

15d. If you swap science students based on test scores, which test do you use to make that determination? ____

15e. If you swap science students based on test scores, what is the score range of the students you teach? ____

15f. Have you taught the same groups of science students since at least October of this school year?

__Yes
__No; please explain why not: ___________________________________

15g. Is there anything else you would like us to know about your science classes?

____
____

Science Instructional Strategies

The following questions are attempting to understand the number of hours that students receive of each type of instruction. Each question asks you to reflect upon the last two weeks (ten full days) of instruction.

16a. **Think back on your last two weeks (10 full days) of instruction:** approximately how many minutes did your students spend doing science in your class? Please enter the total number of minutes. *Be sure to consider all activities, including discussion, lecture, reading, watching video, hands-on activities, worksheets, and activities that integrate science with other subjects.*
Minutes of science instruction ___________

16b. The number in question 16a represents my minutes of instruction
__Daily
__Weekly
__For two weeks

16c. How many science classes (that is different groups of students) do you teach?

__1 (Go to question 16e)

__2 (Go to question 16d)

__3 (Go to question 16d)

__4 (Go to question 16d)

__5 (Go to question 16d)

__6 (Go to question 16d)

__7 (Go to question 16d)

__8 (Go to question 16d)

__Other, please specify ___________________(Go to question 16d)

16d. Is the number in question 16a the sum of the minutes for all science classes or the average minutes per class?

__Sum

__Average

__Other, please specify_________________________

16e. For the remainder of the science instruction section of this survey, please continue to calculate your responses in the same manner as you did for question 16a.
__OK

16f. Is there anything else you would like us to know about the number of minutes of science instruction that you reported? ____________________________

17. Consider the following description of Inquiry-Based Instruction in which students do ***all*** of the following activities as part of the learning process:

- Make observations
- Pose questions
- Examine books and other sources of information to see what is already known
- Plan investigations
- Review what is already known in light of experimental evidence
- Use tools to *gather*, *analyze*, and *interpret data*
- Propose answers, explanations, and predictions
- Communicate the results

During the past two weeks, approximately how many minutes did students participate in **Inquiry-Based Instruction** in your science class?
Minutes of inquiry-based science instruction____________

18. During the past two weeks, approximately how many minutes did students participate in hands-on science activities (involving active participation; applied, as opposed to theoretical)? Please enter the total number of minutes.
Minutes of hands-on science instruction____________

19. During the past two weeks, how many minutes were your students engaged in science activities that required higher-order thinking skills? (that is, where students advance from skills such as *focusing* and *information gathering* to skills such as *integrating* and *evaluating*.) Please enter the total number of minutes.
Minutes of higher-order thinking skills in science____________

20. During the past two weeks, about how much time did you teach using AMSTI supplied print materials? Please enter the total number of minutes. If you do not teach AMSTI, please enter "0."

Minutes using AMSTI supplied science print materials_________

21. During the past two weeks, what type of science assessments did you use in your classroom? Please check all that apply.

___Informal assessments, such as questioning and observation, to gauge student learning

___Formative paper and pencil assessments (that is, assessments that occur regularly throughout the year in order to inform instruction)

___Performance-based assessments (that is, assessing students based on their application of knowledge, skills, and work habits through the performance of tasks that are meaningful and engaging to students)

___Standardized assessments

___Other, please describe ______________

___I did not administer any science assessments

Science Professional Development

For AMSTI: Please include any professional development you have received as part of the AMSTI program or in any way connected with AMSTI.

For non-AMSTI: Please include all non-AMSTI professional development you have received.

22a. The following questions refer to science Professional Development (PD) activities in which you have participated during the past month. If you have not participated in science professional development activities, please enter "0" for number of hours.

During the past month, how much professional development have you received for your science program. Please do not include support or collaboration meetings. Please enter the total hours of training in each box.

AMSTI science __________
Non-AMSTI science _________

22b. To what extent have the science professional development activities increased the following?

1 = Not at all or very little, 2 = To some extent, 3 = A great deal, NA = Not applicable

_____Your ability to incorporate technology into your teaching

_____Your ability to use new teaching methods

_____Your ability to teach basic skills and facts

_____Your classroom management strategies

_____Your ability to teach critical thinking skills to your students

_____Your students' academic achievement

_____The way you assess student work

The next questions are about asking for and receiving support. If you did not ask for or receive support, please enter "0" for total times.

23a. During the past month, how many times did you try contacting someone for support (for example, for mentoring or coaching) with science instruction?

AMSTI science total times__________
Non-AMSTI science total times __________

23b. During the past month, how many times did someone actually provide support (for example, for mentoring or coaching) with science instruction?

AMSTI science total times__________
Non-AMSTI science total times __________

23c. To what extent have the science support activities listed in question 23b increased the following?

1 = Not at all or very little, 2 = To some extent, 3 = A great deal, NA = Not applicable

_____Your ability to incorporate technology into your teaching
_____Your ability to use new teaching methods
_____Your ability to teach basic skills and facts
_____Your classroom management strategies
_____Your ability to teach critical thinking skills to your students
_____Your students' academic achievement
_____The way you assess student work
_____Your ability to incorporate technology into your teaching

24a. During the past month, how frequently have you had collaboration meetings with other teachers (for example, for planning lessons) for science?

1 = Never, 2 = Once or twice, 3 = At least weekly, 4 = Daily, NA = Not applicable

AMSTI Science __________
Non-AMSTI Sciences__________

24b. To what extent have the science collaboration activities listed in question 24a increased the following?

1 = Not at all or very little, 2 = To some extent, 3 = A great deal, NA ability to use new teaching methods

_____Your ability to teach basic skills and facts
_____Your classroom management strategies
_____Your ability to teach critical thinking skills to your students
_____Your students' academic achievement
_____The way you assess student work

25a. During the past two weeks, how many hours (both paid and unpaid time) did you spend planning your Science lessons? Please enter the total number of hours.

Science__________

Science Materials

26a. How well is your classroom equipped with the types of materials for hands-on science you need?

___I have all the types that I need
___I have most the types that I need
___I have some of the types that I need
___I don't have any hands-on science materials

26b. How well is your classroom supplied with quantities of materials for hands-on science?

___I have enough materials for hands-on science for all of my students
___I have enough materials for hands-on science for most of my students
___I have enough materials for hands-on science for some of my students
___I don't have any materials for hands-on science

Technology

27. To what extent do you agree with the following statements about education technology? Mark one box per row.

1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Neither Disagree nor Agree 4 = Somewhat Agree, 5 = Strongly Agree

_____Educational technology can be used to improve instructional practice.
_____Educational technology can be used to improve teachers' subject matter knowledge.
_____Educational Technology can be used to improve student learning.
_____Educational technology can be used to improve students' performance on standardized tests.
_____Educational technology (the availability of) can help to narrow the achievement gap between traditionally underserved students and other students.

28. Approximately how many **computers** are available for students to use in your classroom?

_____One computer for each student
_____One computer for every two students
_____One computer for every three students
_____One computer for every four students
_____One computer for every five students
_____One computer for every six or more students
_____Did not have computers in the classroom
_____Not Applicable

29. How many **graphing calculators** are available for students to use in your classroom?

_____One graphing calculator for each student
_____One graphing calculator for every two students
_____One graphing calculator for every three students
_____One graphing calculator for every four students
_____One graphing calculator for every five students
_____One graphing calculator for every six or more students
_____Did not have graphing calculators in the classrooms
_____Not Applicable

30. How many **scientific calculators** are available for students to use in your classroom?
_____One graphing calculator for each student
_____One graphing calculator for every two students
_____One graphing calculator for every three students
_____One graphing calculator for every four students
_____One graphing calculator for every five students
_____One graphing calculator for every six or more students
_____Did not have graphing calculators in the classrooms
_____Not Applicable

31. How many **basic/4 function calculators** are available for students to use in your classroom?
_____One basic/4 function calculator for each student
_____One basic/4 function calculator for every two students
_____One basic/4 function calculator for every three students
_____One basic/4 function calculator for every four students
_____One basic/4 function calculator for every five students
_____One basic/4 function calculator for every six or more students
_____Did not have basic/4 function calculators in the classrooms
_____Not Applicable

32. How well are your technical support needs met?
_____Not very well
_____Moderately well
_____Very well
_____Not applicable

Additional Information
33. Teachers who participate in the study for a whole school year by completing all four web-based surveys will receive an honorarium. Please provide your mailing address so that we may mail you your stipend check during the summer of 2009.
____
____
____

34. Is there anything else you would like us to know about your math and/or science program, or about this survey?
____

Appendix H. Data cleaning and data file construction

This appendix reports the methods used to clean and construct the data files for the student-level and program implementation data.

Student-level data

Data were obtained at the student, teacher, and school levels. Unique identifiers at each of these levels allowed researchers to link students to teachers and to schools and to link teachers to schools. This linking was necessary for subsequent hierarchical linear modeling analyses of the results.

Data on the student achievement measure were collected and received from Information System Services at ALSDE in a Microsoft Excel file. The engineering department at Empirical Education used a verification tool that automates the process for checking the validity of the data. The specific criteria for each value were inputted into the verification tool and unexpected values flagged. The values were checked to make sure they were within the expected range, fit the appropriate code or format, were unique, and were required. Any values that did not meet the specific criteria were flagged; if unfixable, they were sent back to Information Systems Services for correcting. Once the data had been verified, they were imported into a Microsoft Access database. SAS was used to read and analyze the data.

Program implementation data

All data from professional development training logs, teacher interviews, and principal interviews were entered by research staff into an SPSS database. Double-data entry procedures were completed for all measures, in order to ensure 100 percent accuracy.

All web-based teacher survey data were automatically entered into a Microsoft Excel file from a web-based survey tool used by Empirical Education. Data that required cleaning were cleaned within Excel; the accuracy of data cleaning was verified through double-data cleaning procedures. Data were then transferred into a Microsoft Access database and categorical data were coded. SAS was used to read and analyze the survey data.

Appendix I. Attrition through study stages for samples used in the confirmatory analysis

This appendix reviews the steps through which the analytic sample was selected for each confirmatory outcome. It does not discuss student or teacher crossover or attrition of students from study schools. Student posttests were collected directly from the state; roster information was collected only once, at the start of the school year. Therefore, it was not possible to know whether a given posttest was collected from a student who was in the same school and condition in which the student started, in a different condition or a different school, or in a public school in Alabama that was not participating in the study. Posttests were not obtained from students who were absent from school when the test was given (because they were ill, were temporarily out of school, or had moved out of state) or who transferred to a private school in Alabama.

Selection of the Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome sample

Below appear changes in the numbers of schools, teachers, and students from the point of randomization to the point at which the analytic sample associated with the SAT 10 mathematics problem solving outcome was identified (table I1).

Table I1 Attrition from analytical sample associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome

	AMSTI			Control		
Item	Number of schools	Number of teachers	Number of students	Number of schools	Number of teachers	Number of students
Randomization	41	na	na	41	na	na
Numbers indicated in fall rosters	41	249	12,065	41	233	10,492
Loss because of disability	0	–3	–1,548	0	–4	–1,383
Baseline sample	41	246	10,517	41	229	9,109
Loss because of students transferring from Subexperiment 1 to Subexperiment 2	0	0	–24	0	0	–26
Available cases	41	246	10,493	41	229	9,083
Loss because of missing student identifier	0	0	0	0	0	0
Available cases	41	246	10,493	41	229	9,083
Loss because of missing school identifier	0	0	0	0	0	0
Available cases	41	246	10,493	41	229	9,083
Loss because of missing posttests	0	–2	–471	0	0	–392
Number of cases in sample	41	244	10,022	41	229	8,691

na is not applicable.
Source: Student achievement data from tests administered as part of the state's accountability system.

Random assignment

Forty-one schools were randomly assigned to the AMSTI condition, and 41 schools were randomly assigned to the control condition.

Confirmation of rosters and teacher assignments

Districts provided information about teachers and student rosters at the start of the school year following randomization. The record for the number of teachers and students began at the time roster information was received. The rosters from AMSTI schools included 249 mathematics teachers (with 12,065 students) from grades 4–8. The rosters from the control schools included 233 mathematics teachers (with 10,492 students) from grades 4–8.

Exclusion of data on students with disabilities

All students with disabilities, including those listed on the rosters of “general” mathematics and science classes, were regarded as ineligible to participate in the study. Disability was defined as a mental, emotional, physical, or learning disability, according to the district’s formal designation.[111] All students with disabilities were identified and were receiving special education services. The school districts determined a student’s disability status and whether the student was ineligible for the pretest or required testing modifications because of it.

Data on 1,548 students with disabilities were excluded from AMSTI schools; data on 1,383 students with disabilities were excluded from the control schools. The remaining sample consisted of 41 AMSTI schools (with 246 teachers and 10,517 students) and 41 control schools (with 229 teachers and 9,109 students).

Exclusion of data on students who moved between subexperiments

Data from students whose identifiers appeared twice in the analytic sample were excluded from analysis. These were students who moved between subexperiments. Because the first subexperiment started one year before the second one, moving across experiments would have resulted in unknown levels of exposure to the intervention. Data from 24 students from AMSTI schools and 26 students from control schools were excluded for this reason. Removing these students’ data did not result in a decline in the number of teachers or schools.

Exclusion of data on students without valid identifiers

Student data were excluded if they were missing valid student or school identifiers. It was critical to have this information in order to properly model school membership in the analysis. All students were linked to school identifiers.

---

[111] These designations are also used by the state for accountability purposes to provide the counts of students with disabilities.

Exclusion of data on students without valid posttests

Student data were excluded from analysis if the posttest score was missing or lay outside the range of the posttest scale. Data on 471 students in AMSTI school and 392 students in control schools were excluded because they were missing posttests. In AMSTI schools, the loss of these data resulted in the loss of two teachers but no schools. The loss of the data had no effect on the number of control teachers or schools.

Student data were not excluded if they were missing values for one or more covariates in the model, including the pretest. A dummy variable method was used to handle missing values for covariates, which involved replacing the unobserved value with a constant and, for each covariate, adding an indicator variable with a value of 1 or 0 to signify whether the value of the corresponding covariate was observed or unobserved (missing).

The starting sample included 41 AMSTI schools (with 249 mathematics teachers and 12,065 students) and 41 control schools (with 233 mathematics teachers and 10,492 students). The baseline sample for the analysis of mathematics outcomes, which was limited to students without disabilities, consisted of 41 AMSTI schools (with 246 teachers and 10,517 students) and 41 control schools (with 229 teachers and 9,109 students). The analytic sample for the analysis of mathematics outcomes consisted of 41 AMSTI schools (with 244 teachers and 10,022 students) and 41 control schools (with 229 teachers and 8,691 students).

Selection of the Stanford Achievement Test Tenth Edition (SAT 10) science outcome sample

Below appear changes in the numbers of schools, teachers, and students from the point of randomization to the point at which the analytic sample associated with the SAT 10 science outcome was identified (table I2). This process parallels the one for mathematics described above, except that the student sample was limited to grades 5 and 7.

Table I2 Attrition from analytical sample associated with Stanford Achievement Test Tenth Edition (SAT 10) science outcome

Item	AMSTI			Control		
	Number of schools	Number of teachers	Number of students	Number of schools	Number of teachers	Number of students
Randomization	41	na	na	41	na	na
Numbers indicated in fall rosters	41	233	12,065	41	213	10,492
Loss because not in grades 5 or 7	0	–128	–6,972	0	–116	–6,284
Cases in grades 5 and 7	41	105	5,093	41	97	4,208
Loss because of disability	–2	–2	–613	0	–2	–520
Available cases (baseline sample)	39	103	4,480	41	95	3,688
Loss because of students transferring from Subexperiment 1 to Subexperiment 2)	0	0	–14	0	0	–9
Available cases	39	103	4,466	41	95	3,679
Loss because of missing student identifier	0	0	0	0	0	0
Available cases	39	103	4,466	41	95	3,679
Loss because of missing school identifier	0	0	0	0	0	0
Available cases	39	103	4,466	41	95	3,679
Loss because of missing posttests	0	–1	–384	–1	–5	–233
Number of cases in sample	39	102	4,082	40	90	3,446

na is not applicable.
Source: Student achievement data from tests administered as part of the state's accountability system.

Random assignment phase

Forty-one schools were randomly assigned to the AMSTI condition, and 41 schools were randomly assigned to the control condition.

Confirmation of rosters and teacher assignments

Districts provided information about teachers and student rosters at the start of the school year, following randomization. The record for the number of teachers and students began at the time roster information was received. The rosters from AMSTI schools included 233 science teachers (with 12,065 students) from grades 4–8. The rosters from the control schools included 213 science teachers (with 10,492 students) from grades 4–8.

Participation of specific grade levels

Only data from students in grades 5 and 7 (the grades for which it is mandatory to assess performance on Alabama's science achievement test) were included in the analysis of the impact of AMSTI on science performance. Students in other grades are also tested in science, but they are tested at the discretion of districts, principals, and teachers. Science outcomes from these grades were not included in the analysis, because voluntary participation could lead to selection effects that could have biased the outcomes. After limiting data from students to grades 5 or 7, 41 AMSTI schools (with 105 teachers and 5,093 students) and 41 control schools (with 97 teachers and 4,208 students) remained in the sample.

Exclusion of data on students with disabilities

Data were excluded on 613 students with disabilities from AMSTI schools and 520 students with disabilities from the control schools. The remaining sample consisted of 39 AMSTI schools (with 103 teachers and 4,480 students) and 41 control schools (with 95 teachers and 3,688 students).

Exclusion of data on students who moved between subexperiments

Data from students whose identifiers appeared twice in the analytic sample were excluded from analysis. These were students who moved between subexperiments. Because the first subexperiment started one year before the second one, moving across the experiments would have resulted in unknown levels of exposure to the intervention. Data from 14 students from AMSTI schools and 9 students from control schools were excluded for this reason. Removing these students' data did not result in a decline in the number of teachers or schools.

Exclusion of data on students without valid identifiers

Student data were excluded if they were missing valid student or school identifiers. It was critical to have this information in order to properly model school membership in the analysis. All students were linked to school identifiers.

Exclusion of data on students without valid posttests

Student data were excluded from analysis if a posttest score was missing or lay outside the range of the posttest scale. Data from 384 students in AMSTI school and 233 students in control schools were excluded because they were missing posttests. Exclusion of these data resulted in the loss of one teacher but no schools in the AMSTI condition and five teachers and one school in the control condition.

Student data were not excluded if they were missing values for one or more covariates, including the pretest. A dummy variable method was used to address missing values for covariates, which involved replacing the unobserved value with a constant and, for each covariate, adding an indicator variable with a value of 1 or 0 to signify whether the value of the corresponding covariate is observed or unobserved (missing).

The starting sample included 41 AMSTI schools (with 233 teachers and 12,065 students) and 41 control schools (with 213 teachers and 10,492 students). The baseline sample for the analysis of science outcomes, which was limited to students without disabilities in grades 5 or 7, consisted of 39 AMSTI schools (with 103 teachers and 4,480 students) and 41 control schools (with 95 teachers and 3,688 students). The analytic sample for the analysis of science outcomes consisted of 39 AMSTI schools (with 102 teachers and 4,082 students) and 40 control schools (with 90 teachers and 3,446 students).

Selection of the sample for the active learning score outcome

Below appear changes in the sample from the point at which completed surveys were collected from teachers to the point at which the analytic sample was identified (tables I3 and I4).

Table I3 Attrition from analytical sample associated with active learning in mathematics outcome

	AMSTI		Control	
Item	Number of schools	Number of teachers	Number of schools	Number of teachers
Randomization	41	na	41	na
Loss before completing surveys	0	na	–1	na
Completed surveys (baseline sample)	41	221	40	205
Loss because of missing teacher identifier	0	0	0	0
Available cases	41	221	40	205
Loss because of missing school identifier	0	0	0	0
Available cases	41	221	40	205
Loss because of missing valid active learning in mathematics score	0	–8	0	–13
Number of cases in sample	41	213	40	192

na is not applicable.
Source: Teacher survey data.

Table I4 Attrition from analytical sample associated with active learning in science outcome

Item	AMSTI		Control	
	Number of schools	Number of teachers	Number of schools	Number of teachers
Randomization	41	na	41	na
Loss before completing survey	–1	na	–1	na
Completed surveys (baseline sample)	40	203	40	192
Loss because of missing teacher identifier	0	0	0	0
Available cases	40	203	40	192
Loss because of missing school identifier	0	0	0	0
Available cases	40	203	40	192
Loss because of missing valid active learning in science score	0	–9	–2	–17
Number of cases in sample	40	194	38	175

na is not applicable.
Source: Student achievement data from tests administered as part of the state's accountability system.

Random assignment phase

Forty-one schools were randomly assigned to the AMSTI condition, and 41 schools were randomly assigned to the control condition.

Survey completion

As described in the data collection section, teachers were asked to consent to participate in online surveys. Teachers who did so were asked to complete four monthly surveys between January and April of Year 1. Data from specific questions from these surveys were used to compute a composite variable called the *active learning score* (see the data analysis method section for a description of how the active learning score was computed).

The sample was first limited to teachers who taught the appropriate grades and subjects. All AMSTI schools had at least one mathematics teacher (in non-special education classes in from grades 4–8) who completed the surveys. One school in the control condition dropped out of the study shortly after randomization took place. Although this school was not excluded from the eligible sample related to the student outcomes (because the data were collected at the district and state level), it did not consent to participate in the online surveys and was not included in the baseline sample of the classrooms practice outcomes.

In AMSTI schools, 221 mathematics teachers (in 41 schools) completed the survey. In control schools, 205 mathematics teachers (in 40 schools) completed the survey.

Exclusion of data on teachers without valid identifiers

All mathematics and science teachers in both conditions had valid teacher and school identifiers.

Exclusion of data on teachers without valid active learning scores

If teachers were missing data for all four survey occasions, the teacher's data were removed from analysis (for complete definition of a valid active learning score, see the data analysis method section). Data from 8 mathematics teachers in AMSTI schools and 13 mathematics teachers in control schools were removed from this analysis because of missing valid scores. No schools were eliminated because of missing valid scores. Data from 9 science teachers in AMSTI schools and 17 science teachers in control schools removed from this analysis because of missing valid scores. Data from two control schools were removed from this analysis because of missing valid scores form the teachers.

The baseline sample was limited to teachers who taught the appropriate subjects and grades and completed surveys. For the analysis of active learning score in mathematics outcomes, the sample consisted of 41 AMSTI schools (with 221 teachers) and 40 control schools (with 205 teachers). The analytic sample for the analysis of the active learning score in mathematics outcomes consisted of 41 AMSTI schools (with 213 teachers) and 40 control schools (with 192 teachers).

The baseline sample for the analysis of active learning score in science outcomes consisted of 40 AMSTI schools (with 203 teachers) and 40 control schools (with 192 teachers). The analytic sample for the analysis of the active learning score in science outcomes consisted of 40 AMSTI schools (with 194 teachers) and 38 control schools (with 175 teachers).

Appendix J. Description of degree rank

Teacher quality depends not only on overall years of teaching experience but also on teacher knowledge of content area (Amrein-Beardsley 2006; Center for Public Education 2005). Whether or not mathematics and science teachers have an undergraduate or graduate major or minor in their teaching subject matter is one potential indicator of content knowledge (National Science Foundation 2006a). Out-of-field teaching is defined as “a mismatch between the subjects a teacher teaches and that teacher’s academic training and/or certification” (National Science Foundation 2006b, p. 55). The *Science and Engineering Indicators* (National Science Foundation 2006a) reports that “in 1999–2000, 71 percent of public school teachers who taught mathematics in grades 7–12 had a college major or minor in mathematics, and 77 percent of public school teachers who taught science in these same grades had a college major or minor in science” (p. 34). In Alabama 86 percent of mathematics teachers and 81 percent of science teachers in public school grades 7–12 had a college major or minor in their teaching subject area (National Science Foundation 2006c).

Based on these criteria, a *degree rank* variable was created that categorized the study’s teachers’ postsecondary major and minor degrees according to the content of the degree and current teaching assignment. For elementary teachers, the degree rank was based on the presence or absence of at least one degree in elementary education. For middle school teachers, the degree rank was based on whether teachers had a degree in mathematics or science.

The following codes were used for elementary teachers:

0 = no elementary education degree (out-of-field teacher)
1 = bachelor’s, bachelor’s plus additional coursework, or master’s in elementary education (that is, at least one major in content area)
2 = bachelor’s and additional coursework or master’s in a combination of elementary education, reading, math, or science (that is, two or more majors in content area)

The following codes were used for middle-school teachers:

0 = no degree in math or science content (out-of-field teacher)
1 = bachelor’s, bachelor’s plus additional coursework, or master’s in math or science content (that is, at least one major in content area)
2 = bachelor’s and additional coursework or master’s in a combination of middle school education, math, or science (that is, two or more majors in content area)

Appendix K. Equivalence of Year 1 baseline and analyzed samples for confirmatory student-level and classroom practice outcomes

This appendix examines the equivalence of the baseline sample and the sample used to analyze the student-level and classroom practice outcomes in Year 1 (tables K1–K4).

Table K1 Year 1 mean baseline sample characteristics associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving and science outcomes after one year

		Mathematics			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
Teacher characteristic							
Average of school percent of out-of-field teachers	Average in Each Condition	19.1		26.8	8.8		12.40
	Standard Deviation	23.2		29.6	22.0		29.3
	Sample Size (Schools)	41		39	35		39
	Average Difference		–7.7			–3.6	
	Standard Error		5.9			6.1	
	Test statistic		*t* = 1.30			*t* = 0.59	
	***p*-value**		.20			.56	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	57.0		55.9	69.93		59.4
	Standard Deviation	30.4		30.6	35.7		40.6
	Sample Size (Schools)	41		39	35		39
	Average Difference		1.1			10.5	
	Standard Error		6.8			8.9	
	Test statistic		*t* = 0.15			*t* = 1.18	
	***p*-value**		.88			.24	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	23.9		17.3	21.3		28.2
	Standard Deviation	24.7		22.0	33.3		38.8
	Sample Size (Schools)	41		39	35		39
	Average Difference		6.6			–6.9	
	Standard Error		5.2			8.5	
	Test statistic		*t* = 1.27			*t* = 0.82	
	***p*-value**		.21			.41	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	27.6		24.5	28.5		32.1
	Standard Deviation	26.9		24.6	38.1		37.9
	Sample Size (Schools)	41		39	35		39
	Average Difference		3.0			–3.6	
	Standard Error		5.8			8.8	

		Mathematics			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
	Test statistic		*t* = 0.53			*t* = 0.41	
	***p*-value**		.60			.69	
Average of school percent of teachers with less than four years' teaching experience in subject area	Average in Each Condition	28.6		30.3	34.8		29.7
	Standard Deviation	27.1		26.8	40.0		34.7
	Sample Size (Schools)	41		39	35		39
	Average Difference		–1.7			5.1	
	Standard Error		6.0			8.7	
	Test statistic		*t* = 0.29			*t* = 0.58	
	***p*-value**		.77			.56	
Student Characteristic							
Average of school percent of boys	Average in Each Condition	49.4		49.6	48.1		51.1
	Standard Deviation	4.1		4.1	5.1		6.4
	Sample Size (Schools)	41		41	39		41
	Average Difference		–0.2			–3.0	
	Standard Error		0.9			2.4	
	Test statistic		*t* = 0.25			*t* = 1.29	
	***p*-value**		.80			.02	
Average of school percent of minority students	Average in Each Condition	51.0		46.9	51.8		47.3
	Standard Deviation	34.6		33.2	35.1		33.4
	Sample Size (Schools)	41		41	39		41
	Average Difference		4.1			4.5	
	Standard Error		7.5			7.7	
	Test statistic		*t* = 0.55			*t* = 0.59	
	***p*-value**		.58			.56	
Average of school percent of students proficient in English	Average in Each Condition	98.3		98.6	98.4		98.3
	Standard Deviation	3.3		2.4	3.0		2.7
	Sample Size (Schools)	41		41	39		41
	Average Difference		–0.2			0.1	
	Standard Error		0.6			0.6	
	Test statistic		*t* = 0.36			*t* = 0.23	
	***p*-value**		.72			.82	
Average of school percent of students enrolled in the free or reduced-price lunch	Average in Each Condition	63.2		64.7	64.2		65.3
	Standard Deviation	24.8		24.2	26.3		24.4
	Sample Size (Schools)	41		41	39		41
	Average Difference		–1.5			–1.1	
	Standard Error		5.4			5.7	

		Mathematics			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
program	**Test statistic**		*t* = 0.28			*t* = 0.19	
	***p*-value**		.78			.85	
Average of school percent of students in grade 4	Average in Each Condition	25.7		22.4	na		na
	Standard Deviation	24.7		20.7	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		3.3			na	
	Standard Error		5.0			na	
	Test statistic		*t* = 0.65			na	
	***p*-value**		.52			na	
Average of school percent of students in grade 5	Average in Each Condition	25.0		23.3	61.3		57.5
	Standard Deviation	18.7		19.5	40.6		45.7
	Sample Size (Schools)	41		41	39		41
	Average Difference		1.7			3.8	
	Standard Error		4.2			9.7	
	Test statistic		*t* = 0.41			*t* = 0.39	
	***p*-value**		.68			.69	
Average of school percent of students in grade 6	Average in Each Condition	14.3		18.8	na		na
	Standard Deviation	14.4		16.6	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		–4.5			na	
	Standard Error		3.4			na	
	Test statistic		*t* = 1.32			na	
	***p*-value**		.19			na	
Average of school percent of students in grade 7	Average in Each Condition	16.7		18.1	38.7		42.5
	Standard Deviation	19.0		20.6	40.6		45.7
	Sample Size (Schools)	41		41	39		41
	Average Difference		–1.4			na	
	Standard Error		0.4			na	
	Test statistic		*t* = 0.32			na	
	***p*-value**		.75			na	
Average of school percent of students in grade 8	Average in Each Condition	18.4		17.4	na		na
	Standard Deviation	22.2		19.2	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		0.9			na	
	Standard Error		4.6			na	
	Test statistic		*t* = 0.20			na	

		Mathematics			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
	***p*-value**		.84			na	
School average pretest score and sample size							
SAT10[a]	Pretest Score	637.0		639.8	645.9		649.1
	Standard Deviation	23.2		20.1	19.0		15.6
	Sample Size (Schools)	41		41	39		41
	Average Difference		–2.8			–3.2	
	Standard Error		4.8			3.9	
	Test statistic		t = 0.58			t = 0.83	
	***p*-value**		.56			.41	
Sample Size	Number of schools = 41 Number of teachers = 246 Number of students = 10,517		Number of schools = 41 Number of teachers = 229 Number of students = 9,109		Number of schools = 39 Number of teachers = 103 Number of students = 4,480		Number of schools = 41 Number of teachers = 95 Number of students = 3,688

na is not applicable.

Note: Detail may not sum to totals because of rounding. The number of schools and teachers for the comparisons varied slightly because of missing data. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Grade level for mathematics problem solving and teacher degree rank are categorical variables with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of students across grades for the mathematics problem solving outcome (p = .76) and teacher degree rank (p = .10 for mathematics problem solving, p = .76 for science). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in negative twice the log likelihood statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the p-value was .40 for mathematics problem solving and .22 for science.)

a. The SAT 10 mathematics problem solving pretest was used for the mathematics outcome. The SAT 10 reading pretest was used for the science outcome.

Source: Student achievement data from tests administered as part of the state's accountability system, student demographic data from state data system, and teacher survey data.

Table K2 Year 1 mean baseline sample characteristics associated with active learning outcomes after one year

		Mathematics			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
		(*n*=41)		**(*n*=40)**	**(*n*=40)**		**(*n*=40)**
Average of school percent of out-of-field teachers	Average in Each Condition	19.8		26.1	12.5		20.3
	Standard Deviation	23.1		29.1	18.1		24.7
	Average Difference		–6.2			–7.8	
	Standard Error		5.8			4.8	
	Test statistic		*t* = 1.07			*t* = 1.61	
	***p*-value**		.29			.11	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	56.4		56.9	60.1		55.3
	Standard Deviation	29.0		30.9	29.7		34.1
	Average Difference		0.5			4.8	
	Standard Error		6.7			7.2	
	Test statistic		*t* = 0.08			*t* = 0.67	
	***p*-value**		.94			.50	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	23.8		17.1	27.3		24.4
	Standard Deviation	21.7		22.3	28.5		28.1
	Average Difference		6.8			3.0	
	Standard Error		4.9			6.3	
	Test statistic		*t* = 1.38			*t* = 0.47	
	***p*-value**		.17			.64	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	26.6		27.0	27.9		30.1
	Standard Deviation	26.3		26.9	29.8		25.5
	Average Difference		–0.5			–2.2	
	Standard Error		5.9			6.2	
	Test statistic		*t* = 0.08			*t* = 0.36	
	***p*-value**		.94			.72	
Average of school percent of teachers with less than four years' teaching experience in subject area	Average in Each Condition	29.4		31.3	31.0		32.7
	Standard Deviation	26.9		26.3	29.5		23.3
	Average Difference		–1.9			–1.7	
	Standard Error		5.9			5.9	
	Test statistic		*t* = 0.32			*t* = 0.29	
	***p*-value**		.75			.77	
Sample Size	Number of schools = 41 Number of teachers = 221		Number of schools = 40 Number of teachers = 205		Number of schools = 40 Number of teachers = 203		Number of schools = 40 Number of teachers = 192

Note: Detail may not sum to totals because of rounding. The number of schools and teachers for the comparisons varied slightly because of missing data. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Teacher degree rank is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of teacher degree rank ($p = .07$ for mathematics, $p = .23$ for science). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the *p*-value was .63 for active learning in mathematics and .43 for active learning in science.)
Source: Teacher survey data.

Table K3 Year 1 mean analytic sample characteristics associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving and science outcomes after one year

		Mathematics			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
Teacher characteristic							
Average of school percent of out-of-field teachers	Average in Each Condition	19.5		26.8	8.8		12.7
	Standard Deviation	23.6		29.6	22.0		29.6
	Sample Size (Schools)	41		39	35		38
	Average Difference		–7.3			–3.9	
	Standard Error		6.0			6.2	
	Test statistic		t = 1.22			t = 0.64	
	***p*-value**		.23			.53	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	56.6		55.9	71.4		58.3
	Standard Deviation	30.3		30.6	35.9		40.6
	Sample Size (Schools)	41		39	35		38
	Average Difference		.6			13.0	
	Standard Error		6.8			9.0	
	Test statistic		t = 0.09			t = 1.45	
	***p*-value**		.93			.15	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	23.9		17.3	19.8		29.0
	Standard Deviation	24.7		22.0	33.1		39.0
	Sample Size (Schools)	41		39	35		38
	Average Difference		6.6			–9.1	
	Standard Error		5.2			8.5	
	Test statistic		t = 1.27			t = 1.07	
	***p*-value**		.21			.29	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	27.6		24.5	28.5		32.0
	Standard Deviation	26.9		24.6	38.1		38.4
	Sample Size (Schools)	41		39	35		38
	Average Difference		3.0			–3.6	
	Standard Error		5.8			9.0	
	Test statistic		t = 0.53			t = 0.40	
	***p*-value**		.60			.69	
Average of school percent of teachers with less than four years' teaching experience in	Average in Each Condition	28.6		30.3	34.8		30.9
	Standard Deviation	27.1		26.8	40.0		36.8
	Sample Size (Schools)	41		39	35		38
	Average Difference		–1.7			3.8	
	Standard Error		6.0			9.0	

		Mathematics			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
subject area	**Test statistic**		*t* = 0.29			*t* = 0.43	
	***p*-value**		.77			.67	
Student characteristic							
Average of school percent of boys	Average in Each Condition	49.3		49.3	48.2		50.6
	Standard Deviation	4.2		3.9	5.6		6.9
	Sample Size (Schools)	41		41	39		40
	Average Difference		0.0			–2.4	
	Standard Error		0.9			1.7	
	Test statistic		*t* = 0.04			*t* = 1.41	
	***p*-value**		.97			.10	
Average of school percent of minority students	Average in Each Condition	51.2		46.9	52.0		48.6
	Standard Deviation	34.6		33.4	35.2		33.0
	Sample Size (Schools)	41		41	39		40
	Average Difference		4.3			3.4	
	Standard Error		7.5			7.7	
	Test statistic		*t* = 0.57			*t* = 0.45	
	***p*-value**		.57			.66	
Average of school percent of students proficient in English	Average in Each Condition	98.3		98.7	98.4		98.3
	Standard Deviation	3.3		2.3	3.1		2.6
	Sample Size (Schools)	41		41	39		40
	Average Difference		–0.3			0.1	
	Standard Error		0.63			0.6	
	Test statistic		*t* = 0.51			*t* = 0.14	
	***p*-value**		.61			.89	
Average of school percent of students enrolled in the free or reduced-price lunch program	Average in Each Condition	63.2		64.6	65.5		65.6
	Standard Deviation	24.8		24.3	26.8		24.8
	Sample Size (Schools)	41		41	39		40
	Average Difference		–1.5			–0.1	
	Standard Error		5.4			5.8	
	Test statistic		*t* = 0.27			*t* = 0.01	
	***p*-value**		.79			.99	
Average of school percent of students in grade 4	Average in Each Condition	25.8		22.5	na		na
	Standard Deviation	24.8		20.7	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		3.3			na	
	Standard Error		5.0			na	

		Mathematics			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
	Test statistic		*t* = 0.65			na	
	***p*-value**		.52			na	
Average of school percent of students in grade 5	Average in Each Condition	25.2		23.3	60.9		56.9
	Standard Deviation	18.7		19.7	40.8		46.2
	Sample Size (Schools)	41		41	39		40
	Average Difference		2.0.			4.0	
	Standard Error		4.2			9.8	
	Test statistic		*t* = 0.46			*t* = 0.41	
	***p*-value**		.65			.68	
Average of school percent of students in grade 6	Average in Each Condition	14.5		18.8	na		na
	Standard Deviation	14.5		16.6	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		–4.4			na	
	Standard Error		3.4			na	
	Test statistic		*t* = 1.27			na	
	***p*-value**		.21			na	
Average of school percent of students in grade 7	Average in Each Condition	16.8		18.2	39.1		43.1
	Standard Deviation	18.9		20.6	40.8		46.2
	Sample Size (Schools)	41		41	39		40
	Average Difference		–1.4			na	
	Standard Error		4.4			na	
	Test statistic		*t* = 0.33			na	
	***p*-value**		.74			na	
Average of school percent of students in grade 8	Average in Each Condition	17.8		17.2	na		na
	Standard Deviation	22.4		19.4	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		0.6			na	
	Standard Error		4.6			na	
	Test statistic		*t* = 0.13			na	
	***p*-value**		.90			na	
School average pretest score and sample size							
SAT10[a]	Pretest Score	636.9		639.9	646.4		649.0
	Standard Deviation	23.0		20.2	19.9		15.8
	Sample Size (Schools)	41		41	39		40
	Average Difference		–3.0			–2.6	
	Standard Error		4.8			4.0	

		Mathematics			Science		
Sample characteristic		AMSTI schools		Control schools	AMSTI schools		Control schools
	Test statistic		t = 0.61			t = 0.66	
	p-value		.54			.51	
Sample Size		Number of schools = 41 Number of teachers = 244		Number of schools = 41 Number of teachers = 229	Number of schools = 39 Number of teachers = 102		Number of schools = 40 Number of teachers = 90

na is not applicable.

Note: Detail may not sum to totals because of rounding. The number of schools and teachers for the comparisons varied slightly because of missing data. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Grade level for mathematics problem solving and teacher degree rank are categorical variables with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of students across grades for the mathematics problem solving outcome (p = .76) and teacher degree rank (p = .10 for mathematics problem solving, p = .53 for science). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the p-value was .45 for mathematics problem solving and .48 for science.)

a. The SAT 10 mathematics problem solving pretest was used for the mathematics outcome. The SAT 10 reading pretest was used for the science outcome.

Source: Student achievement data from tests administered as part of the state's accountability system, student demographic data from state data system, and teacher survey data.

Table K4 Year 1 mean analytic sample characteristics associated with active learning outcomes after one year

Sample characteristic		Mathematics: AMSTI schools (*n*=41)	Mathematics	Mathematics: Control schools (*n*=40)	Science: AMSTI schools (*n*=40)	Science	Science: Control schools (*n*=38)
Average of school percent of out-of-field teachers	Average in Each Condition	19.0		28.0	11.7		21.8
	Standard Deviation	23.2		31.4	17.9		27.4
	Average Difference		–9.0			–10.1	
	Standard Error		6.1			5.2	
	Test statistic		t = 1.48			*t* = 1.93	
	***p*-value**		.14			.06	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	56.6		55.7	62.3		51.6
	Standard Deviation	28.7		32.1	30.9		34.3
	Average Difference		0.9			10.6	
	Standard Error		6.8			7.4	
	Test statistic		*t* = 0.13			*t* = 1.44	
	***p*-value**		.90			.15	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	24.5		16.3	26.0		26.6
	Standard Deviation	22.2		21.7	28.6		30.5
	Average Difference		8.2			–0.6	
	Standard Error		4.9			6.7	
	Test statistic		*t* = 1.67			*t* = 0.09	
	***p*-value**		.10			.93	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	27.1		28.5	29.2		28.9
	Standard Deviation	26.7		29.4	31.8		23.7
	Average Difference		–1.5			0.3	
	Standard Error		6.2			6.4	
	Test statistic		*t* = 0.24			*t* = 0.05	
	***p*-value**		.81			.96	
Average of school percent of teachers with less than four years' teaching experience in subject area	Average in Each Condition	28.5		32.1	32.4		36.1
	Standard Deviation	26.9		29.0	31.4		27.1
	Average Difference		–3.6			–3.7	
	Standard Error		6.2			6.7	
	Test statistic		*t* = 0.58			*t* = 0.56	
	***p*-value**		.56			.58	
Sample Size	Number of schools = 41 Number of teachers = 213		Number of schools = 40 Number of teachers = 192	Number of schools = 40 Number of teachers = 194		Number of schools = 38 Number of teachers = 175	

Note: Detail may not sum to totals because of rounding. The number of schools and teachers for the comparisons varied slightly because of missing data. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Teacher degree rank is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of teacher degree rank ($p = .04$ for mathematics, $p = .29$ for science). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the *p*-value was .40 for active learning in mathematics and .34 for active learning in science.)
Source: Teacher survey data.

Appendix L. Internal consistency and validity of active learning measures

Information was obtained about the internal consistency and validity of the active learning measures. For the scale measuring instructional strategies for active learning, Cronbach's alpha was .89 for AMSTI schools and .78 for control schools in mathematics and .88 for AMSTI schools and .92 for control schools in science. The content validity of the active learning scales was confirmed by examining the correlations among the items that comprise each scale and performing confirmatory principal components factor analyses (tables L1 and L2).

Table L1 Correlation coefficients for instructional strategies for active learning for mathematics

Instructional strategies		Inquiry-based instruction	Hands-on instruction	Higher-order thinking
AMSTI schools				
Inquiry-based instruction	*r*	1.00	.77***	.76***
	n	211	211	210
Hands-on instruction	*r*		1.00	.74***
	n		213	212
Higher-order thinking	*r*			1.00
	n			212
Control schools				
Inquiry-based instruction	*r*	1.00	.64***	.52***
	n	190	190	189
Hands-on instruction	*r*		1.00	.53***
	n		192	191
Higher-order thinking	*r*			1.00
	n			191

*** Significant at $p < .01$.
Source: Teacher survey data.

Table L2 Correlation coefficients for instructional strategies for active learning for science

Instructional strategies		Inquiry-based instruction	Hands-on instruction	Higher-order thinking
AMSTI schools				
Inquiry-based instruction	*r*	1.00	.80***	.69***
	n	192	192	192
Hands-on instruction	*r*		1.00	.67***
	n		193	192
Higher-order thinking	*r*			1.00
	n			193
Control schools				
Inquiry-based instruction	*r*	1.00	.77**	.78**
	n	173	173	173
Hands-on instruction	*r*		1.00	.86**
	n		175	174
Higher-order thinking	*r*			1.00
	n			174

** Significant at $p < .05$; ***Significant at $p < .01$.
Source: Teacher survey data.

Appendix M. Number of students and teachers in schools in analytic samples used to analyze Year 1 confirmatory questions

This appendix provides information on the number of students and teachers included in the analytical sampled used to analyze the Year 1 confirmatory questions (table M1).

Table M1 Number of students and teachers in schools in analytic samples used to analyze Year 1 confirmatory questions

Outcome	Total number of schools	Minimum number per school	Maximum number per school	Median number per school	Mean number per school
Student level					
SAT 10 mathematics problem solving	82	51	904	184.5	228.2
SAT 10 science	79	3	320	78.0	95.3
Teacher level					
Active learning in mathematics	81	< 4	26	4.0	5.0
Active learning in science	78	< 4	17	4.0	4.7

Source: Student achievement data from tests administered as part of the state's accountability system and teacher survey data.

Appendix N. Attrition through study stages for samples used in Year 1 exploratory analysis

This appendix reviews the steps through which the sample was selected for the exploratory analysis in Year 1. It examines the attrition in the samples associated with the SAT reading outcomes, teacher content knowledge in mathematics and in science, and student engagement in mathematics and science.

Stanford Achievement Test Tenth Edition (SAT 10) reading outcome

Below are changes in the numbers of schools and students from the point of randomization to the point at which the analytic sample associated with the SAT 10 reading outcome were identified (table N1).[112]

Table N1 Attrition from analytical sample associated with Stanford Achievement Test Tenth Edition (SAT 10) reading outcome

	AMSTI		Control	
Item	Number of schools	Number of students	Number of schools	Number of students
Randomization	41	na	41	na
Numbers indicated in fall rosters	41	12,065	41	10,492
Loss because of disability	0	–1,548	0	–1,383
Available cases (baseline sample)	41	10,517	41	9,109
Loss because of students transferring from Subexperiment 1 to Subexperiment 2	0	–24	0	–26
Available cases	41	10,493	41	9,083
Loss because of missing student identifier	0	0	0	0
Available cases	41	10,493	41	9,083
Loss because of missing school identifier	0	0	0	0
Available cases	41	10,493	41	9,083
Loss because of missing posttests	0	–474	0	–392
Number of cases in sample	41	10,019	41	8,691

na is not applicable.
Source: Student achievement data from tests administered as part of the state's accountability system.

[112] Data on reading teachers were not collected. Therefore, only counts at the school and student levels are presented.

Random assignment phase

Forty-one schools were randomly assigned to the AMSTI condition, and 41 schools were randomly assigned to the control condition.

Confirmation of rosters and teacher assignments

Districts provided information about student rosters at the start of the school year following randomization. The record for the number of students began at the time roster information was received. The rosters from AMSTI schools included 12,065 students from grades 4–8. The rosters from control schools included 10,492 students from grades 4–8.

Exclusion of data on students with disabilities

Data for 1,548 students with disabilities were excluded from AMSTI schools, and data for 1,383 students with disabilities were excluded from control schools. The remaining sample consisted of 41 AMSTI schools (with 10,517 students) and 41 control schools (with 9,109 students).

Exclusion of data on students who moved between subexperiments

Data for students whose identifiers appeared twice in the analytic sample were excluded from analysis. These were students who moved between subexperiments. Because the first subexperiment started one year before the second one, moving across the subexperiments would have resulted in unknown levels of exposure to the intervention. Data for 24 students from AMSTI schools and 26 students from control schools were excluded for this reason. Removing these data did not result in a decline in the number of schools.

Exclusion of data on students without valid identifiers

All students were linked to school identifiers.

Exclusion of data on students without valid posttests

Students' data were excluded from analysis if they were missing a posttest score or if their posttest score lay outside the range of the posttest scale. Data for 474 students in AMSTI schools and 392 students in control schools were excluded because they were missing posttests. This did not result in a decline in the number of schools in either condition.

Treatment of missing data for covariates

Students' data were not excluded if they were missing values for one or more covariates in the model, including the pretest. A dummy variable method was used to address missing values for covariates. This method involved replacing the unobserved value with a constant and, for each covariate, adding an indicator variable with a value of either one or zero to signify whether the value of the corresponding covariate is observed or unobserved (missing).

Summary of attrition in SAT reading sample

The starting sample included 41 AMSTI schools (with 12,065 students) and 41 control schools (with 10,492 students). The baseline sample for the analysis of mathematics outcomes, which was limited to students without disabilities, consisted of 41 AMSTI schools (with 10,517 students) and 41 control schools (with 9,109 students). The analytic sample for the analysis of reading outcomes consisted of 41 AMSTI schools (with 10,019 students) and 41 control schools (with 8,691 students).

Level of teacher content knowledge in mathematics

Below are changes in the sample from the point at which completed teacher surveys were collected to the point at which the analytic sample was identified (table N2).

Table N2 Attrition from analytical sample associated with level of teacher content knowledge in mathematics outcome

	AMSTI		Control	
Item	**Number of schools**	**Number of teachers**	**Number of schools**	**Number of teachers**
Completed surveys (baseline sample)	41	221	40	205
Loss because of missing teacher identifier	0	0	0	0
Available cases	41	221	40	205
Loss because of missing school identifier	0	0	0	0
Available cases	41	221	40	205
Loss because of missing valid teacher content knowledge in mathematics rating	0	–24	0	–18
Number of cases in sample	41	197	40	187

Source: Teacher survey data.

Survey completion

As described in the data collection section, teachers were asked to consent to participate in online surveys. Teachers who consented were asked to complete four monthly surveys between January and April of Year 1. Data from the teacher content knowledge outcome came from the fourth survey.

The eligible sample was first limited to teachers who taught the appropriate grades and subjects (in non-special education classes in grades 4–8) who completed the survey. In AMSTI schools, 221 teachers (in 41 schools) completed the survey. In control schools, 205 teachers (in 40 schools) completed survey.

Exclusion of data on students without valid identifiers

All teachers in both conditions had valid teacher and schools identifiers.

Exclusion of data on because of missing teacher content knowledge rating

If teachers did not complete the fourth survey, did not complete the question on teacher content knowledge, or selected the "not applicable" option on the teacher content knowledge question, data were considered missing. Data on 24 teachers in AMSTI schools and 18 teachers in the control schools were removed from this analysis because of missing valid data.[113] No schools were eliminated for this reason.

Summary of attrition in teacher content knowledge sample

The baseline sample for the analysis of teacher content knowledge in mathematics, which was limited to teachers who completed the survey, consisted of 41 AMSTI schools (with 221 teachers) and 40 control schools (with 205 teachers). The analytic sample for the analysis of teacher content knowledge in mathematics consisted of 41 AMSTI schools (with 197 teachers) and 40 control schools (with 187 teachers).

[113] Data from one control mathematics teacher were removed because the teacher selected "not applicable."

Level of teacher content knowledge in science

Below are changes in the sample from the point at which completed surveys from the teachers were collected to the point at which the analytic sample was identified (table N3).

Table N3 Attrition from analytical sample associated with level of teacher content knowledge in science outcome

	AMSTI		Control	
Item	**Number of schools**	**Number of teachers**	**Number of schools**	**Number of teachers**
Completed surveys (baseline sample)	40	203	40	192
Loss because of missing teacher identifier	0	0	0	0
Available cases	40	203	40	192
Loss because of missing school identifier	0	0	0	0
Available cases	40	203	40	192
Loss because of missing valid teacher content knowledge in science rating	0	–27	–1	–22
Number of cases in sample	40	176	39	170

Source: Teacher survey data.

Survey completion

As described in the data collection section, teachers were asked to consent to participate in online surveys. Teachers who consented were asked to complete four monthly surveys between January and April of Year 1. Data from the teacher content knowledge outcome came from the fourth survey.

The eligible sample was first limited to teachers who taught the appropriate grades and subjects (in non-special education classes in grades 4–8) who completed surveys. In AMSTI schools, 203 teachers (in 40 schools) completed the survey. In control schools, 192 teachers (in 40 schools) completed the survey.

Exclusion of data on students without valid identifiers

All teachers in both conditions had valid teacher and schools identifiers.

Exclusion of data because of missing data on teacher content knowledge

If teachers did not complete the fourth survey, did not complete the question on teacher content knowledge, or selected the "not applicable" option from this survey question, data were considered missing. Data on 27 teachers in AMSTI schools and 22 teachers in control schools were removed from this analysis because of missing valid data.[114] One control school was also removed for this reason.

Summary of attrition in teacher content in science sample

The baseline sample for the analysis of teacher content knowledge in science outcome, which was limited to teachers who completed the survey, consisted of 40 AMSTI schools (with 203 teachers) and 40 control schools (with 192 teachers). The analytic sample for the analysis of teacher content knowledge in science outcomes consisted of 40 AMSTI schools (with 176 teachers) and 39 control schools (with 170 teachers).

Level of student engagement in mathematics

Below are changes in the sample from the point at which completed surveys from teachers were collected to the point at which the analytic sample was identified (table N4).

Table N4. Attrition from analytical sample associated with level of student engagement in mathematics outcome

	AMSTI		Control	
Item	**Number of schools**	**Number of teachers**	**Number of schools**	**Number of teachers**
Completed surveys (baseline sample)	41	221	40	205
Loss because of missing teacher identifier	0	0	0	0
Available cases	41	221	40	205
Loss because of missing school identifier	0	0	0	0
Available cases	41	221	40	205
Loss because of missing valid rating of student engagement in mathematics classrooms	0	–24	0	–17
Number of cases in sample	41	197	40	188

Source: Teacher survey data.

Survey completion

As described in the data collection section, teachers were asked to consent to participate in online surveys. Teachers who consented were asked to complete four monthly surveys

[114] Data from two control science teachers were removed due to selecting "not applicable."

between January and April of Year 1. Data from the student engagement outcome came from the fourth survey.

The eligible sample was first limited to teachers who taught the appropriate grades and subjects (from non-special education classes in grades 4–8) who completed the survey. In AMSTI schools, 221 teachers (in 41 schools) completed the survey. In control schools, 205 teachers (in 40 schools) completed the survey.

Exclusion of data on students without valid identifiers

All teachers in both conditions had valid teacher and schools identifiers.

Exclusion because of missing data on student engagement

If teachers did not complete the fourth survey or did not complete the question on teacher content knowledge, their data were considered missing. Data from 24 teachers in AMSTI schools and 17 teachers in control schools were removed from this analysis because of missing valid data. No schools were removed for this reason.

Summary of attrition in student engagement sample

The baseline sample for the analysis of student engagement in mathematics classroom outcome, which was limited to teachers who completed the survey, consisted of 41 AMSTI schools (with 221 teachers) and 40 control schools (with 205 teachers). The analytic sample for the analysis of student engagement in mathematics outcomes consisted of 41 AMSTI schools (with 197 teachers) and 40 control schools (with 188 teachers).

Level of student engagement in science attrition

Below are changes in the sample from the point at which completed surveys from teachers were returned to the point at which the analytic sample was identified (table N5).

Table N5. Attrition from analytical sample associated with level of student engagement in science outcome

	AMSTI		Control	
Item	**Number of schools**	**Number of teachers**	**Number of schools**	**Number of teachers**
Completed surveys (baseline sample)	40	203	40	192
Loss because of missing teacher identifier	0	0	0	0
Available cases	40	203	40	192
Loss because of missing school identifier	0	0	0	0
Available cases	40	203	40	192
Loss because of missing valid rating of student engagement in science classrooms	0	–27	0	–20
Number of cases in sample	40	176	40	172

Source: Teacher survey data.

Survey completion

As described in the data collection section, teachers were asked to consent to participate in online surveys. Teachers who consented were asked to complete four monthly surveys between January and April of Year 1. Data on the student engagement outcome came from the fourth survey.

The eligible sample was first limited to teachers who taught the appropriate grades and subjects (in non-special education classes in grades 4–8) who completed the survey. In AMSTI schools, 203 teachers (in 40 schools) completed the survey. In control schools, 192 teachers (in 40 schools) completed survey.

Exclusion of data on students without valid identifiers

All teachers in both conditions had valid teacher and schools identifiers.

Exclusion because of missing data on student engagement

If teachers did not complete the fourth survey, did not complete the question on teacher content knowledge, or selected the “not applicable” option from this survey question, data were considered missing. Data on 27 teachers in AMSTI schools and 20 teachers in control schools were removed from this analysis because of missing valid data. No schools were eliminated for this reason.

Summary of attrition in student engagement in science sample

The baseline sample for the analysis of teacher content knowledge in science outcome, which was limited to teachers who completed surveys, consisted of 40 AMSTI schools (with 203 teachers) and 40 control schools (with 192 teachers). The analytic sample for the analysis of student engagement in science outcomes consisted of 40 AMSTI schools (with 176 teachers) and 40 control schools (with 172 teachers).

Appendix O. Tests of equivalence for baseline and analytic samples for Year 1 exploratory outcomes

Table O1 Year 1 mean baseline sample characteristics associated with reading achievement outcome after one year

Characteristic		**AMSTI schools**		**Control schools**
Student Characteristic		**(*n*=41)**		**(*n*=41)**
Average of school percent of boys	Average in Each Condition	49.4		49.6
	Standard Deviation	4.1		4.1
	Average Difference		–0.2	
	Standard Error		0.9	
	Test statistic		t = 0.25	
	***p*-value**		.80	
Average of school percent of minority students	Average in Each Condition	51.0		46.9
	Standard Deviation	34.6		33.2
	Average Difference		4.1	
	Standard Error		7.5	
	Test statistic		*t* = 0.55	
	***p*-value**		.58	
Average of school percent of students proficient in English	Average in Each Condition	98.3		98.6
	Standard Deviation	3.3		2.4
	Average Difference		–0.2	
	Standard Error		0.6	
	Test statistic		*t* = 0.36	
	***p*-value**		.72	
Average of school percent of students enrolled in the free or reduced-price lunch program	Average in Each Condition	63.2		64.7
	Standard Deviation	24.8		24.2
	Average Difference		–1.5	
	Standard Error		5.4	
	Test statistic		*t* = 0.28	
	***p*-value**		.78	
Average of school percent of students in grade 4	Average in Each Condition	25.7		22.4
	Standard Deviation	24.7		20.7
	Average Difference		3.3	
	Standard Error		5.0	
	Test statistic		*t* = 0.65	
	***p*-value**		.52	
Average of school percent of students in grade 5	Average in Each Condition	25.0		23.3
	Standard Deviation	18.7		19.5
	Average Difference		1.7	
	Standard Error		4.2	
	Test statistic		t = 0.41	
	***p*-value**		.68	
Average of school percent of students in grade 6	Average in Each Condition	14.3		18.8
	Standard Deviation	14.4		16.6
	Average Difference		–4.5	
	Standard Error		3.4	
	Test statistic		*t* = 1.32	
	***p*-value**		.19	
Average of school percent of students in	Average in Each Condition	16.7		18.1
	Standard Deviation	19.0		20.6
	Average Difference		–1.4	

Characteristic		AMSTI schools		Control schools
grade 7	Standard Error		4.4	
	Test statistic		t = 0.32	
	***p*-value**		.75	
Average of school percent of students in grade 8	Average in Each Condition	18.4		17.4
	Standard Deviation	22.2		19.2
	Average Difference		0.9	
	Standard Error		4.6	
	Test statistic		t = 0.20	
	***p*-value**		.84	
School Average pretest score and sample size				
SAT 10[a] area	Average in Each Condition	645.1		648.1
	Standard Deviation	21.3		17.0
	Average Difference		–3.0	
	Standard Error		4.3	
	Test statistic		t = 0.71	
	***p*-value**		.48	
Sample size	**AMSTI schools** Number of schools = 41 Number of teachers = 231 Number of students = 10,517		**Control schools** Number of schools = 41 Number of teachers = 210 Number of students = 9,109	

Note: Detail may not sum to totals because of rounding. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Grade level for reading is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of students across grades for the reading outcome (p = .76). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the p-value was .07 for reading).

a. The SAT 10 reading pretest was used for the reading outcome.

Source: Student achievement data from tests administered as part of the state's accountability system, student demographic data from state data system, and teacher survey data.

Table O2 Year 1 mean baseline sample characteristics associated with level of teacher content knowledge outcomes after one year

		Teacher content knowledge in mathematics			Teacher content knowledge in science		
Characteristic		**AMSTI schools (*n*=41)**		**Control schools (*n*=40)**	**AMSTI schools (*n*=40)**		**Control schools (*n*=40)**
Student Characteristic							
Average of school percent of out-of-field teachers	Average in Each Condition	19.8		26.1	12.5		20.3
	Standard Deviation	23.1		29.1	18.1		24.7
	Average Difference		–6.2			–7.8	
	Standard Error		5.8			4.8	
	Test statistic		t = 1.07			t = 1.61	
	***p*-value**		.29			.11	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	56.4		56.9	60.1		55.3
	Standard Deviation	29.0		30.9	29.7		34.1
	Average Difference		–0.5			4.8	
	Standard Error		6.7			7.2	
	Test statistic		*t* = 0.08			*t* = 0.67	
	***p*-value**		.94			.50	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	23.8		17.1	27.3		24.4
	Standard Deviation	21.7		22.3	28.5		28.1
	Average Difference		6.8			3.0	
	Standard Error		4.9			6.3	
	Test statistic		*t* = 1.38			*t* = 0.47	
	***p*-value**		.17			.64	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	26.6		27.0	27.9		30.1
	Standard Deviation	26.3		26.9	29.8		25.5
	Average Difference		–0.5			–2.2	
	Standard Error		5.9			6.2	
	Test statistic		*t* = 0.08			*t* = 0.36	
	***p*-value**		.94			.72	
Average of school percent of teachers with less than four years' teaching experience in subject area	Average in Each Condition	29.4		31.3	31.0		32.7
	Standard Deviation	26.9		26.3	29.5		23.3
	Average Difference		–1.9			–1.7	
	Standard Error		0.3			5.9	
	Test statistic		*t* = 5.90			*t* = 0.29	
	***p*-value**		.75			.77	

		Teacher content knowledge in mathematics			Teacher content knowledge in science		
Characteristic		AMSTI schools		Control schools	AMSTI schools		Control schools
Sample size		Number of schools = 41 Number of teachers = 221		Number of schools = 40 Number of teachers = 205	Number of schools = 40 Number of teachers = 203		Number of schools = 40 Number of teachers = 192

Note: Detail may not sum to totals because of rounding. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Teacher degree rank is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of teacher degree rank (p = .07 for mathematics, p = .23 for science). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the p-value was .28 for teacher content knowledge in mathematics and .09 for teacher content knowledge in science.)

Source: Teacher survey data.

Table O3 Year 1 mean baseline sample characteristics associated with level of student engagement outcomes after one year

		Student engagement in mathematics			Student engagement in science		
Characteristic		**AMSTI schools (*n*=41)**		**Control schools (*n*=40)**	**AMSTI schools (*n*=40)**		**Control schools (*n*=40)**
Average of school percent of out-of-field teachers	Average in Each Condition	19.8		26.0	12.5		20.3
	Standard Deviation	23.1		29.1	18.1		24.7
	Average Difference		–6.2			–7.8	
	Standard Error		5.8			4.8	
	Test statistic		t = 1.07			t = 1.61	
	***p*-value**		.29			.11	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	56.4		56.9	60.1		55.3
	Standard Deviation	29.0		30.9	29.7		34.1
	Average Difference		–0.5			4.8	
	Standard Error		6.7			7.2	
	Test statistic		*t* = 0.08			*t* = 0.67	
	***p*-value**		.94			.50	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	23.8		17.1	27.3		14.4
	Standard Deviation	21.7		22.3	28.5		28.1
	Average Difference		6.8			3.0	
	Standard Error		4.9			6.3	
	Test statistic		*t* = 1.38			*t* = 0.47	
	***p*-value**		.17			.64	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	26.6		27.0	27.9		30.1
	Standard Deviation	26.3		26.9	29.8		25.5
	Average Difference		–0.5			–2.2	
	Standard Error		0.1			6.2	
	Test statistic		*t* =5.91			*t* = 0.36	
	***p*-value**		.94			.72	
Average of school percent of teachers with less than four years' teaching experience in subject area	Average in Each Condition	29.4		31.3	31.0		32.7
	Standard Deviation	26.9		26.3	29.5		23.3
	Average Difference		–1.9			–1.7	
	Standard Error		5.9			5.9	
	Test statistic		*t* = 0.32			*t* = 0.29	
	***p*-value**		.75			.77	

		Student engagement in mathematics			Student engagement in science		
Characteristic		AMSTI schools		Control schools	AMSTI schools		Control schools
Sample size		Number of schools = 41 Number of teachers = 221		Number of schools = 40 Number of teachers = 205	Number of schools = 40 Number of teachers = 203		Number of schools = 40 Number of teachers = 192

Note: Detail may not sum to totals because of rounding. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Teacher degree rank is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of teacher degree rank ($p = .07$ for mathematics, $p = .23$ for science). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the *p*-value was .28 for student engagement in mathematics and .09 for student engagement in science.)

Source: Teacher survey data.

Table O4 Year 1 mean analytic sample characteristics associated with reading achievement outcome after one year

Characteristic		AMSTI schools		Control schools
Student Characteristic		**(*n*=41)**		**(*n*=41)**
Average of school percent of boys	Average in Each Condition	49.3		49.4
	Standard Deviation	4.2		4.0
	Average Difference		–0.1	
	Standard Error		0.9	
	Test statistic		t = 0.11	
	***p*-value**		.91	
Average of school percent of minority students	Average in Each Condition	51.2		46.9
	Standard Deviation	34.6		33.4
	Average Difference		4.3	
	Standard Error		7.5	
	Test statistic		*t* = 0.57	
	***p*-value**		.57	
Average of school percent of students proficient in English	Average in Each Condition	98.3		98.7
	Standard Deviation	3.3		2.2
	Average Difference		–0.3	
	Standard Error		0.6	
	Test statistic		*t* = 0.53	
	***p*-value**		.60	
Average of school percent of students enrolled in the free or reduced-price lunch program	Average in Each Condition	63.2		64.6
	Standard Deviation	24.9		24.3
	Average Difference		–1.4	
	Standard Error		5.4	
	Test statistic		*t* = 0.26	
	***p*-value**		.80	
Average of school percent of students in grade 4	Average in Each Condition	25.8		22.5
	Standard Deviation	24.8		20.7
	Average Difference		3.3	
	Standard Error		5.0	
	Test statistic		*t* = 0.65	
	***p*-value**		.52	
Average of school percent of students in grade 5	Average in Each Condition	25.3		23.3
	Standard Deviation	18.7		19.7
	Average Difference		2.0	
	Standard Error		4.2	
	Test statistic		t = 0.47	
	***p*-value**		.64	
Average of school percent of students in grade 6	Average in Each Condition	14.5		18.8
	Standard Deviation	14.5		16.6
	Average Difference		–4.4	
	Standard Error		3.4	
	Test statistic		*t* = 1.27	
	***p*-value**		.21	
Average of school percent of students in grade 7	Average in Each Condition	16.8		18.2
	Standard Deviation	19.0		20.6
	Average Difference		–1.4	
	Standard Error		4.4	
	Test statistic		*t* = 0.31	
	***p*-value**		.76	

Characteristic		AMSTI schools		Control schools
Average of school percent of students in grade 8	Average in Each Condition	17.7		17.2
	Standard Deviation	22.4		19.4
	Average Difference		0.5	
	Standard Error		4.6	
	Test statistic		t = 0.11	
	***p*-value**		.91	
School Average pretest score and sample size				
SAT 10[a] area	Average in Each Condition	645.1		648.3
	Standard Deviation	21.2		17.0
	Average Difference		–3.2	
	Standard Error		4.2	
	Test statistic		t = 0.76	
	***p*-value**		.45	
Sample size	**AMSTI schools** Number of schools = 41 Number of teachers = 231 Number of students = 10,019		**Control schools** Number of schools = 41 Number of teachers = 210 Number of students = 8,691	

Note: Detail may not sum to totals because of rounding. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Grade level for reading is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of students across grades for the reading outcome (p = .73). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the p-value was .054 for reading).

a. The SAT 10 reading pretest was used for the reading outcome.

Source: Student achievement data from tests administered as part of the state's accountability system, student demographic data from state data system, and teacher survey data.

Table O5 Year 1 mean analytic sample characteristics associated with level of teacher content knowledge outcomes after one year

		Teacher content knowledge in mathematics			Teacher content knowledge in science		
Characteristic		AMSTI schools (*n*=41)		Control schools (*n*=40)	AMSTI schools (*n*=40)		Control schools (*n*=39)
Average of school percent of out-of-field teachers	Average in Each Condition	16.5		25.0	12.7		20.6
	Standard Deviation	22.4		29.3	19.2		27.4
	Average Difference		–8.5			–7.9	
	Standard Error		5.8			5.3	
	Test statistic		t = 1.46			t = 1.49	
	***p*-value**		.15			.14	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	55.6		58.3	60.4		54.8
	Standard Deviation	29.8		32.1	32.2		34.0
	Average Difference		–2.7			5.6	
	Standard Error		6.9			7.4	
	Test statistic		*t* = 0.40			*t* = 0.76	
	***p*-value**		.69			.45	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	27.9		16.7	26.9		24.6
	Standard Deviation	28.8		22.2	29.2		30.2
	Average Difference		11.2			2.3	
	Standard Error		5.7			6.7	
	Test statistic		*t* = 1.96			*t* = 0.34	
	***p*-value**		.053			.74	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	27.3		28.3	28.2		27.1
	Standard Deviation	29.3		28.5	32.1		24.7
	Average Difference		–1.0			1.1	
	Standard Error		6.4			6.5	
	Test statistic		*t* = 0.16			*t* = 0.16	
	***p*-value**		.87			.87	
Average of school percent of teachers with less than four years' teaching experience in subject area	Average in Each Condition	28.1		30.9	30.4		33.9
	Standard Deviation	29.6		28.0	31.8		28.1
	Average Difference		–2.8			–3.5	
	Standard Error		6.4			6.8	
	Test statistic		*t* = 0.44			*t* = 0.52	
	***p*-value**		.66			.61	

		Teacher content knowledge in mathematics			Teacher content knowledge in science		
Characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
Sample size		Number of schools = 41 Number of teachers = 197		Number of schools = 40 Number of teachers = 187	Number of schools = 40 Number of teachers = 176		Number of schools = 39 Number of teachers = 170

Note: Detail may not sum to totals because of rounding. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Teacher degree rank is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of teacher degree rank (p = .04 for mathematics, p = .23 for science). For the analysis associated with teacher content knowledge in mathematics, the number of teachers responding in each category were 28 with no degree, 113 with one degree, and 52 with more than one degree in the content area for AMSTI teachers and 39 with no degree, 104 with one degree, and 31 with more than one degree in the content area for control teachers. They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the p-value was .504 for teacher content knowledge in mathematics and .29 for teacher content knowledge in science.)

Source: Teacher survey data

Table O6 Year 1 mean analytic sample characteristics associated with level of student engagement outcomes after one year

		Teacher content knowledge in mathematics			Teacher content knowledge in science		
Characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
		(*n*=41)		**(*n*=40)**	**(*n*=40)**		**(*n*=40)**
Average of school percent of out-of-field teachers	Average in Each Condition	16.5		25.3	12.7		20.5
	Standard Deviation	22.4		29.4	19.2		27.0
	Average Difference		–8.8			–7.9	
	Standard Error		5.8			5.2	
	Test statistic		t = 1.53			t = 1.50	
	***p*-value**		.13			.14	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	55.6		58.1	60.4		55.6
	Standard Deviation	29.8		32.2	32.2		34.1
	Average Difference		–2.5			4.9	
	Standard Error		6.9			7.4	
	Test statistic		t = 0.36			t = 0.66	
	***p*-value**		.72			.51	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	27.9		16.6	26.9		23.9
	Standard Deviation	28.8		22.1	29.2		30.0
	Average Difference		11.3			3.0	
	Standard Error		5.7			6.6	
	Test statistic		t = 1.98			t = 0.45	
	***p*-value**		.051			.65	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	27.3		28.2	28.2		28.7
	Standard Deviation	29.3		28.5	32.1		26.8
	Average Difference		–0.9			–0.5	
	Standard Error		6.4			6.6	
	Test statistic		t = 0.14			t = 0.08	
	***p*-value**		.89			.94	
Average of school percent of teachers with less than four years' teaching experience in subject area	Average in Each Condition	28.1		31.3	30.4		32.8
	Standard Deviation	29.6		28.0	31.8		28.2
	Average Difference		–3.2			–2.4	
	Standard Error		6.4			6.7	
	Test statistic		t = 0.50			t = 0.36	
	***p*-value**		.62			.72	

		Teacher content knowledge in mathematics			Teacher content knowledge in science		
Characteristic		AMSTI schools		Control schools	AMSTI schools		Control schools
Sample size		Number of schools = 41 Number of teachers = 197		Number of schools = 40 Number of teachers = 188	Number of schools = 40 Number of teachers = 176		Number of schools = 40 Number of teachers = 172

Note: Detail may not sum to totals because of rounding. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Teacher degree rank is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of teacher degree rank ($p = .04$ for mathematics, $p = .19$ for science).

For the analysis associated with student engagement in mathematics, the number of teachers responding in each category were 28 with no degree, 113 with one degree, and 52 with more than one degree in the content area for the AMSTI teachers and 40 with no degree, 104 with one degree, and 31 with more than one degree in the content area for control teachers. They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the *p*-value was .04 for student engagement in mathematics and .17 for student engagement in science).

Source: Teacher survey data.

Appendix P. Statistical power analyses for moderator analyses

This appendix describes the statistical power analyses for the exploratory analyses of the moderating effects of specific student-level covariates. The minimum detectable effect size can be expressed as a constant, *factor* (α, β, df), multiplied by the standard error of the impact estimate and divided by the standard deviation of the outcome measure (Bloom 2005; Schochet 2008):

$$MDES(\hat{\beta}_T) = Factor(\alpha, \beta, df) * \frac{\sqrt{Variance(IMPACT)}}{\sigma} \qquad \text{(P1)}$$

where

- $Factor(\alpha, \beta, df)$ is a constant that is a function of the significance level (α), statistical power (β), and the number of degrees of freedom (*df*).[115]
- $\sqrt{Variance(IMPACT)}$ is the standard error of the program impact.
- σ is the standard deviation of the outcome measure.

In calculating the standard error of the program impact, it is necessary to figure in the cluster randomized design. In the experimental design, students were nested in schools and schools were randomized to conditions. The standard error of the impact estimate can be expressed approximately as

$$s.e.(\hat{IMPACT}) = \sqrt{\frac{\tau^2(1-R_c^2)}{P(1-P)J} + \frac{\sigma^2(1-R_i^2)}{P(1-P)nJ}} \qquad \text{(P2)}$$

where

- τ^2 is the school-level variance in the outcome.
- σ^2 is the student-level variance in the outcome.
- R_I^2 is the proportion of the random variance within schools that is reduced by the covariates (their student-level explanatory power).

[115] This study used a matched-pairs design. The degrees of freedom available for estimating the impact are equal to the number of clusters/2 – 1 (Bloom 2005). In this study, with 82 clusters (schools) and 2 treatment conditions (intervention and control), there were 82/2 –1 = 40 degrees of freedom. For a multiplier for alpha of .05, a power of .80, and 40 degrees of freedom, the multiplier for the minimum detectable effect size in this study is 2.87 (Schochet 2008). (The actual degrees of freedom for estimating differential effects will depend on the number of schools with the relevant subgroups.)

- R_C^2 is the proportion of the random variance between schools that is reduced by the covariates (their school-level explanatory power).
- J is the number of schools.
- P is the proportion of schools assigned to the treatment condition.
- n is the average number of students sampled from each school.

Equation Q3 (which assumes that the impacts for subgroup A and subgroup B are independent quantities) was used to compute the standard error of the estimate of the difference in the impact of AMSTI for the two subgroups:

$$s.e.(\hat{IMPACT}_{A-B}) = \sqrt{Var(\hat{IMPACT}_A) + Var(\hat{IMPACT}_B)} \qquad \text{(P3)}$$

where

- $s.e.(\hat{IMPACT}_{A-B})$ is the estimated standard error for the estimate of the program impact for subgroup A minus the program impact for subgroup B (that is, this is the standard error for the estimate of the interaction between the indicator of treatment status and the subgroup indicator).
- $Var(\hat{IMPACT}_A)$ is the variance of the impact estimate for subgroup A.
- $Var(\hat{IMPACT}_B)$ is the variance of the impact estimate for subgroup B.

The standard error was substituted for the estimate of the differential impact into equation (1) to compute the minimum detectable differential effect size in the impact for subgroup A compared with subgroup B:

$$MDD(\hat{IMPACT}_{A-B}) = Factor(\alpha, \beta, df) * s.e.(\hat{IMPACT}_{A-B}) / \sigma .$$

The confirmatory impact analysis yielded estimates of the critical parameters needed to compute an estimate of this value. The sample data were used to estimate the minimum detectable differential effect size values for the moderator analyses. (Because the reading outcomes were not analyzed at the time the power analysis was conducted, the sample statistics were not available to estimate minimum detectable differential effect sizes for this outcome. Sample sizes and parameter values were expected to be similar for reading and math problem solving; the results for math problem solving were therefore used as a guide for expected power for reading.)

Below is the specification of the covariates (including moderators) for the five moderator analyses for the mathematics problem solving outcome (the moderator analyses for science and reading outcomes follow a similar scheme [table P1].) Listwise deletion rather than the dummy variable method was used to handle missing values for the moderator in a given analysis; a dummy variable was therefore not used for the corresponding moderator in each analysis. For consistency with the benchmark model for estimating the impact on mathematics problem solving, the reading pretest was not modeled, except in the analysis of its moderating effect.

Table P1 Specification of covariates used in moderator analyses for mathematics problem solving outcome

	Analysis of the moderating effect of				
Covariate	**Mathematics problem solving pretest**	**Racial/ethnic minority status**	**Free or reduced-priced lunch status**	**Gender**	**Reading pretest**
(1) Student-level pretest score in mathematics problem solving	MOD	C	C	C	N
(2) Racial/ethnic minority status (coded 0 for White, 1 for minority)	C	MOD	C	C	C
(3) Free or reduced-price lunch status (coded 0 if student enrolled in the free or reduced-price lunch program, 1 otherwise)	C	C	MOD	C	C
(4) Gender (coded 0 for girls, 1 for boys)	C	C	C	MOD	C
(5) Student-level pretest score in reading	N	N	N	N	MOD
(6) Proficient in English (coded 0 if student proficient in English, 1 otherwise)	C	C	C	C	C
(7) Member of grade 4 (coded 1 if student belongs to grade 4, 0 otherwise)	C	C	C	C	C
(8) Member of grade 5 (coded 1 if student belongs to grade 5, 0 otherwise)	C	C	C	C	C
(9) Member of grade 7 (coded 1 if the student belongs to grade 7, 0 otherwise)	C	C	C	C	C
(10) Member of grade 8 (coded 1 if student belongs to grade 8, 0 otherwise)	C	C	C	C	C
(11) Dummy variable indicating missing value for (1)	N	C	C	C	N
(12) Dummy variable indicating missing value for (2)	C	N	C	C	C
(13) Dummy variable indicating missing value for (3)	C	C	N	C	C

	Analysis of the moderating effect of				
Covariate	**Mathematics problem solving pretest**	**Racial/ethnic minority status**	**Free or reduced-priced lunch status**	**Gender**	**Reading pretest**
(14) Dummy variable indicating missing value for (4)	C	C	C	N	C
(15) Dummy variable indicating missing value for (6)	C	C	C	C	C
(16) Dummy variable indicating missing value for (7)	C	C	C	C	C
(17) Dummy variable indicating missing value for (8)	C	C	C	C	C
(18) Dummy variable indicating missing value for (9)	C	C	C	C	C
(19) Dummy variable indicating missing value for (10)	C	C	C	C	C

Note: MOD is moderator. C is a covariate other than the moderator. N is not modeled.

Shown below are the achieved values of critical characteristics for the power analysis for estimating differential impacts for the mathematics problem solving outcome (table P2). The results are used as a guide for the expected power for the corresponding analyses for the reading outcome. Table P3 shows the same results for the science outcome.

Table P2 Sample-based assumptions used in estimating power for moderator analyses of mathematics problem solving outcome

Item	Gender		Free or reduced-priced lunch program		Racial/ethnic minority		Mathematics problem solving pretest[b]		Reading pretest[b]	
	Boys	Girls	Not enrolled	Enrolled	Yes	No	High	Low	High	Low
School-level variance (τ^2)	467.0	362.0	354.4	302.5	278.4	382.6	177.2	145.1	189.3	158.9
Student-level variance (σ^2)	1,572.4	1,296.0	1,614.7	1,145.4	1,159.4	1,4967.0	1,170.7	695.1	1,235.2	815.3
Unconditional intraclass correlation	.23	.22	.18	.21	.19	.20	.13	.17	.13	.16
School-level R^2 (R^2_C)	.97	.96	.96	.95	.95	.96	.87	.89	.81	.89
Student-level R^2 (R^2_I)	.57	.59	.59	.52	.55	.60	.08	.17	.07	.14
Number of schools (J)	82	82	81	82	82	74	82	82	82	82
Number of students/school (n)	112	116	103	127	94	140	109	102	106	105
Proportion treatment schools (P)	.50	.50	.49	.50	.50	.50	.50	.50	.50	.50
Proportion control schools ($1 - P$)	.50	.50	.51	.50	.50	.50	.50	.50	.50	.50
Control group standard deviation (σ)	44.2	40.6	45.0	35.8	36.7	44.1	37.8	27.0	39.2	29.3
Statistical power (β)	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80
Alpha level (α)	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
Degrees of freedom (df)	40	40	38	40	40	27	40	40	40	40
Factor (α, β, df)[a]	2.87	2.87	2.87	2.87	2.87	2.92	2.87	2.87	2.87	2.87
Standard error (impact)	1.0	1.0	1.0	1.0	1.0	1.0	1.3	1.0	1.5	1.1
Standard error (differential)	1.4	na	1.4	na	1.4	na	1.6	na	1.9	na
Minimum detectable differential effect size	0.09	na	0.10	na	0.10	na	0.14	na	0.16	na

na is not applicable.

a. The value for *Factor (α, β, df)* comes from table 1 of Schochet (2008).

b. For the purpose of the power analysis, the pretest was dichotomized into values above the median (high) and values below the median (low). In the moderator analyses, the pretest was divided into three categories: low, for scores in stanines 1–3; middle, for scores in stanines 4–6; and high, for scores in stanines 7–9. Dividing the pretest into three levels instead of two should increase statistical power; the results presented here can therefore be considered conservative. The cutpoints for the stanines were based on the pretest scale scores for the sample.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table P3 Sample-based assumptions used in estimating power for moderator analyses of science outcome

Item	Gender		Free or reduced-priced lunch program		Racial/ethnic minority		Reading pretest[b]	
	Boys	Girls	Not enrolled	Enrolled	Yes	No	High	Low
School-level variance (τ^2)	264.0	195.5	128.8	147.4	136.1	149.1	122.9	94.0
Student-level variance (σ^2)	951.6	688.5	842.1	718.5	646.0	845.6	645.9	561.8
Unconditional intraclass correlation	.22	.22	.13	.17	.17	.15	.16	.14
School-level R^2 (R^2_C)	.90	.89	.85	.83	.82	.86	.79	.55
Student-level R^2 (R^2_I)	.44	.43	.42	.38	.39	.44	.07	.06
Number of schools (J)	79	79	74	79	78	66	79	79
Number of students/school (n)	47	48	44	54	41	61	44	44
Proportion treatment schools (P)	.49	.49	.49	.49	.49	.48	.49	.49
Proportion control schools ($1 - P$)	.51	.51	.51	.51	.51	.52	.51	.51
Control group standard deviation (σ)	33.8	28.6	30.8	28.8	28.1	31.1	26.7	24.7
Statistical power (β)	.80	.80	.80	.80	.80	.80	.80	.80
Alpha level (α)	.05	.05	.05	.05	.05	.05	.05	.05
Degrees of freedom (df)	34	34	27	34	35	20	34	34
Factor (α, β, df)[a]	2.89	2.89	2.92	2.89	2.89	2.95	2.89	2.89
Standard error (impact)	1.4	1.2	1.3	1.3	1.3	1.3	1.4	1.7
Standard error (differential)	1.9	na	1.8	na	1.9	na	2.8	na
Minimum detectable differential effect size	.17	na	.18	na	.18	na	.25	na

na is not applicable.

a. The value for *Factor (α, β, df)* comes from table 1 of Schochet (2008).

b. For the purpose of the power analysis, the pretest was dichotomized into values above the median (high) and values below the median (low). In the moderator analyses, the pretest was divided into three categories: low, for scores in stanines 1–3; middle, for scores in stanines 4–6; and high, for scores in stanines 7–9. Dividing the pretest into three levels instead of two should increase statistical power; the results presented here can therefore be considered conservative. The cutpoints for the stanines were based on the pretest scale scores for the sample.

Source: S achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix Q. Derivation and motivation of the Bell-Bradley estimator when measuring estimated two-year effect of the Alabama Math, Science, and Technology Initiative (AMSTI)

Bell and Bradley (2008) proposed approximating the "untreated" control group outcome level for grade *g* in Year 2 with Y_{2g}^{C*}, defined as:

$$Y_{2g}^{C*} = Y_{2g}^{C} - (Y_{1g}^{T} - Y_{1g}^{C}) \qquad \text{(Q1)}$$

where

- Y_{2g}^{C*} = approximate untreated control mean outcome level in Year 2 for grade *g*,
- Y_{2g}^{C} = observed control mean outcome in Year 2 for grade *g*,
- Y_{1g}^{T} = observed treatment group mean outcome in Year 1 for grade *g*, and
- Y_{1g}^{C} = observed control mean outcome in Year 1 for grade *g*.

Using Y_{2g}^{C*}, one can approximate the effect of two years of AMSTI on the treatment group, compared with no AMSTI, using the experimental structure of the data with the following equation:

$$F_{2g}^{T*} = Y_{2g}^{T} - Y_{2g}^{C*} = Y_{2g}^{T} - Y_{2g}^{C} + Y_{1g}^{T} - Y_{1g}^{C} \qquad \text{(Q2)}$$

where F_{2g}^{T*} = approximate mean effect of two years of AMSTI on the treatment group compared with no AMSTI in Year 2 for grade *g* and Y_{2g}^{T} = observed treatment group mean outcome in Year 2 for grade *g*.

The estimated mean effect of one year of AMSTI on the treatment group compared with no AMSTI for grade g in Year 1 can be defined as

$$F_{-g}^{T} = Y_{1g}^{T} - Y_{1g}^{C}.$$

The estimated mean effect of two years of AMSTI on the treatment group compared with one year of AMSTI, in Year 2 for grade *g*, can be defined as

$$F_{2g}^{T} = Y_{2g}^{T} - Y_{2g}^{C}.$$

Substituting both equations into equation (R2) yields

$$F_{2g}^{T*} = \mathrm{F}_{2g}^{T} + F_{1g}^{T}. \qquad \text{(Q3)}$$

The Bell-Bradley two-year effect estimate is thus the sum of consecutive years' effect estimates on the treatment group students at a fixed grade level, *g*. Intuitively, the second part of this sum, F_{1g}^{T}, is an unbiased estimate of the impact of AMSTI on students in their first year of program exposure; the first part of the sum, F_{2g}^{T}, adds an unbiased estimate of the incremental effect of the second year of AMSTI over and above the impact of the first year. F_{2g}^{T} estimates the incremental effect of a second year because the control group portion of F_{2g}^{T}—Y_{2g}^{C} in equation (2)—had the first year of AMSTI while the treatment group portion of F_{2g}^{T}—Y_{2g}^{T} in equation (2)—had the first year plus a second year.

Formal assessment of the approximation error in the F_{2g}^{T*} estimator requires consideration of the expected value of F_{2g}^{T*} in equation (2). Specifically,

$$E(F_{2g}^{T*}) = E(Y_{2g}^{T}) – E(Y_{2g}^{C}) + E(Y_{1g}^{T}) – E(Y_{1g}^{C}) = U_{2g}^{T} – M_{2g}^{T} – U_{2g}^{C} – M_{2g}^{C} + U_{1g}^{T}–+ M_{1g}^{T} – U_{1g}^{C} – M_{1g}^{C}, \quad \text{(Q4)}$$

where

- U_{2g}^{T} = expected outcome of the treatment group if untreated, in Year 2 for grade *g*.
- M_{2g}^{T} = expected impact of two years of AMSTI on the treatment group, compared with zero years, in Year 2 for grade *g*.
- U_{2g}^{C} = expected outcome of the control group if untreated in Year 2 for grade *g*.
- M_{2g}^{C} = expected impact of one year of AMSTI on the control group, compared with zero years, in Year 2 for grade *g*.
- U_{1g}^{T} = expected outcome of the treatment group if untreated in Year 1 for grade *g*.
- M_{1g}^{T} = expected impact of one year of AMSTI on the treatment group, compared with zero years, in Year 1 for grade *g*.
- U_{1g}^{C} = expected outcome of the control group if untreated in Year 1 for grade *g*.
- M_{1g}^{C} = expected impact of zero years of AMSTI on the control group in Year 2 for grade *g*.

By virtue of random assignment, the expected untreated outcome levels are the same for the treatment and control groups in any given year and grade level:

$$U_{2g}^{T} = U_{2g}^{C}. \quad \text{(Q5)}$$

$$U_{1g}^{T} = U_{1g}^{C}. \quad \text{(Q6)}$$

Furthermore, the expected impact on the control group in Year 1—when it receives no intervention—is zero:

$$M_{1g}^{C} = 0. \quad \text{(Q7)}$$

Substituting equations (Q5), (Q6), and (Q7) into equation (Q4) yields

$$E(F_{2g}^{T*}) = M_{2g}^{T} – M_{2g}^{C} + M_{1g}^{T}. \quad \text{(Q8)}$$

F_{2g}^{T*} is unbiased for M_{2g}^{T}, the impact of two years of AMSTI compared with no AMSTI, if

$M_{2g}^{C} = M_{1g}^{T}$. (Q9)

Thus, the lack of bias of the Bell-Bradley estimator rests on the equivalence of impacts in the control and treatment groups when each group receives its first year of the AMSTI intervention.

Standard error for estimator for the two-year impact.

The standard error of the estimator of this quantity (R2), which uses outcomes for grade *g*, for two consecutive birth cohorts of students, and which assumes groups (in this study, schools) are randomized, is:

$$SE\left(\hat{F}_{2g}^{T*}\right) = \sqrt{\frac{2\sigma^2}{P(1-P)nJ} + \frac{4\tau^2}{P(1-P)J}} \qquad \text{(Q10)}$$

where σ^2 = variance of individual outcomes around a group mean

τ^2 = variance of the group mean outcome across groups

J = number of groups randomized

n = number of individuals per group

P = proportion of groups randomized to the treatment group.

The derivation of this formula is provided in Bell and Bradley (in press).

Appendix R. Attrition through study stages for samples contributing to estimation of two-year effects

Estimating two-year impacts with the Bell-Bradley method requires samples used to estimate the effect for each outcome: one with which to estimate the Year 1 effect (Year 1 sample[116]) and one with which to estimate the differential effect in Year 2 (Year 2 sample[117]). This appendix explains how the sample for each of these components was developed to estimate the SAT 10 mathematics problem solving and science outcomes.

Selection of the Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving sample

Below are changes in the numbers of schools, teachers, and students from the point of randomization to the point at which the analytic sample associated with the SAT 10 mathematics problem solving outcome was identified (table R1).

Table R1 Attrition from Year 1 and Year 2 analytical samples contributing to estimation of two-year effect (associated with Stanford Achievement Test Tenth Edition [SAT 10)] mathematics problem solving outcome)

	Year 1						Year 2					
	AMSTI			Control			AMSTI			Control (with one year of AMSTI implementation)		
Item	Schools	Teachers	Students	Schools	Teachers	Students	Schools	Teachers	Students	Schools	Teachers	Students
Randomization	41	na	na	41	na	na	41	na	na	41	na	na
Numbers indicated in fall rosters	41	249	12,065	41	233	10,492	41	217	11,574	41	170	9,857
Loss because of disability	0	–3	–1,548	0	–4	–1,383	0	–4	–1,434	0	–5	–1,245
Available cases (baseline sample)	41	246	10,517	41	229	9,109	41	213	10,139	41	165	8,612
Loss because of students moving from Subexperiment 1 to Subexperiment 2 (between Year 1 and Year 2)	0	–0	–24	0	0	–26	0	0	–30	0	0	–25
Available cases	41	246	10,493	41	229	9,083	41	213	10,109	41	165	8,587

[116] The Year 1 sample includes AMSTI schools in their first year of AMSTI implementation and control schools with no AMSTI implementation (across subexperiments).
[117] The Year 2 sample includes AMSTI schools in their second year of AMSTI implementation and control schools in their first year of AMSTI implementation (across subexperiments).

	Year 1						Year 2					
	AMSTI			Control			AMSTI			Control (with one year of AMSTI implementation)		
Loss because of missing student identifier	0	0	0	0	0	0	0	0	0	0	0	0
Available cases	41	246	10,493	41	229	9,083	41	213	10,109	41	165	8,587
Loss because of missing school identifier	0	0	0	0	0	0	0	0	0	0	0	0
Available cases	41	246	10,493	41	229	9,083	41	213	10,109	41	165	8,587
Loss because of missing posttests	0	–2	–471	0	0	–392	0	0	–270	0	0	–222
Available cases	41	244	10,022	41	229	8,691	41	213	9,839	41	165	8,365
Loss because of duplicate records	0	–0	–23	0	0	–13	0	0	–6	0	0	–16
Available cases	41	244	9,999	41	229	8,678	41	213	9,833	41	165	8,349
Loss because of repeating or skipping a grade	0	–1	–479	0	0	–204	–1	–5	–447	0	–1	–205
Number of cases in sample	41	243	9,520	41	229	8,474	40	208	9,386	41	164	8,144

na is not applicable.
Source: Student achievement data from tests administered as part of the state's accountability system.

Random assignment phase

Both the Year 1 and Year 2 samples began with 41 schools randomly assigned to the AMSTI condition and 41 schools randomly assigned to the control condition. (The 41 control schools in Year 1 are the same 41 schools that are the 41 control schools in Year 2, but with one year of AMSTI implementation.)

Confirmation of rosters and teacher assignments

Districts provided information about teachers and student rosters at the start of the school year following randomization in Year 1 and at the start of the Year 2 school year. The record for the number of teachers and students began at the time that roster information was received.

The Year 1 rosters from AMSTI schools included 249 teachers with 12,065 students. The Year 1 rosters from the control schools included 233 teachers with 10,492 students.

The Year 2 rosters from AMSTI schools included 217 mathematics teachers and 11,574 students in grades 4–8. The Year 2 rosters from the control schools included 170 teachers and 9,857 students in grades 4–8.

Exclusion of data on students with disabilities

Data for 1,548 students from AMSTI schools and 1,383 students from control schools were excluded from the Year 1 sample because the students were classified as having disabilities. The remaining Year 1 sample consisted of 41 AMSTI schools (with 246 teachers and 10,517 students) and 41 control schools (with 229 teachers and 9,109 students).

Data for 1,434 students from AMSTI schools and 1,245 students from control schools were excluded from the Year 2 sample because the students were classified as having disabilities. The remaining Year 2 sample consisted of 41 AMSTI schools (with 213 mathematics teachers and 10,139 students in grades 4–8) and 41 control schools (with 165 mathematics teachers and 8,612 students in grades 4–8).

Exclusion of data on students who moved between subexperiments

Data from students whose identifiers appeared twice in the analytic sample were excluded from analysis. These were students who moved between subexperiments (between Year 1 and Year 2). Because the first subexperiment started one year before the second one, moving across experiments would have resulted in unknown levels of exposure to the intervention.

Data for 24 students from AMSTI schools and 26 students from control schools were excluded from the Year 1 sample because the students moved between subexperiments. Removing these data did not result in a decline in the number of teachers or schools.

Data for 30 students from AMSTI schools and 25 students from the control schools were excluded from the Year 2 sample because the students moved between subexperiments. Removing these data did not result in a decline in the number of teachers or schools.

Exclusion of data on students without valid identifiers

Students' data were excluded if they were missing valid student or school identifiers. It was critical to have this information in order to properly model school membership in the analysis. All students in the Year 1 and Year 2 samples were linked to school identifiers.

Exclusion of data on students without valid identifiers

Students' data were excluded from analysis if they were missing a posttest score or if their posttest score lay outside the range of the posttest scale. Data for 471 students from AMSTI schools and 392 students from control schools were excluded from the Year 1 sample because the students were missing posttests. Exclusion of these data resulted in the loss of two teachers from intervention schools and no teachers from control schools.

Data for 270 students in AMSTI schools and 222 students in control schools were excluded from the Year 2 sample because the students were missing posttests. The exclusion of these data did not result in teacher or school exclusions in Year 2 for either condition.

Treatment of missing data for covariates

Students' data were not excluded if they were missing values for one or more covariates in the model, including the pretest. A dummy variable method was used to address missing values for covariates. This method involved replacing the unobserved value with a constant and, for each covariate, adding an indicator variable with a value of 1 or 0 to signify whether the value of the corresponding covariate was observed or unobserved (missing).

Exclusion of students with duplicate records and students who skipped or repeated a grade

Students' data were excluded from analysis if duplicate records were found or the student was determined to have skipped or repeated a grade. As with the one-year impact analysis, the two-year impact analysis for science included students from grades 5 and 7 only. The end of Year 1 and end of Year 2 student samples consisted largely of different sets of students. The exceptions were students who repeated grade 5 or grade 7 or who skipped a grade, going from grade 5 in Year 1 to grade 7 in Year 2. To avoid having more than one outcome per student in cases in which repeated measures analyses were not used (that is, where the data were analyzed in wide form instead of long form and assuming independence in outcomes across rows), researchers excluded students who would have appeared twice in the dataset because they skipped or repeated grades. For consistency, they eliminated students using the same criteria in the two-year impact analysis of the mathematics problem solving outcome.

For the one-year confirmatory impact analyses, students who would go on to repeat or skip grades the following year could not be distinguished from the rest of the student population. Therefore, those results apply to all students, including skippers and repeaters. The results of the two-year impact analysis apply only to students who did not repeat or skip grades. For this reason, the sample contributing to the Year 1 component of the estimated two-year effect is slightly different from the sample used to compute the one-year impact in the confirmatory analyses.

From the Year 1 sample, in AMSTI schools 23 students were excluded because of duplicate records and 479 were excluded because of skipping or repeating a grade. Exclusion of these data resulted in the loss of one teacher. In control schools, 13 students were excluded because of duplicate records and 204 were excluded because of skipping or repeating a grade. Exclusion of these data resulted in the loss of no teachers or schools.

From the Year 2 sample, in AMSTI schools 6 students were excluded because of duplicative records and 447 were excluded because of skipping or repeating a grade. Exclusion of these data resulted in the loss of five teachers and one school. From the Year 2 sample, in control schools data from 16 students were excluded because of duplicative records and data from 205 students were excluded because of skipping or repeating a grade. Exclusion of these data resulted in the loss of one teacher.

Summary of selection of mathematics problem solving sample

The starting sample for Year 1 included 41 AMSTI schools (with 251 mathematics teachers and 12,198 students) and 41 control schools (with 234 mathematics teachers and 10,514 students). The baseline sample for the analysis of mathematics outcomes, which was limited to students without disabilities in grades 4–8, consisted of 41 AMSTI schools (with 246 teachers and 10,517 students) and 41 control schools (with 229 teachers and 9,109 students). The analytic sample for the analysis of mathematics outcomes consisted of 41 AMSTI schools (with 243 teachers and 9,520 students) and 41 control schools (with 229 teachers and 8,474 students).

The starting sample for Year 2 included 41 AMSTI schools (with 217 mathematics teachers and 11,574 students) and 41 control schools (with 170 mathematics teachers and 10,081 students). The baseline sample for the analysis of mathematics outcomes, which was limited to students without disabilities in grades 4–8, consisted of 41 AMSTI schools (with 213 teachers and 10,139 students) and 41 control schools (with 165 teachers and 8,612 students). The analytic sample for the analysis of mathematics outcomes consisted of 40 AMSTI schools (with 208 teachers and 9,386 students) and 41 control schools (with 164 teachers and 8,144 students).

Selection of the Stanford Achievement Test Tenth Edition (SAT 10) science sample

Below are changes in the numbers of schools, teachers, and students from the point of randomization to the point at which the analytic sample associated with the SAT 10 science outcome was identified (table R2).

Table R2 Attrition from Year 1 and Year 2 analytical samples contributing to estimation of two-year effect (associated with Stanford Achievement Test Tenth Edition [SAT 10)] science outcome)

	Year 1						Year 2					
	AMSTI			Control			AMSTI			Control (with one year of AMSTI implementation)		
Item	Schools	Teachers	Students	Schools	Teachers	Students	Schools	Teachers	Students	Schools	Teachers	Students
Randomization	41	na	na	41	na	na	41	206	na	41	156	na
Numbers indicated in fall rosters	41	233	12,065	41	213	10,492	41	206	11,574	41	156	10,081
Loss because not in grades 5 and 7	0	–128	–6,972	0	–116	–6,284	–2	–111	–6,688	0	–93	–6,140
Available cases	41	105	5,093	41	97	4,208	39	95	4,886	41	63	3,941
Loss because of disability	–2	–2	–613	0	–2	–520	0	–4	–633	0	–2	–526
Available cases (baseline sample)	39	103	4,480	41	95	3,688	39	91	4,253	41	61	3,415
Loss because of students moving from Subexperiment 1 to Subexperiment 2 (between Year 1 and Year 2)	0	0	–14	0	0	(9	0	0	–14	0	0	–11
Available cases	39	103	4,466	41	95	3,679	39	91	4,239	41	61	3,404
Loss because of missing student identifier	0	0	0	0	0	0	0	0	0	0	0	0
Available cases	39	103	4,466	41	95	3,679	39	91	4,239	41	61	3,404
Loss because of missing school identifier	0	0	0	0	0	0	0	0	0	0	0	0
Available cases	39	103	4,466	41	95	3,679	39	91	4,239	41	61	3,404
Loss because of missing posttests	0	–1	–384	–1	–5	–233	0	–4	–261	–1	–1	–157
Available cases	39	102	4,082	40	90	3,446	39	87	3,978	40	60	3,247
Loss because of duplicate records	0	0	–7	0	0	–8	0	0	–2	0	0	–2
Available cases	39	102	4,075	40	90	3,438	39	87	3,976	40	60	3,245
Loss because of repeating or skipping a grade	0	–1	–161	0	0	–74	–1	–2	–133	0	0	–84
Number of cases in sample	39	101	3,914	40	90	3,364	38	85	3,843	40	60	3,161

na is not applicable.
Source: Student achievement data from tests administered as part of the state's accountability system.

Random assignment phase

Both the Year 1 and Year 2 samples began with 41 schools randomly assigned to the AMSTI condition and 41 schools randomly assigned to the control condition. (The 41 control schools in Year 1 are the same 41 schools that are the 41 control schools in Year 2, but with one year of AMSTI implementation.)

Confirmation of rosters and teacher assignments

Districts provided information about teachers and student rosters at the start of the school year following randomization in Year 1 and at the start of the Year 2 school year. The record for the number of teachers and students began at the time that roster information was received.

The Year 1 rosters from AMSTI schools included 233 teachers with 12,065 students. The Year 1 rosters from control schools included 213 teachers with 10,492 students. Data from students not in grade 5 or 7 were removed from the sample, resulting in 105 teachers and 5,093 students in 41 AMSTI schools and 97 teachers and 4,208 students in 41 control schools.

The Year 2 rosters from AMSTI schools included 206 teachers and 11,574 students. The Year 2 rosters from control schools included 156 teachers and 10,081 students. Retaining data from students only in grade 5 or 7 resulted in 95 teachers and 4,886 students in 39 AMSTI schools and 63 teachers and 3,941 students in 41control schools.

Exclusion of data on students with disabilities

Data for 613 students from AMSTI schools and 520 students from control schools were excluded from the Year 1 sample because the students were classified as having disabilities. Exclusion of these data resulted in the loss of two teachers and two schools in the AMSTI condition and two teachers in the control condition. The remaining Year 1 sample consisted of 39 AMSTI schools (with 103 teachers and 4,480 students) and 41 control schools (with 95 teachers and 3,688 students).

Data for 633 students from AMSTI schools and 526 students from control schools were excluded from the Year 2 sample because the students were classified as having disabilities. Exclusion of these data resulted in the loss of four teachers in AMSTI schools and two teachers in control schools. The remaining Year 2 sample consisted of 39 AMSTI schools (with 91 science teachers and 4,253 students in grades 5 or 7) and 41 control schools (with 61 science teachers and 3,415 students in grades 5 or 7).

Exclusion of data on students who moved between subexperiments

Data from students whose identifiers appeared twice in the analytic sample were excluded from analysis. These were students who moved between subexperiments (between Year 1 and Year 2). Because the first subexperiment started one year before the second one, moving between the two subexperiments would have resulted in unknown levels of exposure to the intervention.

Data for 14 students from AMSTI schools and 9 students from control schools were excluded from the Year 1 sample because students moved between subexperiments. From the Year 2 sample, data for 14 students from AMSTI schools and 11 students from control schools were excluded because students moved between subexperiments. Removing these students' data did not result in the loss of teachers or schools for either year.

Exclusion of data on students without valid identifiers

Students' data were excluded if they were missing valid student or school identifiers. It was critical to have this information in order to properly model school membership in the analysis. All students in the Year 1 and Year 2 samples were linked to school identifiers.

Exclusion of data on students without valid posttests

Students' data were excluded from analysis if they were missing a posttest score or if their posttest score lay outside the range of the posttest scale. Data for 384 students from AMSTI schools and 233 students from control schools were excluded from the Year 1 sample because students were missing posttests. Exclusion of these data resulted in the loss of one teacher from the AMSTI condition and five teachers and one school from the control condition.

Data for 261 students from AMSTI schools and 157 students from control schools were excluded from the Year 2 sample because students were missing posttests. Exclusion of these data resulted in the loss of four teachers in AMSTI condition and one teacher and one school in the control condition.

Treatment of missing data for covariates

Students' data were not excluded if they were missing values for one or more covariates in the model, including the pretest. A dummy variable method was used to address missing values for covariates. This method involved replacing the unobserved value with a constant and, for each covariate, adding an indicator variable with a value of either 1 or 0 to signify whether the value of the corresponding covariate is observed or unobserved (missing).

Exclusion of students with duplicate records and students who skipped or repeated a grade

Students' data were excluded from analysis if duplicate records were found or the student was determined to have skipped or repeated a grade. From the Year 1 sample, in AMSTI schools 7 students were excluded because of duplicate records, and 161 students were excluded because of skipping or repeating a grade. Exclusion of these data resulted in the loss of one teacher. In control schools, 8 students were excluded from the control condition because of duplicate records and 74 students were excluded because of skipping or repeating a grade. Exclusion of these data resulted in the loss of no teachers or schools.

From the Year 2 sample, in AMSTI schools data from 2 students were excluded because of duplicative records and data from 133 students were excluded because of skipping or repeating a grade. Exclusion of these data resulted in the loss of two teachers and one school. From the Year 2 sample, in control schools 2 students were excluded because of duplicative records and 84 students were excluded because of students skipping or repeating a grade. Exclusion of these data resulted in the loss of no teachers or schools from the control condition.

Summary of selection of science sample

The starting sample for Year 1 included 41 AMSTI schools (with 233 science teachers and 12,065 students) and 41 control schools (with 213 science teachers and 10,492 students). The baseline sample for the analysis of science outcomes, which was limited to students without disabilities in grade 5 or 7, consisted of 39 AMSTI schools (with 103 teachers and 4,480 students) and 41 control schools (with 95 teachers and 3,688 students). The analytic sample for the analysis of science outcomes consisted of 39 AMSTI schools (with 101 teachers and 3,914 students) and 40 control schools (with 90 teachers and 3,364 students).

The starting sample for Year 2 included 41 AMSTI schools (with 206 science teachers and 11,574 students) and 41 control schools (with 156 science teachers and 10,081 students). The baseline sample for the analysis of science outcomes, which was limited to students without disabilities in grade 5 or 7, consisted of 39 AMSTI schools (with 91 teachers and 4,253 students) and 41 control schools (with 61 teachers and 3,415 students). The analytic sample for the analysis of science outcomes consisted of 38 AMSTI schools (with 85 teachers and 3,843 students) and 40 control schools (with 60 teachers and 3,161 students).

Appendix S. Examination of equivalence in baseline and analytic samples used in the estimation of two-year effects

This appendix provides data on the mean characteristics of the baseline (table S1) and analytical (table S2) samples contributing to the Year 2 component of the Bell-Bradley estimate for both the mathematics problem solving and science outcomes.

Table S1 Mean characteristics of baseline sample contributing to the Year 2 component of the Bell-Bradley estimate (associated with Stanford Achievement Test Tenth Edition [SAT 10] mathematics problem solving and science outcomes)

		Mathematics problem solving			**Science**		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
Student Characteristic							
Average of school percent of boys	Average in Each Condition	50.1		49.6	51.3		49.2
	Standard Deviation	4.9		3.5	6.8		5.7
	Sample Size (Schools)	41		41	39		41
	Average Difference		0.5			2.1	
	Standard Error		1.0			1.4	
	Test statistic		*t* = 0.51			*t* = 1.53	
	***p*-value**		.61			.13	
Average of school percent of minority students	Average in Each Condition	50.7		47.2	51.5		48.0
	Standard Deviation	34.4		33.4	35.5		34.0
	Sample Size (Schools)	41		41	39		41
	Average Difference		3.6			4.0	
	Standard Error		7.5			7.8	
	Test statistic		*t* = 0.47			*t* = 0.52	
	***p*-value**		.64			.60	

[118] The Year 1 sample includes AMSTI schools in their first year of AMSTI implementation and control schools with no AMSTI implementation (across subexperiments).

[119] The Year 2 sample includes AMSTI schools in their second year of AMSTI implementation and control schools in their first year of AMSTI implementation (across subexperiments).

		Mathematics problem solving			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
Average of school percent of students proficient in English	Average in Each Condition	98.1		98.7	98.2		99.0
	Standard Deviation	3.7		2.2	3.9		2.2
	Sample Size (Schools)	41		41	39		41
	Average Difference		–0.6			–0.7	
	Standard Error		0.7			0.7	
	Test statistic		*t* = 0.88			*t* = 0.93	
	***p*-value**		.38			.35	
Average of school percent of students enrolled in the free or reduced-price lunch program	Average in Each Condition	64.9		65.9	64.7		65.0
	Standard Deviation	24.5		23.9	24.4		25.0
	Sample Size (Schools)	41		41	39		41
	Average Difference		–1.0			–0.3	
	Standard Error		5.3			5.5	
	Test statistic		*t* = 0.20			*t* = 0.05	
	***p*-value**		.85			.96	
Average of school percent of students in grade 4	Average in Each Condition	24.2		22.9	na		na
	Standard Deviation	23.5		23.0	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		–1.3			na	
	Standard Error		5.1			na	
	Test statistic		*t* = 0.26			na	
	***p*-value**		.80			na	
Average of school percent of students in grade 5	Average in Each Condition	26.2		22.9	61.7		56.4
	Standard Deviation	20.0		20.2	40.6		46.4
	Sample Size (Schools)	41		41	39		41
	Average Difference		3.4			5.3	
	Standard Error		4.4			9.8	
	Test statistic		*t* = 0.76			*t* =0.54	
	***p*-value**		.45			.59	
Average of school percent of students in grade 6	Average in Each Condition	14.6		18.3	na		na
	Standard Deviation	14.6		16.5	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		–3.8			na	
	Standard Error		3.4			na	
	Test statistic		*t* = 1.09			na	
	***p*-value**		.28			na	

		Mathematics problem solving			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
Average of school percent of students in grade 7	Average in Each Condition	16.1		17.9	38.3		43.6
	Standard Deviation	18.3		19.9	40.6		46.4
	Sample Size (Schools)	41		41	39		41
	Average Difference		–1.8			na	
	Standard Error		4.2			na	
	Test statistic		*t* = 0.43			na	
	***p*-value**		.67			na	
Average of school percent of students in grade 8	Average in Each Condition	18.9		18.0	na		na
	Standard Deviation	22.5		19.9	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		0.9			na	
	Standard Error		4.7			na	
	Test statistic		*t* = 0.19			na	
	***p*-value**		.85			na	
School average pretest score and sample size							
SAT10[a]	Pretest Score	630.5		631.9	632.6		635.1
	Standard Deviation	24.1		21.5	19.5		17.7
	Sample Size (Schools)	41		41	38		41
	Average Difference		–1.4			–2.6	
	Standard Error		5.1			4.2	
	Test statistic		*t* = 0.28			*t* = 0.61	
	***p*-value**		.78			.54	
Sample Size	Number of schools = 41 Number of teachers = 213 Number of students = 10,139			Number of schools = 41 Number of teachers = 165 Number of students = 8,612	Number of schools = 39 Number of teachers = 91 Number of students = 4,253		Number of schools = 41 Number of teachers = 61 Number of students = 3,415

na is not applicable

Note: Detail may not sum to totals because of rounding. The number of schools and teachers for the comparisons varied slightly because of missing data. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Grade level for mathematics problem solving is a categorical variables with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of students across grades for the mathematics problem solving outcome (p = .81). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the p-value was .78 for mathematics problem solving and .87 for science.)

a. The SAT 10 mathematics problem solving pretest was used for the mathematics outcome. The SAT 10 reading pretest was used for the science outcome.

Source: Student achievement data from tests administered as part of the state's accountability system, student demographic data from state data system, and teacher survey data.

Table S2 Mean characteristics of analytic samples contributing to the Year 2 component of the Bell-Bradley estimate (associated with Stanford Achievement Test Tenth Edition [SAT 10] mathematics problem solving and science outcomes)

		Mathematics problem solving			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
Student Characteristic							
Average of school percent of boys	Average in Each Condition	49.7		49.5	50.7		49.1
	Standard Deviation	5.2		3.4	6.9		5.3
	Sample Size (Schools)	40		41	38		40
	Average Difference		0.2			1.5	
	Standard Error		1.0			1.4	
	Test statistic		t = 0.21			t = 1.10	
	***p*-value**		.84			.28	
Average of school percent of minority students	Average in Each Condition	49.6		47.1	50.4		45.9
	Standard Deviation	34.0		33.4	35.2		33.4
	Sample Size (Schools)	40		41	38		40
	Average Difference		2.5			4.5	
	Standard Error		7.5			7.8	
	Test statistic		t = 0.33			t = 0.58	
	***p*-value**		.74			.56	
Average of school percent of students proficient in English	Average in Each Condition	98.1		98.7	98.3		98.9
	Standard Deviation	3.7		2.1	3.7		2.2
	Sample Size (Schools)	40		41	38		40
	Average Difference		–0.6			–0.5	
	Standard Error		0.7			0.7	
	Test statistic		t = 0.88			t = 0.79	
	***p*-value**		.38			.43	
Average of school percent of students enrolled in the free or reduced-price lunch program	Average in Each Condition	63.7		65.2	63.5		63.4
	Standard Deviation	24.4		24.1	24.3		24.9
	Sample Size (Schools)	40		41	38		40
	Average Difference		–1.5			0.1	
	Standard Error		5.4			5.6	
	Test statistic		t = 0.29			t = 0.02	
	***p*-value**		.78			.98	
Average of school percent of students in grade 4	Average in Each Condition	25.0		23.3	na		na
	Standard Deviation	23.6		23.1	na		na
	Sample Size (Schools)	40		41	na		na
	Average Difference		–1.7			na	

		Mathematics problem solving			Science		
Sample characteristic		**AMSTI schools**		**Control schools**	**AMSTI schools**		**Control schools**
	Standard Error		5.2			na	
	Test statistic		*t* = 0.33			na	
	***p*-value**		.74			na	
Average of school percent of students in grade 5	Average in Each Condition	26.6		23.4	63.2		56.0
	Standard Deviation	20.4		20.5	41.3		46.6
	Sample Size (Schools)	40		41	38		40
	Average Difference		3.3			7.3	
	Standard Error		4.5			10.0	
	Test statistic		*t* = 0.72			*t* = 0.73	
	***p*-value**		.47			.47	
Average of school percent of students in grade 6	Average in Each Condition	15.0		18.1	na		na
	Standard Deviation	15.6		16.6	na		na
	Sample Size (Schools)	40		41	na		na
	Average Difference		–3.1			na	
	Standard Error		3.6			na	
	Test statistic		*t* = 0.85			na	
	***p*-value**		.40			na	
Average of school percent of students in grade 7	Average in Each Condition	15.3		17.5	36.8		44.0
	Standard Deviation	18.5		19.6	41.3		46.6
	Sample Size (Schools)	40		41	38		40
	Average Difference		–2.2			na	
	Standard Error		4.2			na	
	Test statistic		*t* = 0.52			na	
	***p*-value**		.60			na	
Average of school percent of students in grade 8	Average in Each Condition	18.1		17.8	na		na
	Standard Deviation	22.8		20.1	na		na
	Sample Size (Schools)	40		41	na		na
	Average Difference		0.3			na	
	Standard Error		4.8			na	
	Test statistic		*t* = 0.06			na	
	***p*-value**		.95			na	
School average pretest score and sample size							
SAT10[a]	Pretest Score	630.3		632.1	631.7		635.2
	Standard Deviation	24.6		22.0	21.2		17.7
	Sample Size (Schools)	40		41	36		38

		Mathematics problem solving			Science		
Sample characteristic		AMSTI schools		Control schools	AMSTI schools		Control schools
	Average Difference		–1.8			–3.5	
	Standard Error		5.2			4.5	
	Test statistic		*t* = 0.34			*t* = 0.77	
	***p*-value**		.73			.44	
Sample Size		Number of schools = 40 Number of teachers = 208 Number of students = 9,386		Number of schools = 41 Number of teachers = 164 Number of students = 8,144	Number of schools = 38 Number of teachers = 85 Number of students = 3,843		Number of schools = 40 Number of teachers = 60 Number of students = 3,161

na is not applicable.

Note: Detail may not sum to totals because of rounding. The number of schools and teachers for the comparisons varied slightly because of missing data. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Grade level for mathematics problem solving is a categorical variables with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of students across grades for the mathematics problem solving outcome (p = .87). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the *p*-value was .19 for mathematics problem solving and .29 for science.)

a. The SAT 10 mathematics problem solving pretest was used for the mathematics outcome. The SAT 10 reading pretest was used for the science outcome.

Source: Student achievement data from tests administered as part of the state's accountability system, student demographic data from state data system, and teacher survey data.

Appendix T. Estimation model for two-year effects of the Alabama Math, Science, and Technology Initiative (AMSTI)

This appendix explains how the two-year effects of AMSTI were estimated. It describes the time structure of the data and how it relates to the model specification, presents the multilevel equations of the model and discusses the estimation process, and explains how the two-year impact was estimated from the model's estimated parameters.

The same two-level model used in the one-year analysis was used. As noted in chapter 2, for mathematics problem solving, multiple observations over time existed for most students in the analysis file. Multiple observations occurred because the Bell-Bradley method calculates treatment-control group differences in outcomes for the same students in consecutive grades. For example, to obtain an estimate for grade 6 students, the Bell-Bradley method combines data for grade 6 students in the year of interest with data on the preceding year's grade 6 students (table T1). The preceding year's grade 6 students are also analyzed as the current year's grade 7 students, resulting in two observations for every student in this cohort.[120] To obtain unbiased standard errors that take account of multiple observations per student over time (that is, nesting of scores within students), post hoc adjustments were made to the conventional standard errors for the Bell-Bradley estimators (see chapter 5 for details of the adjustment made). Science test scores were available only for grade 5 and 7 students, so no students appear in the analysis sample more than once (those who skipped or repeated grades were removed.) There was therefore no need to consider the nesting of scores within students (that is, repeated measures) and hence no need to make post hoc adjustments to standard errors.

Table T1 Multiple observations per student used in analyzing two-year effects

Outcome/birth cohort	Grade level observed during 2006/07 school year	Grade level observed during 2007/08 school year
Mathematics achievement		
1992/93	Grade 8	na
1993/94	Grade 7	Grade 8
1994/95	Grade 6	Grade 7
1995/96	Grade 5	Grade 6
1996/97	Grade 4	Grade 5
1997/98	na	Grade 4
Science achievement		
1993/94	Grade 7	na
1994/95	na	Grade 7
1995/96	Grade 5	na
1996/97	na	Grade 5

na is not available.
Note: Table illustrates repeated measures for the Subexperiment 1 sample.

[120] Only single observations were used for the youngest and oldest cohorts, namely, students who were in grade 9 in the second year of the study and students who did not enter the study window (that is, grade 4) until the second year.

Specification of the model, with students at level 1 and schools at level 2, is as follows:[121]

Level 1 (student level):

The two-year level 1 model differs from the one-year level 1 model because it includes an indicator of whether a student belongs to the first student age cohort or the second (one year younger) cohort (COHORT).[122]

$$y_{ij} = \beta_{0j} + \beta_{1j} COHORT_{ij} + \sum_{p=1}^{12} \beta_{pj} COV_{pij} + e_{ij} .$$

$COHORT_{1ij}$ indicates the birth cohort to which the student belongs. It takes on the value of 1 for students in the younger birth cohort (whose Year 2 estimated effect is of interest) and 0 for the older birth cohort (whose Year 1 estimated effect is of interest). COV_{pij} is the value of the covariate *p*. Covariates include the pretest and indicators of racial/ethnic minority status, free or reduced-price lunch status, English proficiency, grade level, and gender. Dummy variables for each covariate indicate whether the value of the covariate is missing. e_{ij} is the random effect associated with student *i* in school *j*, conditioning on the other variables in the model.

Level 2 (school level):

Level 2 is specified the same way as in the one-year effect model, but the coefficient for the cohort effect is expressed as a linear function of the treatment effect (that is, it is allowed to differ between AMSTI and control schools):[123]

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} \overline{X}_j + \gamma_{02} T_j + \sum_{m=3}^{42} \gamma_{0m} I_{(m-3)} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} T_j \qquad\qquad p = 1,\ldots,12 \\ \beta_{pj} &= \gamma_{p0} . \end{aligned}$$

[121] The model used to estimate the effect of AMSTI on student performance in science is presented. The effect on mathematics problem solving was estimated using a similar model but one that included a larger number of fixed effects, because grades 4–8 were included in that analysis, not just grades 5 and 7. Therefore, four dummy variables were used to indicate grade, with grade 6 serving as the reference grade.

[122] As shown in appendix Q, estimation using the Bell-Bradley method requires data from two consecutive age cohorts of students, one cohort providing an AMSTI–control difference in mean outcomes for Year 2 and the other providing an AMSTI–control difference in outcomes at the same grade level a year earlier (in Year 1). The students used to calculate the second difference were born a year earlier than the students used to calculate the first difference.

[123] When the different levels of the model are brought together, this difference translates into different-size effects for the two birth cohorts—an essential distinction in using the model to obtain the two components of the Bell-Bradley estimator.

$\overline{X}_j$ is the mean pretest score for school *j*.

T_j indicates whether a school is assigned to the AMSTI or control condition in year-1 (coded 1 if the school is assigned to AMSTI and 0 if the school is assigned to control).

$I_{(m-3)}$ indicates the matched pair to which a school belongs; it takes on a value of 0 or 1. There are 40 indicators for the 41 pairs.

$u_{0,j}$ is the random effect of school *j*, conditioning on the other variables in the model.

Substituting the level 2 equations into level 1 yields the following mixed model:

$$y_{ij} = \gamma_{00} + \gamma_{01}\overline{X}_j + \gamma_{02}T_j + \sum_{m=3}^{42}\gamma_{0m}I_{(m-3)} + \sum_{p=1}^{12}\gamma_{p0}COV_{pij}$$
$$+\gamma_{10}COHORT_{ij} + \gamma_{11}T_j * COHORT_{ij} + u_{0j} + e_{ij}.$$

The MIXED procedure from the SAS Institute (2006) was used to estimate the model.[124] Each outcome variable was regressed against the fixed effects, including a school-level indicator variable for treatment status. Adjustment for random imbalances between the two groups is provided in this model by COV_{pij} the student-level covariates also used in the one-year impact analyses, and $\overline{X}_j$, the school mean of the pretest. To account for the matched-pairs component of the random assignment design, the model includes variables, $I_{(m-3)}$, which identify the matched pairs within which schools were randomized.

The regression-adjusted average difference between treatment and control group outcomes in the older birth cohort at the end of Year 1 is estimated as: $\hat{\gamma}_{02}$, which is the regression-adjusted version of $\overline{Y}_{1g}^T - \overline{Y}_{1g}^C$ from appendix Q. The regression-adjusted average difference between treatment and control group outcomes in the younger birth cohort at the end of Year 2 is estimated $\hat{\gamma}_{02} + \hat{\gamma}_{11}$, where $\hat{\gamma}_{02}$ is defined as above and $\hat{\gamma}_{11}$ is the regression-adjusted version of $\overline{Y}_{2g}^T - \overline{Y}_{2g}^C - \overline{Y}_{1g}^T - \overline{Y}_{1g}^C$, making $\hat{\gamma}_{02} + \hat{\gamma}_{11}$ the regression-adjusted version of $\overline{Y}_{2g}^T - \overline{Y}_{2g}^C$ from appendix Q.

As explained in appendix Q, the Bell-Bradley estimate of the effect of two years of AMSTI compared with no AMSTI is simply the sum of the two-regression adjusted differences for students in the same grade level *g* in consecutive years, constructed as

$$EFFECT = \hat{\gamma}_{02} + (\hat{\gamma}_{02} + \hat{\gamma}_{11}) = 2 * \hat{\gamma}_{02} + \hat{\gamma}_{11}.$$

---

[124] Singer (1998) describes the procedure for conducting mixed-model analyses of hierarchical data sets using SAS.

Exploratory measures of the effect of two years of AMSTI compared with no AMSTI are reported based on this formula for both mathematics problem solving and science. Also reported are the effect sizes associated with each estimate (calculated as the estimate of effect divided by the pooled standard deviation of the outcome measure in Year 2) and the differences in percentile standing, as described in the discussion of the one-year impact analyses. Although two outcomes are analyzed, adjustments for multiple comparisons are not made, because both analyses are exploratory. The results are presented in chapter 5.

Appendix U. Topics and instructional methods used at the Alabama Math, Science, and Technology Initiative (AMSTI) summer institute

Trainers at the AMSTI summer institute covered a variety of topics and used a variety of instructional methods. The number of days each trainer covered a topic to a moderate extent or more were added and the sum divided by the number of trainers for that grade/subject level in order to determine the average number of days each topic was covered to a moderate extent or more. For the grade 5 trainings, there were four trainers (tables U1 and U2). For the grade 7 trainings, there were two trainers (table U3). Only topics, and not summary statistics, are listed for grade 7 trainings, because presenting summary statistics on only two informants posed a threat to participant confidentiality.

Table U1 Grade 5 mathematics topics covered to a moderate extent or more by Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1

Topic	Average number of days (out of 5)	Standard deviation	Standard error
Mathematical thinking	1.00	0.82	0.41
Landmarks in the number system	1.50	1.00	0.50
Fractions	2.75	0.50	0.25
Percents	2.75	0.50	0.25
Decimals	2.75	1.26	0.63
Computation strategies	1.50	0.58	0.29
Estimation strategies	0.75	0.96	0.48

Note: Statistics are based on self-report of four trainers.
Source: Professional development training logs.

Table U2 Grade 5 science topics covered to a moderate extent or more by Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1

Topic	Average number of days (out of 5)	Standard deviation	Standard error
Science notebooks	2.50	1.29	0.65
Microscopes: lenses, practice, and slide preparation	2.00	0.82	0.41
Observation skills	2.00	0.82	0.41
Flippers investigation	1.25	0.50	0.25
Looking at living things: blepharisma	1.25	0.50	0.25
Looking at living things: vinegar eels	1.25	0.50	0.25
Looking at living things: volvox	1.25	0.50	0.25
Lifeboats investigation (capacity)	1.00	0.00	0.00
Looking at living things: hay and grass infusions	1.00	0.82	0.41
Plane sense investigation (catapults)	1.00	0.82	0.41
Swingers investigation (pendulum)	1.00	0.00	0.00
Globe: soil	0.75	0.50	0.25
Globe: atmosphere	0.50	0.58	0.29

Note: Statistics are based on self-report of four trainers.
Source: Professional development training logs.

Table U3 Grade 7 mathematics and science topics covered by Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1

Mathematics	Science
Enlarging figures using rubber-band stretchers and coordinating plotting	Science notebook
Visualizing similar and distorted transformations informally	Describing and naming organisms, producing scientific drawings, and observation skills
Identifying similar figures by side lengths and angles	Pond ecosystems: constructing and observing
Recognizing scale factors for similar figures	Plants: reproduction, leaf structure, transpiration, and flower structure
Reptiles: building and dividing shapes	Plant and animal cells: observing, drawing, and measuring
Understanding the relationship between similarity and equivalent fractions	Cell division: understanding, and creating a model
Understanding areas of similar figures	Protists: observing, drawing, and measuring
Understanding similar triangles: rules	Fungi: molds, mold formation, fungal garden, yeast cells
Understanding similar rectangles: rules	Daphnia: drawing and experimenting
Solving for unknown lengths with scale factors	Hydra: sketching and observing, feeding, and reproduction
Making connections to the real world	Seeds: harvesting and preparing, observing new sprouts
Making connections to algebra	Globe: identification of living organisms
Making connections to geometry	Globe: measurement of living organisms
Using geometry software	
Using writing in mathematics	

Source: Professional development training logs.

Instructional methods used during the summer institutes included hands-on activities, lesson demonstrations, small group discussions, and other methods (tables U4–U6).

Table U4 Use of instructional methods more than 25 percent of the time by grade 5 mathematics Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1

Instructional method	Average number of days (out of 5)	Standard deviation	Standard error
Hands-on activities	4.0	2.0	1.0
Lesson demonstration	3.5	2.4	1.2
Small group discussion	2.5	2.4	1.2
Skills practice	1.5	1.9	1.0
Writing in math	1.3	1.5	0.8
Whole group discussion	1.0	1.2	0.6
Lecture	0	0	0
Computer-based instruction	0	0	0

Note: Statistics are based on self-report of four trainers.
Source: Professional development training logs.

Table U5 Use of instructional methods more than 25 percent of the time by grade 5 science Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1

Instructional method	Average number of days (out of 5)	Standard deviation	Standard error
Hands-on activities	4.5	0.6	0.3
Lesson demonstration	3.5	1.9	1.0
Skills practice	2.8	2.1	1.0
Small group discussion	2.8	2.1	1.0
Whole group discussion	1.5	1.3	0.7
Lecture	0.8	1.0	0.5
Computer-based instruction	0.5	1.0	0.5
Writing in science	0.3	0.5	0.2

Note: Statistics are based on self-report of four trainers.
Source: Professional development training logs.

Table U6 Use of instructional methods more than 25 percent of the time by grade 7 mathematics and science Alabama Math, Science, and Technology Initiative (AMSTI) trainers in Year 1

Instructional method	Average number of days (out of 10)	Standard deviation	Standard error
Hands-on activities	7.8	2.2	0.8
Lesson demonstration	6.3	4.5	1.6
Skills practice	5.5	4.2	1.5
Whole group discussion	4.8	5.0	1.8
Writing in science	4.3	3.8	1.3
Small group discussion	4.0	4.6	1.6
Lecture	2.5	4.4	2.2
Computer-based instruction	1.5	1.7	0.6

Note: Statistics are based on self-report of two trainers.
Source: Professional development training logs.

Appendix V. Parameter estimates on probability scale for odds-ratio tests of differences between Alabama Math, Science, and Technology Initiative (AMSTI) and control conditions in Year 1 (associated with summer professional development and in-school support outcomes)

This appendix includes the parameter estimates for the probability scale for odds-ratio tests of differences between AMSTI and control conditions in Year 1 associated with summer professional development and in-school support outcomes (table V1) as well as the odds-ratio tests of differences between AMSTI and control conditions in Year 1 (table V2).

Table V1 Parameter estimates on probability scale for odds-ratio tests of differences between Alabama Math, Science, and Technology Initiative (AMSTI) and control conditions in Year 1 associated with summer professional development and in-school support outcomes

Model/condition	Marginal probability	Standard error
If mathematics teachers received summer professional development		
AMSTI	.86	0.04
Control	.24	0.05
If science teachers received summer professional development		
AMSTI	.87	0.04
Control	.24	0.06
If mathematics teachers requested support for instruction		
AMSTI	.30	0.04
Control	.33	0.04
If mathematics teachers received support for instruction		
AMSTI	.44	0.04
Control	.23	0.04
If science teachers requested support for instruction		
AMSTI	.54	0.05
Control	.40	0.05
If science teachers received support for instruction		
AMSTI	.65	0.04
Control	.25	0.04

Note: The standard error estimates for the marginal probabilities account for the clustering of students within schools. Both the logit parameter estimates and their standard errors were translated into probability units using an inverse link function and used to form a test statistic for the null hypothesis of no impact.
Source: Teacher survey data.

Table V2 Parameter estimates for odds-ratio tests of differences between Alabama Math, Science, and Technology Initiative (AMSTI) and control conditions in Year 1

Model	Coefficient (estimated log-odds)	Standard error	Degrees of freedom	Odds ratio	*p*-value
If mathematics teachers received summer professional development	3.0	0.4	55.96	19.9	< .01
If science teachers received summer professional development	2.8	0.4	49.53	16.6	< .01
Access to mathematics materials and manipulatives	1.6	0.3	57.08	5.1	< .01
Access to science materials and manipulatives	1.1	0.3	73.03	2.9	< .01
If mathematics teachers requested support for instruction	–0.1	0.3	66.77	0.9	.62
If mathematics teachers received support for instruction	0.6	0.3	72.85	1.8	.05
If science teachers requested support for instruction	1.0	0.3	58.90	2.7	< .01
If science teachers received support for instruction	1.7	0.3	65.96	5.7	< .01

Note: The odds ratio compares the odds of an event for teachers exposed to AMSTI with the odds of the same event for teachers not exposed to AMST; it indicates how many times greater the odds of the event are for teachers exposed to AMSTI. All entries in this table, except those for access to material and manipulatives, refer to dichotomous outcomes. The odds therefore correspond to an event happening or not. Access to materials and manipulatives are measured using an ordinal scale. For these outcomes, the odds are based on probabilities of selecting a given response category or ones of a lower order. Teachers in AMSTI schools have higher odds of selecting response options indicating greater access to materials. Both the logit parameter estimates and their standard errors were translated into probability units using an inverse link function and used to form a test statistic for the null hypothesis of no impact.
Source: Teacher survey data.

Appendix W. Descriptive statistics for variables that change to a binary scale used in the Alabama Math, Science, and Technology Initiative (AMSTI) and control conditions in Year 1

Below are the descriptive statistics for variables that change to a binary scale in Year 1 (table W1).

Table W1 Descriptive statistics for variables that change to binary scale in Year 1

Variable	Mean	Standard deviation	Skewness	Kurtosis
Hours of summer professional development received (mathematics)	25.4	29.6	1.1	1.3
Hours of summer professional development received (science)	34.0	77.2	6.8	57.7
Average number of times mathematics teachers requested support	0.5	1.3	3.3	12.5
Average number of times science teachers requested support	0.4	1.1	5.2	37.0
Average number of times mathematics teachers received support	1.5	7.2	10.7	122.5
Average number of times science teachers received support	0.8	2.9	10.9	153.9

Note: Skewness is a measure of the asymmetry of a probability distribution. Kurtosis is a measure of the "peakedness" of the distribution.
Source: Teacher survey data.

Appendix X. Comparison of assumed parameter values and observed sample statistics for statistical power analysis after one year

The tables in this appendix compare the assumed parameter values and observed sampled statistics for statistical power analysis associated with one-year outcomes for SAT 10 mathematics problem solving (table X1), SAT 10 science (table X2), active learning in mathematics (table X3), and active learning in science (table X4).

Table X1 Comparison of assumed parameter values and observed sample statistics for statistical power analysis associated with Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving outcome after one year

Variable	Assumed parameter value (design phase)	Observed sample statistic (analysis phase)
Minimum detectable effect size	0.20	0.06[a]
School-level intraclass correlation (ρ)	.22	.22[b]
Proportion of between-school variance in posttest explained by covariates in model (R_c^2)	.64	.97
Proportion of within-school variance in posttest explained by covariates in model (R_i^2)	0	0.58
Number of schools (J)	66	82
Number of students per school (n)	280	228 (on average)
Proportion of schools assigned to treatment (P)	.50	.50

Note: Sample statistics should be interpreted with caution, as standard errors are not reported.

a. The observed (sample-based) minimum detectable effect size was estimated using the following formula (Bloom, Richburg-Hayes, and Black 2007). The multiplier is based on an alpha level of .025, reflecting the adjustment for multiple comparisons.

$$MDES \approx 3.18 * \sqrt{\frac{\rho(1-R_c^2)}{P(1-P)J} + \frac{(1-\rho)(1-R_i^2)}{P(1-P)nJ}}$$

b. Estimated as the ratio of the school-level variance component for the posttest to the sum of the variance components for schools and students in a two-level model that includes a fixed effect for treatment only.

Source: Student achievement data from tests administered as part of the state's accountability system (for observed sample statistics).

Table X2 Comparison of assumed parameter values and observed sample statistics for statistical power analysis associated with Stanford Achievement Test Tenth Edition (SAT 10) science outcome after one year

Variable	Assumed parameter value (design phase)	Observed sample statistic (analysis phase)
Minimum detectable effect size	0.22[a]	0.13[b]
School-level intraclass correlation (ρ)	.22	.22[c]
Proportion of between-school variance in posttest explained by covariates in model (R_c^2)	.64	.88
Proportion of within-school variance in posttest explained by covariates in model (R_i^2)	0	.44
Number of schools (J)	66	79
Number of students per school (n)	112	95 (on average)
Proportion of schools assigned to treatment (P)	.50	.49

Note: Sample statistics should be interpreted with caution, as standard errors are not reported.
a. The student sample for science is about two-fifths that for math problem solving, because science outcomes are analyzed for grades 5 and 7 only. The smaller sample size leads to a slightly larger assumed minimum detectable effect size for science than for math problem solving.
b. Estimated using formula displayed in table X1.
c. Estimated as the ratio of the school-level variance component for the posttest to the sum of the variance components for schools and students in a two-level model that includes a fixed effect for treatment only.
Source: Student achievement data from tests administered as part of the state's accountability system (for observed sample statistics).

Table X3 Comparison of assumed parameter values and observed sample statistics for statistical power analysis associated with active learning in mathematics outcome after one year

Variable	Assumed parameter value (design phase)	Observed sample statistic (analysis phase)
Minimum detectable effect size	0.38	0.37[a]
School-level intraclass correlation (ρ)	.20	.12[b]
Proportion of between-school variance in outcome explained by covariates in model (R_c^2)	0	.14
Proportion of within-school variance in outcome explained by the covariates in model (R_i^2)	0	.00
Number of schools (J)	66	81
Number of teachers per school (n)	8	5
Proportion of schools assigned to treatment (P)	.50	.51

Note: Sample statistics should be interpreted with caution, as standard errors are not reported.
a. Calculated using formula in table X1.
b. Estimated as the ratio of the school-level variance component for outcome measure to the sum of the variance components for schools and teachers in two-level model that includes fixed effect for treatment only.
Source: Teacher survey data (for observed sample statistics).

Table X4 Comparison of assumed parameter values and observed sample statistics for statistical power analysis associated with active learning in science outcome after one year

Variable	Assumed parameter value (design phase)	Observed sample statistic (analysis phase)
Minimum detectable effect size	0.38	0.35[a]
School-level intraclass correlation (ρ)	.20	.13 [b]
Proportion of between-school variance in outcome explained by covariates in model (R_c^2)	0	.49
Proportion of within-school variance in outcome explained by the covariates in model (R_i^2)	0	.03
Number of schools (J)	66	78
Number of teachers per school (n)	8	5
Proportion of schools assigned to treatment (P)	.50	.51

Note: Sample statistics should be interpreted with caution, as standard errors are not reported.
a. Estimated using formula in table X1.
b. Estimated as the ratio of the school-level variance component for outcome measure to the sum of the variance components for schools and teachers in a two-level model that includes fixed effect for treatment only.
Source: Teacher survey data (for observed sample statistics).

Appendix Y. Parameter estimates for Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving after one year

This appendix (tables Y1–Y3) presents the full set of effect estimates from the analytic model described in chapter 2. They include the impact estimate (reported in results chapters) as well as the effects of covariates that were used in the models to increase precision. Also included are the effects of dummy variables used to indicate missing values for the covariates, the effects of dummy variables used to indicate matched pairs, and variance components for the random effects from the models. (The dummy variable method requires setting missing values for the covariates to 0; therefore, the effect estimates associated with the covariates are influenced by these substitutions). Tables in appendixes Z–AB, AI–AK, and AM–AO contain similar types of results.

Table Y1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after one year

Fixed effects model	Variance	Standard error	Degrees of freedom[a]	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	241.7	22.8	39	10.62	< .01
Adjusted average AMSTI effect across all schools	2.1	0.7	39	3.10	< .01
Average effect associated with school-level average pretest on student outcome across all schools	0.7	0.0	39	19.18	< .01
Average effect associated with student-level pretest deviation from school average pretest on student outcome across all schools[b]	0.7	0.0	19E3	37.78	< .01
Average effect associated with male gender on student outcome across all schools	0.3	0.4	19E3	0.64	.52
Average effect associated with eligibility for free or reduced-price lunch on student outcome across all schools[b]	–5.6	0.5	19E3	–11.13	< .01
Average effect associated with racial/ethnic status on student outcomes across all schools[b]	–3.9	0.5	19E3	–7.48	< .01
Average effect associated with English proficiency on student outcomes across all schools[b]	1.0	2.5	19E3	0.38	.71
Average effect of being in grade 4 (relative to grade 6) on student outcomes across all schools	–14.8	2.1	152	–6.90	< .01
Average effect of being in grade 5 (relative to grade 6) on student outcomes across all schools	–9.6	1.8	152	–5.33	< .01

Fixed effects model	Variance	Standard error	Degrees of freedom[a]	*t*-value	*p*-value
Average effect of being in grade 7 (relative to grade 6) on student outcomes across all schools	–4.5	1.5	152	–3.02	.03
Average effect of being in grade 8 (relative to grade 6) on student outcomes across all schools	7.6	1.5	152	4.97	< .01
Effect associated with dummy variable indicating missing value for student pretest	–6.1	1.8	19E3	–3.45	< .01
Effect associated with dummy variable indicating missing value for indicator of eligibility for free or reduced-price lunch	–17.3	13.9	19E3	–1.24	.21
Effect associated with dummy variable indicating missing value for indicator of racial/ethnic minority status	–4.1	4.6	19E3	–0.90	.37
Effect associated with dummy variable indicating missing value for indicator of English proficiency	1.6	4.7	19E3	0.34	.74

Note: Table excludes effect estimates for matched pairs.

a. Degrees of freedoms may be expressed as approximations (19E3 is equivalent to roughly 19,000).

b. The dummy variable approach to handling missing data involves setting missing values for covariates as constants. These effects are estimated with missing values set to zero; the effect estimate should be interpreted accordingly.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table Y2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for students within schools	602.0	6.2	96.48	< .01
Variance component for schools	14.4	4.1	3.53	< .01

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table Y3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t*-value	*p*-value
1	3.2	2.8	1.16	.25
2	–0.9	2.4	–0.39	.70
3	3.0	2.2	1.34	.19
4	9.7	2.7	3.65	< .01
5	–6.1	1.1	–5.48	< .01
6	8.7	1.8	4.76	< .01
7	–4.2	1.3	–3.34	< .01
8	–6.9	1.5	–4.52	< .01
9	–0.1	2.4	–0.04	.97
10	1.3	2.6	0.51	.62
11	–6.1	2.7	–2.23	.03
12	–6.8	2.5	–2.77	< .01
13	5.9	2.1	2.82	< .01
14	2.9	2.0	1.42	.16
15	1.3	1.5	0.87	.39
16	–1.0	1.6	–0.58	.56
17	–0.3	1.3	–0.22	.83
18	–4.9	2.6	–1.85	.07
19	4.0	3.8	1.04	.30
20	1.0	1.0	0.95	.35
21	–1.2	1.4	–0.89	.38
22	3.4	2.5	1.32	.20
23	5.1	4.2	1.24	.22
24	0.4	2.5	0.16	.88
25	7.7	3.0	2.61	.01
26	5.3	1.3	4.06	< .01
27	5.2	1.4	3.86	< .01
28	–0.8	3.7	–0.21	.84
29	10.1	2.6	3.82	< .01
30	4.7	3.8	1.22	.23
31	0.3	2.9	0.10	.92
32	1.4	5.0	0.28	.78
33	12.5	2.2	5.76	< .01
34	8.2	3.5	2.37	.02
35	2.7	1.7	1.59	.12
36	–18.0	2.3	–8.02	< .01
37	5.8	1.6	3.73	< .01
38	–1.5	2.6	–0.56	.58
39	–1.7	4.0	–0.44	.67
40	4.3	1.1	3.92	< .01

Note: The degrees of freedom associated with each estimate is 39.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix Z. Parameter estimates for Stanford Achievement Test Tenth Edition (SAT 10) science after one year

This appendix (tables Z1–Z3) presents various estimates from the model used to estimate the impact of AMSTI on science achievement after one year.

Table Z1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after one year

Fixed effects model	Variance	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	307.0	50.6	36	6.07	< .01
Adjusted average AMSTI effect across all schools	1.6	0.9	36	1.73	.09
Average effect associated with school-level average pretest on student outcome across schools[a]	0.5	0.1	36	6.78	< .01
Average effect associated with student-level pretest deviation from school average pretest on student outcome across schools	0.6	0.0	7,439	39.12	< .01
Average effect associated with male gender on student outcome across all schools	5.5	0.6	7,439	9.13	< .01
Average effect associated with eligibility for free or reduced-price lunch on student outcome across all schools[a]	–3.4	0.6	7,439	–5.67	< .01
Average effect associated with racial/ethnic status on student outcomes across all schools[a]	–4.7	0.7	7,439	–6.94	< .01
Average effect associated with English proficiency on student outcomes across all schools[a]	1.4	2.6	7,439	0.57	.57
Average effect of being in grade 5 (relative to grade 7) on student outcomes across all schools	–8.7	2.0	17	–4.27	< .01
Effect associated with dummy variable indicating missing value for pretest	–5.1	1.5	7,439	–3.52	< .01
Effect associated with dummy variable indicating missing value for indicator of eligibility for free or reduced-price lunch	–8.7	1.6	7,439	–5.61	< .01
Effect associated with dummy variable indicating missing value for indicator of racial/ethnic minority status	–5.0	5.7	7,439	–0.86	.39
Effect associated with dummy variable indicating missing value for indicator of English proficiency	6.4	6.9	7,439	0.93	.35

Note: Table excludes effect estimates for matched pairs.
a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant; these effects are estimated with missing values set to zero; therefore, the effect estimate should be interpreted accordingly.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table Z2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for students within schools	464.1	7.6	60.98	< .01
Variance component for schools	28.3	8.6	3.28	< .01

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table Z3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t*-value	*p*-value
1	3.6	2.4	1.49	.14
2	6.6	3.2	2.04	.05
3	13.7	3.5	3.97	< .01
4	12.7	3.4	3.71	< .01
5	1.2	0.8	1.49	.14
6	12.9	2.7	4.78	< .01
7	7.3	5.3	1.38	.18
8	–2.2	4.4	–0.50	.62
9	–1.5	0.7	–2.27	.03
10	0.3	1.4	0.18	.86
11	–0.8	3.5	–0.23	.82
12	1.9	2.4	0.79	.43
13	5.9	2.5	2.41	.02
14	2.0	2.9	0.71	.49
15	4.8	2.2	2.18	.04
16	11.3	4.9	2.31	.03
17	6.8	3.2	2.10	.04
18	4.4	0.7	5.97	< .01
19	5.4	2.4	2.30	.03
20	5.3	2.6	2.02	.05
21	4.0	3.4	1.19	.24
22	8.4	4.7	1.79	.08
23	7.5	2.8	2.66	.01
24	9.9	3.5	2.80	< .01
25	6.9	2.2	3.14	< .01
26	17.9	3.6	5.03	< .01
27	4.3	2.2	1.98	.06
28	4.4	7.2	0.61	.55
29	10.0	2.2	4.50	< .01
30	8.6	2.1	4.11	< .01
31	–6.8	4.2	–1.62	.11
32	7.8	6.2	1.26	.22
33	11.8	5.1	2.30	.03
34	1.5	2.8	0.52	.61
35	10.8	2.9	3.74	< .01

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t*-value	*p*-value
36	–12.1	8.9	–1.35	.19
37	12.8	2.4	5.36	< .01
38	–0.9	2.2	–0.42	.68
39	1.5	1.6	0.94	.36
40	6.7	2.5	2.71	.01

Note: The degrees of freedom associated with each estimate is 36.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AA. Parameter estimates for active learning in mathematics after one year

This appendix (tables AA1–AA3) presents various estimates of the effect of AMSTI on teaching of active learning in mathematics after one year.

Table AA1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) effect on teaching for active learning in mathematics after one year

Fixed effects model	Variance	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	171.0	91.0	39	1.88	.07
Adjusted average AMSTI effect across all schools	49.8	11.5	39	4.34	< .01
Average effect associated with degree rank 0 on teacher outcome across all schools[a]	–20.9	22.7	318	–0.92	.36
Average effect associated with degree rank 1 on teacher outcome across all schools[a]	–28.4	20.1	318	–1.41	.16
Average effect associated with years teaching experience on teacher outcome across all schools[a]	–3.4	1.2	318	–2.78	< .01
Average effect associated with years teaching experience with math on teacher outcome across all schools[a]	3.1	1.4	318	2.23	.03
Effect associated with dummy variable indicating missing value for degree rank	–1.9	54.1	318	–0.03	.97
Effect associated with dummy variable indicating missing value for years teaching experience (also indicates missing value for years of teaching experience in subject area)	–47.2	63.2	318	–0.75	.46

Note: Table excludes effect estimates for matched pairs.

a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant. These effects are estimated with missing values set to zero; the effect estimate should be interpreted accordingly.

Source: Teacher survey data.

Table AA2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teaching for active learning in mathematics after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for teachers within schools	14,965.0	1,182.2	12.66	< .01
Variance component for schools	1,661.4	1,279.4	1.30	.10

Source: Teacher survey data.

Table AA3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning in mathematics after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41002)	Standard error	*t*–value	*p*–value
1001	41.1	102.0	0.40	.69
2001	8.7	97.4	0.09	.93
3001	16.2	89.3	0.18	.86
4001	36.0	89.4	0.40	.69
5001	18.3	101.5	0.18	.86
6001	–45.8	89.0	–0.51	.61
7001	–45.4	89.5	–0.51	.62
8001	4.8	88.9	0.05	.96
9001	116.3	129.8	0.90	.38
10001	–67.5	90.9	–0.74	.46
11001	49.7	119.8	0.41	.68
12001	1.2	90.0	0.01	.99
13001	57.1	91.6	0.62	.54
14001	42.1	95.5	0.44	.66
15001	–42.0	98.6	–0.43	.67
16001	21.6	92.9	0.23	.82
17001	–46.3	89.9	–0.51	.61
18001	133.5	110.6	1.21	.24
19001	9.8	90.1	0.11	.91
20001	–21.3	92.8	–0.23	.82
21002	–71.4	95.4	–0.75	.46
22002	–72.3	89.4	–0.81	.42
23002	–40.8	98.7	–0.41	.68
24002	–27.3	91.9	–0.30	.77
25002	–52.7	90.1	–0.59	.56
26002	–70.2	94.0	–0.75	.46
27002	–20.5	94.5	–0.22	.83
28002	–67.4	89.1	–0.76	.45
29002	–39.2	92.8	–0.42	.68
30002	–108.4	90.9	–1.19	.24
31002	45.1	90.6	0.50	.62
32002	–2.3	89.8	–0.03	.98
33002	–7.9	90.0	–0.09	.93
34002	–45.1	121.9	–0.37	.71
35002	7.7	97.0	0.08	.94
36002	45.1	105.7	0.43	.67
37002	–36.9	89.4	–0.41	.68
38002	–85.9	90.1	–0.95	.35
39002	–105.2	88.9	–1.18	.24
40002	–32.7	91.9	–0.36	.72

Note: The degrees of freedom associated with each estimate is 39.
Source: Teacher survey data.

Appendix AB. Parameter estimates for active learning in science after one year

This appendix (tables AB1–AB3) presents various estimates of the effect of AMSTI on teaching of active learning in science after one year.

Table AB1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teaching for active learning in science after one year

Fixed effects model	Variance	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	196.7	73.5	36	2.68	.01
Adjusted average AMSTI effect across all schools	40.1	11.8	36	3.41	< .01
Average effect associated with degree rank 0 on teacher outcome across all schools[a]	–15.0	21.2	285	–0.71	.48
Average effect associated with degree rank 1 on teacher outcome across all schools[a]	–25.8	15.5	285	–1.67	.10
Average effect associated with years teaching experience on teacher outcome across all schools[a]	–3.0	1.3	285	–2.29	.02
Average effect associated with years teaching experience with science on teacher outcome across all schools[a]	4.6	1.5	285	3.19	< .01
Effect associated with dummy variable indicating missing value for degree rank	–83.6	22.3	285	–3.75	< .01
Effect associated with dummy variable indicating missing value for years teaching experience (also indicates missing value for years of teaching experience in subject area)	37.2	29.3	285	1.27	.21

Note: This table excludes effect estimates for matched pairs.
a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant. These effects are estimated with missing values set to zero; the effect estimate should be interpreted accordingly.
Source: Teacher survey data.

Table AB2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teaching for active learning in science after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for teachers within schools	14,479.0	1,203.8	12.03	< .01
Variance component for schools	1,171.3	1,118.9	1.05	.15

Source: Teacher survey data.

Table AB3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning in science after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41002)	Standard error	*t*-value	*p*-value
1001	4.8	94.9	0.05	.96
2001	–92.8	76.9	–1.21	.24
3001	34.8	75.0	0.46	.65
4001	–52.5	77.2	–0.68	.50
5001	28.5	73.1	0.39	.70
6001	35.1	76.9	0.46	.65
7001	–101.7	75.9	–1.34	.19
8001	–89.5	92.2	–0.97	.34
9001	–28.7	85.0	–0.34	.74
10001	–75.5	76.2	–0.99	.33
11001	150.2	101.2	1.48	.15
12001	–59.3	77.5	–0.77	.45
13001	–64.0	91.0	–0.70	.49
14001	–98.1	82.6	–1.19	.24
15001	–79.4	79.5	–1.00	.33
16001	–65.6	78.5	–0.83	.41
17001	–78.7	76.1	–1.03	.31
18001	–18.1	73.3	–0.25	.81
19001	145.8	73.6	1.98	.06
20001	–75.3	79.9	–0.94	.35
21002	–109.5	75.6	–1.45	.16
22002	–39.7	81.7	–0.49	.63
23002	–83.7	85.6	–0.98	.34
24002	–17.3	107.3	–0.16	.87
25002	–111.2	73.6	–1.51	.14
26002	–68.6	77.6	–0.88	.38
27002	–101.1	74.9	–1.35	.19
28002	117.1	75.8	1.55	.13
29002	–38.5	73.7	–0.52	.61
30002	–109.7	76.5	–1.43	.16
31002	–45.4	85.4	–0.53	.60
32002	–93.2	74.7	–1.25	.22
33002	–65.4	73.5	–0.89	.38
34002	–116.1	82.7	–1.40	.17
35002	–105.7	86.2	–1.23	.23
36002	–2.2	79.3	–0.03	.98
37002	–52.4	74.3	–0.70	.49
38002	–69.9	75.5	–0.92	.36
39002	33.7	74.5	0.45	.65
40002	–131.8	73.6	–1.79	.08

Note: The degrees of freedom associated with each estimate is 36.
Source: Teacher survey data.

Appendix AC. Sensitivity analyses of effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement after one year

Below is the result of sensitivity analyses of the one-year effect of AMSTI on SAT 10 mathematics problem solving achievement (table AC1).

Table AC1 Sensitivity analyses of the one-year effect of Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving achievement

Model	Effect estimate	Standard error	*p*-value
Benchmark	2.1	0.7	.004†
Gain score used as outcome variable	1.9	0.8	.020†
Pretest and pairs are only covariates (cases without a pretest were listwise deleted)	1.4	0.8	.087
Pretest and pairs are only covariates (dummy variable approach to handling missing data used with the benchmark model was applied to missing pretests)	1.9	0.8	.022†
Listwise deletion of cases with a missing value for any covariate	1.7	0.7	.016†
Maximum likelihood instead of restricted maximum likelihood estimation	2.0	0.7	.004†
Grade levels weighted equally	2.0	0.7	.007†
Subexperiments weighted equally	2.1	0.7	.003†
Reading pretest included as additional covariate	2.1	0.7	.005†
Schools weighted equally in model 1[a]	2.5	0.8	.003†
Schools weighted equally in model 2[a]	2.5	0.7	.001†

†Remains significant after adjusting for multiple comparisons. Adjusted significance level is .025.
a. For both model 1 and model 2, school-average posttests were regressed against school average values of the covariates used in the benchmark model. These analyses therefore give equal weight to schools in the sample. In contrast, the benchmark model implicitly weights schools by the inverses of the variances associated with them. In model 1 for each covariate, an additional covariate was included to indicate the proportion of students in a school with a missing value for that covariate. In model 2 the covariates indicating the school proportions of students with missing values were not included.
Source: Students achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AD. Sensitivity analyses of effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement after one year

Below is the result of sensitivity analyses of the one-year effect of AMSTI on SAT 10 science achievement (table AD1).

Table AD1 Sensitivity analyses for one-year effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on Stanford Achievement Test Tenth Edition (SAT 10) science achievement

Model	Effect estimate	Standard error	*p*-value
Benchmark	1.6	0.9	.092
Pretest and pairs are only covariates (cases without pretest listwise deleted)	1.0	0.9	.273
Pretest and pairs are only covariates (dummy variable approach to handling missing data used with benchmark model applied to missing pretests)	1.2	1.0	.233
Listwise deletion of cases with missing value for any covariate	1.3	0.9	.139
Maximum likelihood instead of restricted maximum likelihood estimation	1.5	0.9	.106
Grade-levels weighted equally	1.6	1.0	.114
Subexperiments weighted equally	1.6	0.9	.098
Math problem solving pretest included as an additional covariate	1.6	0.9	.093
Schools weighted equally in model 1[a]	3.2	1.2	.011†
Schools weighted equally in model 2[a]	2.0	1.1	.067

†Remains significant after adjusting for multiple comparisons. Adjusted significance level is .025.

a. For both model 1 and model 2, school-average posttests were regressed against school average values of the covariates used in the benchmark model. These analyses therefore give equal weight to schools in the sample. In contrast, the benchmark model implicitly weights schools by the inverses of the variances associated with them. In model 1 for each covariate, an additional covariate was included to indicate the proportion of students in a school with a missing value for that covariate. In model 2 the covariates indicating the school proportions of students with missing values were not included.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AE. Sensitivity analyses of effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms after one year

Below is the result of sensitivity analyses of the one-year effect of AMSTI on active learning instructional strategies in mathematics classrooms (table AE1).

Table AE1 Sensitivity analyses for one-year effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in mathematics classrooms

Model	Effect estimate	Standard error	*p*-value
Benchmark	49.8	11.5	< .001†
Listwise deletion of cases with missing value for any covariate	50.8	11.5	< .001†
Maximum likelihood instead of restricted maximum likelihood estimation[a]	46.9	11.3	< .001†
Subexperiments weighted equally	48.0	11.2	< .001†
Schools weighted equally in model 1[b]	58.7	15.1	< .001†
Schools weighted equally in model 2[b]	53.5	14.2	< .001†
Teachers with responses to fewer than four items removed[c]	54.1	10.4	< .001†

†Remains significant after adjusting for multiple comparisons. Adjusted significance level is .025.
a. The school-level random effect could not be estimated with this model.
b. For both model 1 and model 2, school average posttests were regressed against school average values of the covariates used in the benchmark model. These analyses therefore give equal weight to schools in the sample. In contrast, the benchmark model implicitly weights schools by the inverses of the variances associated with those schools. In model 1 for each covariate, an additional covariate to indicate the school proportion of teachers with a missing value for that covariate was included. In model 2 the covariates indicating school proportions of teachers with missing values were not included.
c. Forty-five of 405 teachers responded to fewer than four items and were excluded from this analysis.
Source: Teacher survey data.

Appendix AF. Sensitivity analyses of effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms after one year

Below is the result of sensitivity analyses of the one-year effect of AMSTI on active learning instructional strategies in science classrooms (table AF1).

Table AF1 Sensitivity analyses for one-year effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on active learning instructional strategies in science classrooms

Model	Effect estimate	Standard error	*p*-value
Benchmark	40.1	11.8	.002†
Listwise deletion of cases with missing value for any covariate	40.3	12.3	.002†
Maximum likelihood instead of restricted maximum likelihood estimation[a]	40.7	12.0	.002†
Subexperiments weighted equally	39.8	11.6	.001†
Schools weighted equally in model 1[b]	37.3	12.3	.005†
Schools weighted equally in model 2[b]	41.6	11.8	.001†
Teachers with responses to fewer than four items removed[c]	37.7	12.5	.005†

†Remains significant after adjusting for multiple comparisons. Adjusted significance level is .025.
a. The school-level random effect could not be estimated with this model.
b. For both model 1 and model 2, school average posttests were regressed against school average values of the covariates used in the benchmark model. These analyses therefore give equal weight to schools in the sample. In contrast, the benchmark model implicitly weights schools by the inverses of the variances associated with those schools. In model 1 for each covariate, an additional covariate to indicate the school proportion of teachers with a missing value for that covariate was included. In model 2 the covariates indicating school proportions of teachers with missing values were not included.
c. Thirty-three of 369 teachers responded to fewer than four items and were excluded from this analysis.
Source: Teacher survey data.

Appendix AG. Tests for violations of factors associated with assumption of equal first year effects on students in Alabama Math, Science, and Technology Initiative (AMSTI) and control schools

As demonstrated in appendix Q, the Bell-Bradley method used to compute two-year impact estimates provides unbiased estimates of impact if the effect of AMSTI after the first year of implementation is stable over time—that is, if initial exposure to AMSTI for treatment group students in Year 1 had the same effect as initial exposure to AMSTI for control group students in Year 2. This assumption depends on a number of factors related to the intervention, the participants, and the context in which the intervention was implemented (see list in box AG1 below). Some of these factors can be tested using evaluation data; others cannot. This appendix provides the results of 66 such tests, related to nine of those factors, to examine the plausibility of the core assumption.[125] Of the 66 tests conducted, 61 suggest the assumption is true, and 5 indicate ways in which it may be violated. Taken in total, the tests suggest that it is reasonable to rely on the assumption and apply the Bell-Bradley method to the AMSTI data.

Below (box AG1) is the evidence from the different tests, organized around the 11 factors introduced in chapter 2 that underlie the assumption of equal impacts in the first year of implementation for both the treatment and control groups. The rest of the appendix examines the test evidence in greater detail.

Box AG1 Factors that need to remain stable over time for effect of the Alabama Math, Science, and Technology Initiative (AMSTI) to be the same in the first year of implementation in both groups of schools (treatment and control) that received it
Intervention-related factors
Sponsor's guidelines for required parameters of intervention's design and implementation Professional development unchanged; alignment with state content standards unchanged.
District's desire and evolving ability to support the intervention Science kit rotation unchanged.
Factors that could be affected by random assignment to the control group
Composition of schools in sample that did not implement intervention when the time came to do so Two of 18 indicators of school characteristics had a statistically significant difference for the SAT 10 mathematics problem solving sample; 1 of 13 indicators of school characteristics had a statistically significant difference for the SAT 10 science sample.
Composition of teachers in implementing schools that did not participate in intervention or data collection when the time came to do so Same shares of teachers did not implement.

[125] These are all of the tests that can be constructed from data available to the evaluation.

Effort by schools to implement intervention One of 12 indicators of school effort differed significantly.
Contextual factors
Characteristics of age cohorts[a] One of 12 indicators tested had a statistically significant difference for the SAT 10 mathematics problem solving sample; none of the 7 indicators tested had statistically significant difference for the SAT 10 science sample.
Alternative programs and schools available in the community[a] See result under "Characteristic of age cohorts."
Existing school services apart from the intervention[a] See result under "Characteristic of age cohorts."
Configuration of courses and curricula taught in school, and system used by school to allocate students at a given grade level to courses or curricula options[a] See result under "Characteristic of age cohorts."
Characteristics of potential teachers in the community Not assessed.
Teacher hiring and course/class assignment practices Not assessed.

a. A single test was performed to test the net effect on the characteristics of students in the AMSTI and control group samples.
Source: Review of evaluation's professional development reports, review of percentage of state content covered by AMSTI print materials, examination of factors associated with efforts to implement AMSTI, examination of differences between control schools that implemented and those that did not, and examination of student characteristics in first year of implementation.

There could be changes in unobserved factors over time that affect the internal validity of the method. For example, researchers did not collect information on characteristics of potential teachers in the community, or on teacher hiring practices and assignment practices.

Sponsor's guidelines for the parameters of the intervention

If the sponsor changed the intervention's design or parameters between the two years, there would be no reason to expect effects in the first year of implementation in the AMSTI schools to be the same as in the second year, when control schools implemented AMSTI for the first time. A review of two professional development reports from the AMSTI evaluation (Sawyer et al. 2009; Sawyer et al. 2010) suggests that the AMSTI professional development received by teachers was similar programmatically (hours of training, types of materials used, topic coverage, and trainer teaching methods). However, AMSTI math and science specialists in the Alabama State Department of Education told the research team that they were making changes to the AMSTI curriculum in order to address 100 percent of the Alabama Course of Study (state content standards). Comparison of AMSTI printed materials with the Alabama Course of Study revealed that curriculum coverage for the initial year of the intervention was constant for the duration of the study.[126] The review uncovered no differences in the coverage of the state content standards between the first year and the second year.

District support

Implementation of AMSTI involves district support; differences in this support could lead to different impacts of AMSTI in the first year of implementation. District support for AMSTI could be provided several ways. Data on whether the schools in each sample received the same number and types of science kits during their first year of AMSTI were examined to determine whether district support had changed. Review of the kit rotation schedules provided for each AMSTI region did not reveal any evidence that the kits received by schools in their first year of AMSTI implementation differed between the intervention and control groups. Other forms of district-level support were not observed. They may or may not have changed between the two years.

Composition of schools that did not implement Alabama Math, Science, and Technology Initiative (AMSTI)

Random assignment ensures that there were no systematic differences between treatment and control group schools in their background characteristics. If all treatment schools and all control schools implemented AMSTI, the characteristics of implementing schools should also be equivalent, as required by the Bell-Bradley estimator. However, three control schools did not implement AMSTI. Therefore, it is necessary to examine whether, and how, the control schools that failed to implement AMSTI were different from control schools that implemented AMSTI, in order to determine whether the omission of these schools from the control group implementation sample results in a sample that does not match the treatment group implementation sample (which consisted of all treatment group schools). Comparison of characteristics of implementing and nonimplementing control group schools determine whether the implementing subset is typical of the whole control group—and hence of the entire treatment group—on measured variables.[127]

Nine characteristics were assessed for the SAT 10 mathematics problem solving sample (table AG1). Teacher characteristics included degree rank, total years of teaching experience, and years of teaching experience in subject area. Student characteristics included the percentages of boys, minority students, students proficient in English, students enrolled in the free or reduced-price lunch program, and students in specific grades as well as mean pretest scores. A joint significance test of all covariates did not reject the null hypothesis that no differences exist between the samples on these characteristics (p = .26). For two of the nine characteristics, the difference between control group schools that implemented AMSTI and control group schools that did not implement AMSTI was statistically significant: nonimplementing schools had a

[126] Only the initial year of implementation matters to the reliability of the Bell-Bradley estimation technique.
[127] Differences may still exist with regard to unmeasured characteristics.

smaller proportion of students enrolled in the free or reduced-price lunch program (20.56 percent compared with 68.22 percent, $p < .01$) and higher school average pretests (661.80 compared with 638.10, $p = .05$). For the SAT 10 science outcome, the same characteristics were examined. Nonimplementing schools were found to have had a smaller proportion of students enrolled in the free or reduced-price lunch program (21.92 percent compared with 68.72 percent, $p < .01$). Differences in neither the joint significance test ($p = .11$) nor any other differences were statistically significant.[128, 129]

[128] Statistical tests of 31 indicators were conducted, 18 for the SAT 10 mathematics problem solving sample and 13 for the SAT10 science sample.

[129] Table AG1 shows incomplete data are for 10 of the tests (5 for mathematics and 5 for science), the teacher characteristics variables, which were obtained from the web-based surveys. One of the three nonimplementing control schools did not participate in the surveys. Teacher characteristic information on nonimplementing schools cannot be displayed for this group without risking the confidentiality of the two schools in the group that did complete the surveys. Student characteristic data were collected from all three nonimplementing schools.

Table AG1 Mean characteristics of nonimplementing and implementing control group schools

		SAT 10 mathematics problem solving sample			SAT 10 science sample		
Baseline characteristic		**Nonimplementing schools**		**Implementing schools**	**Nonimplementing schools**		**Implementing schools**
Teacher characteristic							
Average of school percent of out-of-field teachers	Average in Each Condition	‡		25.2	‡		12.5
	Standard Deviation	‡		29.5	‡		30.2
	Sample Size (Schools)	‡		37	‡		36
	Average Difference		‡			‡	
	Standard Error		‡			‡	
	Test statistic		*t* = 1.43			*t* = 0.19	
	***p*-value**		.16			.85	
Average of school percent of teachers with one degree in teaching content area	Average in Each Condition	‡		57.1	‡		60.7
	Standard Deviation	‡		31.0	‡		41.0
	Sample Size (Schools)	‡		37	‡		36
	Average Difference		‡			‡	
	Standard Error		‡			‡	
	Test statistic		*t* = 1.07			*t* = 0.77	
	***p*-value**		.29			.45	
Average of school percent of teachers with two or more degrees in content area	Average in Each Condition	‡		17.6	‡		26.9
	Standard Deviation	‡		22.4	‡		38.2
	Sample Size (Schools)	‡		37	‡		36
	Average Difference		‡			‡	
	Standard Error		‡			‡	
	Test statistic		*t* = 0.40			*t* = 0.98	
	***p*-value**		.69			.33	
Average of school percent of teachers with less than four years' total teaching experience	Average in Each Condition	‡		24.9	‡		34.3
	Standard Deviation	‡		24.9	‡		38.8
	Sample Size (Schools)	‡		37	‡		36
	Average Difference		‡			‡	
	Standard Error		‡			‡	
	Test statistic		*t* = 0.46			*t* = 0.63	
	***p*-value**		.65			.53	
Average of school percent of teachers with less than four years' teaching experience in	Average in Each Condition	‡		30.2	‡		31.7
	Standard Deviation	‡		27.6	‡		35.5
	Sample Size (Schools)	‡		37	‡		36
	Average Difference		‡			‡	
	Standard Error		‡			‡	

		SAT 10 mathematics problem solving sample			SAT 10 science sample		
Baseline characteristic		**Nonimplementing schools**		**Implementing schools**	**Nonimplementing schools**		**Implementing schools**
subject area	**Test statistic**		*t* = 0.16			*t* = 0.59	
	***p*-value**		.87			.56	
Student characteristic							
Average of school percent of boys	Average in Each Condition	50.6		49.5	51.8		51.1
	Standard Deviation	4.2		4.2	5.4		6.5
	Sample Size (Schools)	3		38	3		38
	Average Difference		1.1			0.7	
	Standard Error		2.5			3.9	
	Test statistic		*t* = 0.44			*t* = 0.19	
	***p*-value**		.66			.85	
Average of school percent of minority students	Average in Each Condition	18.1		49.1	17.9		49.6
	Standard Deviation	10.9		33.4	10.9		33.5
	Sample Size (Schools)	3		38	3		38
	Average Difference		–31.1			–31.7	
	Standard Error		19.5			19.6	
	Test statistic		*t* = 1.59			*t* = 1.61	
	***p*-value**		.12			.11	
Average of school percent of students proficient in English	Average in Each Condition	97.3		98.7	97.5		98.3
	Standard Deviation	1.9		2.4	1.7		2.8
	Sample Size (Schools)	3		38	3		38
	Average Difference		–1.3			-0.8	
	Standard Error		1.5			1.6	
	Test statistic		*t* = 0.93			*t* = 0.52	
	***p*-value**		.36			.61	
Average of school percent of students enrolled in the free or reduced-price lunch program	Average in Each Condition	20.6		68.2	21.9		68.7
	Standard Deviation	19.1		21.0	20.1		21.4
	Sample Size (Schools)	3		38	3		38
	Average Difference		–47.7			-46.8	
	Standard Error		12.5			12.8	
	Test statistic		*t* = 3.80			*t* = 3.66	
	***p*-value**		< .01			< .01	
Average of school percent of students in grade 4	Average in Each Condition	22.9		22.4	na		na
	Standard Deviation	19.9		21.0	na		na
	Sample Size (Schools)	3		38	na		na
	Average Difference		0.5			na	
	Standard Error		12.5			na	

		SAT 10 mathematics problem solving sample			SAT 10 science sample		
Baseline characteristic		**Nonimplementing schools**		**Implementing schools**	**Nonimplementing schools**		**Implementing schools**
	Test statistic		*t* = –0.04			na	
	***p*-value**		.97			na	
Average of school percent of students in grade 5	Average in Each Condition	21.2		23.4	66.7		56.8
	Standard Deviation	18.4		19.8	57.7		45.5
	Sample Size (Schools)	3		38	3		38
	Average Difference		–2.2			9.9	
	Standard Error		11.9			27.7	
	Test statistic		*t* = 0.19			*t* = 0.36	
	***p*-value**		.85			.72	
Average of school percent of students in grade 6	Average in Each Condition	33.8		17.6	na		na
	Standard Deviation	2.0		16.7	na		na
	Sample Size (Schools)	3		38	na		na
	Average Difference		16.2			na	
	Standard Error		9.8			na	
	Test statistic		*t* = –1.66			na	
	***p*-value**		.11			na	
Average of school percent of students in grade 7	Average in Each Condition	11.4		18.6	33.3		43.2
	Standard Deviation	19.7		20.8	57.7		45.5
	Sample Size (Schools)	3		38	3		38
	Average Difference		–7.3			na	
	Standard Error		12.4			na	
	Test statistic		*t* = 0.58			na	
	***p*-value**		.56			na	
Average of school percent of students in grade 8	Average in Each Condition	10.7		18.0	na		na
	Standard Deviation	18.6		19.4	na		na
	Sample Size (Schools)	3		38	na		na
	Average Difference		–7.2			na	
	Standard Error		11.6			na	
	Test statistic		*t* = 0.62			na	
	***p*-value**		.54			na	
School average pretest score and sample size							
SAT10[a]	Pretest Score	661.8		638.1	662.9		648.0
	Standard Deviation	23.6		19.1	13.9		15.4
	Sample Size (Schools)	3		38	3		38
	Average Difference		23.7			14.9	
	Standard Error		11.6			9.2	

		SAT 10 mathematics problem solving sample			SAT 10 science sample		
Baseline characteristic		**Nonimplementing schools**		**Implementing schools**	**Nonimplementing schools**		**Implementing schools**
	Test statistic		*t* = 2.04			*t* = –1.62	
	***p*-value**		.05			.11	
Sample Size	Number of schools = 3 Number of teachers = 31 Number of students = 1,486		Number of schools = 38 Number of teachers = 198 Number of students = 7,623		Number of schools = 3 Number of teachers = 7 Number of students = 496		Number of schools = 38 Number of teachers = 88 Number of students = 3,192

na is not applicable.
‡ Data values for groups of two schools or less, or information from which such data could be inferred, are not provided, in order to protect school confidentiality.
Note: Detail may not sum to totals because of rounding. The number of schools, teachers, and students for the comparisons varied slightly, depending on whether a characteristic was reported. For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Grade level for mathematics problem solving and teacher degree rank are categorical variables with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of students across grades for the mathematics problem solving outcome (p = .78) and teacher degree rank (p = .16 for mathematics problem solving, p = .32 for science). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the non-implementing group. The second modeled the log odds of belonging to the non-implementing group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the *p*-value was .26 for mathematics problem solving and .11 for science.)
a. The SAT 10 mathematics problem solving pretest was used for the mathematics outcome. The SAT 10 reading pretest was used for the science outcome.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system and teacher survey data.

Characteristics of teachers who did not participate in Alabama Math, Science, and Technology Initiative (AMSTI)

No data are available on the characteristics of nonparticipating teachers. However, the proportion of teachers that participated in some form of AMSTI training was almost identical the first year the program came to their schools (86 percent of teachers in schools that introduced AMSTI in Year 1 and 85 percent of teachers in schools that introduced AMSTI in Year 2). These data suggest that nonparticipation may have been driven by similar factors in the two groups.

Efforts by schools to implement Alabama Math, Science, and Technology Initiative (AMSTI)

The intervention and control schools could have differed in the level or type of effort put forth when implementing AMSTI in its first year. If this were the case, the effect for the control schools after one year of implementation may not be equal to the impact for the AMSTI schools in their first year of implementation.

Researchers used data collected from teacher surveys to assess whether the commitment to implementing AMSTI differed significantly in the AMSTI and control groups in their first years of implementation. The lack of significant differences for 11 of 12 indicators of effort tested suggests that the schools generally did not differ in the implementation of AMSTI during the first year (table AG2). The one significant difference found (in responses to the question “so far this school year, how much AMSTI professional development have you received for your science program?”) may be due to chance.

Table AG2 Implementation of the Alabama Math, Science, and Technology Initiative (AMSTI) by Alabama Math, Science, and Technology Initiative and control group schools during first year of Alabama Math, Science, and Technology Initiative intervention

Question	Number of teachers in AMSTI schools, Year 1 (first year of implementing AMSTI)	Number of teachers in control schools, Year 2 (first year of implementing AMSTI)	Estimate of difference	*p*-value
How much AMSTI professional development have you received for your math program? (number of hours of AMSTI summer training)	125	122	4.3 (4.6)	.36
How much AMSTI professional development have you received for your science program? (number of hours of AMSTI summer training)	114	116	19.1 (14.2)	.19
So far this school year, how much AMSTI professional development have you received for your math program?	206	156	0.5 (0.4)	.21
So far this school year, how much AMSTI professional development have you received for your science program?	184	145	1.0 (0.3)	< .01
So far this school year, how many times did you try contacting someone for support (for example, mentoring or coaching) with AMSTI math instruction?	207	161	0.2 (0.3)	.54
So far this school year, how many times did you try contacting someone for support (for example, mentoring or coaching) with AMSTI science instruction?	190	148	0.5 (0.3)	.07
So far this year, how many times did someone actually provide support (for example, mentoring or coaching) with AMSTI math instruction?	207	161	0.6 (0.3)	.09
So far this year, how many times did someone actually provide support (for example, mentoring or coaching) with AMSTI science instruction?	190	148	0.6 (0.3)	.06
Think back on your last two weeks (10 full days) of instruction: approximately how many minutes did your students spend doing math in your class? Please be sure to consider all activities, including discussion, lecture, reading, watching video, hands-on activities, and activities that integrate math with other subjects.	209	162	–17.7 (21.0)	.41

Question	Number of teachers in AMSTI schools, Year 1 (first year of implementing AMSTI)	Number of teachers in control schools, Year 2 (first year of implementing AMSTI)	Estimate of difference	*p*-value
Think back on your last two weeks (10 full days) of instruction: approximately how many minutes did your students spend doing science in your class? Please be sure to consider all activities, including discussion, lecture, reading, watching video, hands-on activities, and activities that integrate math with other subjects.	194	147	–12.4 (21.6)	.57
During the past two weeks, about how much time did you teach using AMSTI supplied mathematics print materials?	208	162	0.0 (0.0)	.62
During the past two weeks, about how much time did you teach using AMSTI supplied science print materials?	189	153	0.0 (0.1)	.08

Note: Numbers in parentheses are standard errors. Estimates were generated from analytic models that used the factor as the outcome.
Source: Teacher survey data.

Characteristics of student cohorts

The student populations participating in AMSTI in its first year need to be matched between AMSTI and control group schools to ensure that the intervention was provided to similar students. To check this condition, researchers examined the six baseline student characteristics available in the study data, namely, the percentages of boys, minority students, students proficient in English, students enrolled in the free or reduced-price lunch program, and students in specific grades as well as mean pretest scores.[130]

Below are compared student and school baseline characteristics for AMSTI schools (and students) when they were first exposed to AMSTI (Year 1) and control schools (and students) when they were first exposed to AMSTI (Year 2) (table AG3). As the table illustrates, for the SAT 10 mathematics problem solving sample, the AMSTI and control groups differed to a statistically significant extent on one or more of the variables examined by the joint test ($p < .01$). The similarities in the set of variables for the AMSTI versus control group are thus not strong enough to accept the null hypothesis of no difference. However, when the same background variables are considered one by one, the evidence never reaches this same level of certainty that the null of no difference must be rejected. For the SAT 10 science sample, neither the joint significance test ($p = .96$) nor any other background variables differed at a statistically significant level.

[130] Statistical tests of 19 indicators were conducted, 12 for the SAT 10 mathematics problem solving sample and 7 for the SAT 10 science sample.

Table AG3 Mean background characteristics of students in first year of Alabama Math, Science, and Technology Initiative (AMSTI) implementation in Year 1 and Year 2

		SAT 10 mathematics problem solving outcome			SAT 10 science outcome		
Baseline characteristic		**AMSTI schools (Year 1)**		**Control schools (Year 2)**	**AMSTI schools (Year 1)**		**Control schools (Year 2)**
Student characteristic							
Average of school percent of boys	Average in Each Condition	49.1		49.5	48.0		49.1
	Standard Deviation	4.1		3.8	5.8		5.3
	Sample Size (Schools)	41		41	39		40
	Average Difference		–0.5			–1.2	
	Standard Error		0.9			1.3	
	Test statistic		*t* = 0.53			*t* = 0.93	
	***p*-value**		.60			.36	
Average of school percent of minority students	Average in Each Condition	51.1		47.1	51.6		45.9
	Standard Deviation	34.6		33.4	35.4		33.4
	Sample Size (Schools)	41		41	39		40
	Average Difference		4.0			5.7	
	Standard Error		7.5			7.7	
	Test statistic		*t* = 0.53			*t* = 0.73	
	***p*-value**		.60			.46	
Average of school percent of students proficient in English	Average in Each Condition	98.3		98.7	98.4		98.9
	Standard Deviation	3.4		2.1	3.1		2.2
	Sample Size (Schools)	41		41	39		40
	Average Difference		–0.5			–0.5	
	Standard Error		0.6			0.6	
	Test statistic		*t* = 0.72			*t* = 0.77	
	***p*-value**		.48			.44	
Average of school percent of students enrolled in the free or reduced-price lunch program	Average in Each Condition	63.1		65.2	65.5		63.4
	Standard Deviation	24.9		24.1	26.8		24.9
	Sample Size (Schools)	41		41	39		40
	Average Difference		–2.2			2.1	
	Standard Error		5.4			5.8	
	Test statistic		*t* = 0.40			*t* = 0.36	
	***p*-value**		.69			.72	
Average of school percent of students in grade 4	Average in Each Condition	25.8		23.3	na		na
	Standard Deviation	24.8		23.1	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		2.6			na	

		SAT 10 mathematics problem solving outcome			SAT 10 science outcome		
Baseline characteristic		**AMSTI schools (Year 1)**		**Control schools (Year 2)**	**AMSTI schools (Year 1)**		**Control schools (Year 2)**
	Standard Error		5.3			na	
	Test statistic		*t* = 0.48			na	
	***p*-value**		.63			na	
Average of school percent of students in grade 5	Average in Each Condition	25.2		23.4	61.2		56.0
	Standard Deviation	19.2		20.5	40.7		46.6
	Sample Size (Schools)	41		41	39		40
	Average Difference		1.9			5.2	
	Standard Error		4.4			9.9	
	Test statistic		*t* = 0.42			*t* = 0.53	
	***p*-value**		.67			.60	
Average of school percent of students in grade 6	Average in Each Condition	13.3		18.1	na		na
	Standard Deviation	14.3		16.6	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		–4.8			na	
	Standard Error		3.4			na	
	Test statistic		*t* = 1.39			na	
	***p*-value**		.17			na	
Average of school percent of students in grade 7	Average in Each Condition	18.4		17.5	38.8		44.0
	Standard Deviation	21.6		19.6	40.7		46.6
	Sample Size (Schools)	41		41	39		40
	Average Difference		1.0			na	
	Standard Error		4.6			na	
	Test statistic		*t* = 0.21			na	
	***p*-value**		.83			na	
Average of school percent of students in grade 8	Average in Each Condition	17.2		17.8	na		na
	Standard Deviation	22.6		20.1	na		na
	Sample Size (Schools)	41		41	na		na
	Average Difference		–0.6			na	
	Standard Error		4.7			na	
	Test statistic		*t* = 0.13			na	
	***p*-value**		.90			na	
School average pretest score and sample size							
SAT10[a]	Pretest Score	628.6		618.9	632.7		635.1
	Standard Deviation	22.1		23.9	19.5		17.7
	Sample Size (Schools)	40		37	39		38
	Average Difference		9.8			2.4	

		SAT 10 mathematics problem solving outcome			SAT 10 science outcome		
Baseline characteristic		**AMSTI schools (Year 1)**		**Control schools (Year 2)**	**AMSTI schools (Year 1)**		**Control schools (Year 2)**
	Standard Error		5.2			4.3	
	Test statistic		*t* = 1.87			*t* = 0.56	
	***p*-value**		.07			.58	
Sample Size	Number of schools = 41 Number of teachers = 243 Number of students = 9,520		Number of schools = 41 Number of teachers = 164 Number of students = 8,144		Number of schools = 39 Number of teachers = 101 Number of students = 3,914		Number of schools = 40 Number of teachers = 60 Number of students = 3,161

na is not applicable.

Note: Detail may not sum to totals because of rounding. The number of schools, teachers and students for the comparisons varied slightly because of missing data.

For binary and continuously distributed variables, school means were computed and the hypothesis of no difference between the AMSTI and control averages of the variables tested. Grade level for mathematics problem solving is a categorical variable with more than two levels. In addition to testing for a difference between conditions in school proportions of cases for each response category, researchers tested the hypothesis of no difference between conditions in the distribution of students across grades for the mathematics problem solving outcome ($p = .94$). They also examined baseline equivalence overall. To do this they ran two logistic regressions. The first modeled the log odds of belonging to the treatment group. The second modeled the log odds of belonging to the treatment group conditioning on all the covariates that had been individually tested for equivalence. Based on the difference between the models in the deviance statistic, the hypothesis of no difference in model fit between the model with the covariates and the model without covariates was not rejected (the *p*-value was < .01 for mathematics problem solving and .96 for the science).

a. The SAT 10 mathematics problem solving pretest was used for the mathematics outcome. The SAT 10 reading pretest was used for the science outcome.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AH. Post hoc adjustment to standard error for estimate of two-year effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on mathematics achievement after two years

For the analysis of the two-year impact of AMSTI on mathematics achievement, the standard error produced by conventional software does not appropriately account for the fact that the analysis uses multiple observations for a given student (one each from two consecutive school years). To account for this nesting of test scores within students, researchers multiplied the initial standard error by 1.02, an adjustment factor derived from the following formula:

Let $$\lambda = \frac{\sqrt{\frac{4\sigma_J^2(1-R_J^2)}{n_j} + \frac{4\sigma_i^2(1-R_i^2)}{n_i} + \frac{4\sigma_t^2(1-R_t^2)}{n_t}}}{\sqrt{\frac{4\sigma^2(1-R^2)}{n_t}}}$$

where

- σ_J^2 = variance at school level.
- R_J^2 = variance at school level explained by covariates.
- σ_i^2 = variance at individual level.
- R_i^2 = variance at individual level explained by covariates.
- σ_t^2 = variance at score (time) level.
- R_t^2 = variance at score (time) level explained by covariates.

Assume that σ^2 = 1 and $R_i^2 = R_t^2$. Then

$$\lambda = \sqrt{\frac{4\sigma^2(1-R^2)}{n_i}}$$

$$\lambda^2 = \frac{\frac{\sigma_j}{n_j} + \frac{\sigma_i}{n_i} + \frac{\sigma_t}{n_t}}{\frac{\sigma}{n_i}}$$

$$\lambda = \sqrt{\frac{\frac{11.92}{82} + \frac{659.25}{25485} + \frac{298.66}{35524}}{\frac{11.92}{82} + \frac{(659.25 + 298.66)}{35524}}} = 1.02.$$

Appendix AI. Parameter estimates for effect of the Alabama Math, Science, and Technology Initiative (AMSTI) after two years

This appendix presents various estimates of the effect of AMSTI on mathematics problem solving (tables AI1–AI3) and science (tables AI4–AI6) achievement after two years.

Table AI1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after two years

Fixed effects model	Coefficient	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for the reference pair	373.4	44.5	39	8.38	< .01
Average effect associated with school-level average pretest on student outcome across all schools	0.5	0.1	39	6.75	< .01
Average effect associated with student-level pretest deviation from school average pretest on student outcome across all schools	0.6	0.0	35E3	29.07	< .01
Adjusted average AMSTI effect across all schools	–2.3	1.0	39	–2.34	.03
Average effect associated with being in the second cohort	0.8	0.6	79	1.34	.18
Average effect for interaction between birth cohort and treatment status	0.8	1.0	79	0.81	.42
Effect associated with dummy variable indicating missing value for student pretest	–4.4	1.3	35E3	–3.44	< .01
Average effect associated with male gender on student outcome across all schools	0.4	0.5	35E3	0.75	.45
Average effect associated with eligibility for free or reduced-price lunch status on student outcomes across all schools	–9.6	0.6	35E3	–15.40	< .01
Average effect associated with being a minority on student outcomes across all schools	–7.1	0.6	35E3	–12.64	< .01
Average effect associated with English proficiency on student outcomes across all schools	4.5	3.2	35E3	1.38	.17
Average effect associated with being in grade 4 (relative to grade 6) on student outcomes across all schools	–24.7	2.2	152	–11.47	< .01
Average effect associated with being in grade 5 (relative to grade 6) on student outcomes across all schools	–13.5	1.5	152	–9.17	< .01
Average effect associated with being in grade 7 (relative to grade 6) on student outcomes across all schools	1.8	1.2	152	1.47	.14

Fixed effects model	Coefficient	Standard error	Degrees of freedom	*t*-value	*p*-value
Average effect associated with being in grade 8 (relative to grade 6) on student outcomes across all schools	10.2	1.5	152	6.85	< .01
Effect associated with dummy variable indicating missing value for indicator of eligibility for free or reduced-lunch price	–27.1	16.1	35E3	–1.69	.09
Effect associated with dummy variable indicating missing value for indicator of racial/ethnic minority status	–3.0	4.5	35E3	–0.67	.50
Effect associated with dummy variable indicating missing value for indicator of English proficiency	–0.7	5.6	35E3	–0.12	.90

Note: Table excludes effect estimates for matched pairs. Degrees of freedoms may be expressed as approximations (35E3 is equivalent to roughly 35,000). Estimation of two-year effects for mathematics problem solving involved modeling repeated outcomes for a subsample of students. Specifically, students in grades 4–7 could contribute up to two scores (from consecutive grades). A dataset was created with one row per observation per student. Because most students contributed two observations, the dataset ended up containing more than 35,000 rows. The degrees of freedom for specific main effects in the model reflect this number of rows.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AI2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after two years

Random effects model	Coefficient	Standard error	*Z*-value	*p*-value
Variance component for students	862.4	6.5	133.09	< .0001
Variance component for students within schools	27.5	7.0	3.95	< .0001

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system

.

Table AI3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student mathematics problem solving achievement after two years

	Coefficient	Standard error	*t*-value	*p*-value
1	0.4	3.5	0.10	.92
2	0.6	3.0	0.21	.84
3	10.0	3.1	3.27	< .01
4	24.2	2.2	11.20	< .01
5	–4.0	3.4	–1.31	.20
6	18.2	2.3	7.92	< .01
7	–1.6	2.3	–0.68	.50
8	–7.1	1.5	–4.81	< .01
9	–3.7	2.6	–1.42	.16
10	5.2	3.8	1.37	.18
11	0.1	1.6	0.07	.95
12	–0.6	2.5	–0.22	.82
13	14.7	4.7	3.16	< .01
14	6.5	3.0	2.19	.03
15	5.6	1.9	2.88	< .01
16	0.4	2.1	0.18	.86
17	4.0	1.9	2.08	.04
18	–0.5	3.4	–0.14	.89
19	7.8	5.1	1.52	.14
20	4.3	1.7	2.56	.01
21	1.2	2.3	0.54	.59
22	11.0	2.6	4.27	< .01
23	11.8	4.0	2.93	< .01
24	9.7	2.9	3.38	< .01
25	22.9	7.8	2.94	< .01
26	0.1	3.5	0.04	.97
27	7.3	4.1	1.80	.08
28	1.5	3.8	0.38	.70
29	14.4	5.1	2.80	< .01
30	2.9	3.8	0.77	.45
31	4.8	4.2	1.14	.26
32	8.3	6.5	1.29	.20
33	20.8	4.0	5.15	< .01
34	2.1	4.1	0.53	.60
35	6.2	4.2	1.49	.15
36	–5.7	4.1	–1.41	.17
37	13.8	2.0	6.84	< .01
38	1.3	2.6	0.52	.61
39	–0.5	3.7	–0.14	.89
40	8.9	1.8	5.08	< .01

Note: The degrees of freedom associated with each estimate is 39.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AI4 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after two years

Fixed effects model	Coefficient	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	427.9	52.4	37	8.17	< .01
Average effect associated with school-level average pretest on student outcome across all schools	0.4	0.1	37	4.28	< .01
Average effect associated with student-level pretest deviation from school average pretest on student outcome across all schools	0.5	0.0	14E3	43.94	< .01
Adjusted average AMSTI effect across all schools	–2.5	1.1	37	–2.33	.03
Average effect associated with being in second birth cohort	0.8	0.7	75	1.14	.26
Average effect associated with interaction between birth cohort and treatment status	1.1	1.2	75	0.88	.38
Effect associated with dummy variable indicating missing value for student pretest	–2.2	1.4	14E3	–1.52	.13
Average effect associated with male gender on student outcome across all schools	5.0	0.6	14E3	9.07	< .01
Effect associated with dummy variable indicating missing value for indicator that student has free or reduced-lunch price status	–12.5	10.1	14E3	–1.23	.22
Average effect associated with eligibility for free or reduced-price lunch on student outcome across all schools	–5.0	0.6	14E3	–8.79	< .01
Effect associated with dummy variable indicating missing value for indicator of racial/ethnic minority status	4.9	8.8	14E3	0.56	.57
Average effect associated with being a minority on student outcomes across all schools	–6.9	0.8	14E3	–9.16	< .01
Effect associated with dummy variable indicating missing value for indicator of English proficiency	–5.5	9.3	14E3	–0.59	.56
Average effect associate with English proficiency on student outcomes across all schools	3.9	1.5	14E3	2.65	< .01
Average effect of being in grade 5 (relative to grade 7) on student outcomes across all schools	–8.1	2.7	18	–2.99	< .01

Note: Table excludes effect estimates for matched pairs. Degrees of freedoms may be expressed as approximations (that is, 14E3 is equivalent to roughly 14,000). The dummy variable approach to handling missing data involves setting missing values for covariates to a constant. These effects are estimated with missing values set to zero; the effect estimate should be interpreted accordingly.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AI5 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after two years

Random effects model	Coefficient	Standard error	*Z*-value	*p*-value
Variance component for students	523.4	6.2	84.23	< .01
Variance component for students within schools	31.1	8.6	3.60	< .01

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AI6 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student science achievement after two years

	Estimate	Standard error	*t*- value	*p*-value
1	1.6	3.7	0.44	.66
2	1.1	3.4	0.33	.74
3	18.6	3.9	4.84	< .01
4	18.3	3.8	4.86	< .01
5	–1.0	2.0	–0.48	.63
6	14.8	2.6	5.66	< .01
7	2.1	4.0	0.52	.61
8	–4.6	2.3	–2.04	.05
9	–2.0	1.5	–1.34	.19
10	3.3	1.8	1.84	.07
11	3.8	1.9	2.03	.05
12	2.9	3.5	0.84	.41
13	11.6	3.2	3.60	< .01
14	1.3	4.3	0.30	.76
15	7.1	2.3	3.10	< .01
16	7.5	5.3	1.41	.17
17	8.6	1.9	4.60	< .01
18	7.3	1.7	4.17	< .01
19	6.4	2.2	2.91	< .01
20	6.5	3.1	2.07	.05
21	3.7	3.9	0.95	.35
22	9.9	4.1	2.42	.02
23	9.1	3.6	2.55	.02
24	11.4	5.7	2.00	.05
25	8.2	3.4	2.38	.02
26	6.5	3.1	2.10	.04
27	7.9	2.2	3.66	< .01
28	6.4	4.6	1.37	.18
29	7.8	4.5	1.74	.09
30	11.6	1.9	6.17	< .01
31	–10.2	5.7	–1.78	.08
32	4.2	7.6	0.56	.58
33	15.3	7.2	2.12	.04
34	–4.3	9.3	–0.46	.65
35	5.0	3.4	1.48	.15
36	–8.2	5.9	–1.40	.17

	Estimate	Standard error	*t*- value	*p*-value
37	15.3	3.5	4.43	< .01
38	–1.5	2.1	–0.75	.46
39	4.7	1.6	3.01	< .01
40	8.2	4.1	2.00	.05

Note: The degrees of freedom associated with each estimate is 37.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AJ. Parameter estimates for effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on student reading achievement after one year

This appendix (tables AJ1–AJ3) presents various estimates of the effect of AMSTI on student reading achievement after one year.

Table AJ1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student reading achievement after one year

Fixed effects model	Variance estimates	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	228.6	31.7	39	7.20	< .01
Adjusted average AMSTI effect across all schools	2.3	0.5	39	4.97	< .01
Average effect associated with school-level average pretest on student outcome across schools	0.7	0.1	39	13.80	< .01
Average effect associated with student-level pretest deviation from school average pretest on student outcome across schools[a]	0.7	0.0	18,615	74.11	< .01
Average effect associated with male gender on student outcome across all schools	–2.3	0.3	18,615	–6.86	< .01
Average effect associated with eligibility for free or reduced-price lunch on student outcome across all schools[a]	–5.4	0.4	18,615	–12.95	< .01
Average effect associated with being a minority on student outcomes across all schools[a]	–4.1	0.5	18,615	–8.65	< .01
Average effect associated with English proficiency on student outcomes across all schools[a]	6.9	2.2	18,615	3.19	< .01
Effect associated with dummy variable indicating missing value for student pretest	–5.9	1.9	18,615	–3.15	< .01
Effect associated with dummy variable indicating missing value for indicator of English proficiency	5.2	4.3	18,615	1.21	.23
Effect associated with dummy variable indicating missing value for indicator of eligibility for free or reduced-price lunch	–16.5	6.8	18,615	–2.44	.02
Effect associated with dummy variable indicating missing value for indicator of racial/ethnic minority status	–2.2	3.4	18,615	–0.64	.52
Average effect of being in grade 4 (relative to grade 6) on student outcomes across all schools	–3.9	1.5	18,615	–2.72	< .01
Average effect of being in grade 5 (relative to grade 6) on student outcomes across all schools	–5.3	1.2	18,615	–4.53	< .01

Fixed effects model	Variance estimates	Standard error	Degrees of freedom	*t*-value	*p*-value
Average effect of being in grade 7 (relative to grade 6) on student outcomes across all schools	4.5	1.6	18,615	2.75	< .01
Average effect of being in grade 8 (relative to grade 6) on student outcomes across all schools	0.4	1.3	18,615	0.35	.73

Note: This table excludes effect estimates for matched pairs.

a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant. ; These effects are estimated with missing values set to zero; therefore, the effect estimate should the effect estimate should be interpreted accordingly.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AJ2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student reading achievement after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for students within schools	449.0	4.7	96.45	< .01
Variance component for schools	9.4	3.0	3.11	< .01

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AJ3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student reading achievement after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t*–value	*p*-value
1	–2.5	4.1	–0.60	.55
2	–3.7	4.1	–0.90	.37
3	2.0	4.0	0.51	.62
4	3.4	4.0	0.84	.41
5	–7.2	3.8	–1.90	.07
6	–0.8	4.1	–0.19	.85
7	–12.2	3.8	–3.16	< .01
8	–4.9	4.3	–1.13	.27
9	–1.9	3.7	–0.50	.62
10	–3.6	4.0	–0.90	.37
11	–7.4	4.0	–1.87	.07
12	–4.6	4.0	–1.17	.25
13	–2.1	4.4	–0.49	.63
14	–6.2	4.1	–1.53	.13
15	–6.3	3.9	–1.64	.11
16	–6.7	3.8	–1.73	.09
17	–3.8	3.7	–1.02	.31
18	–7.7	3.7	–2.07	.05
19	–4.9	5.1	–0.96	.34
20	–2.9	3.8	–0.77	.45
21	–5.9	3.9	–1.51	.14
22	–2.9	4.0	–0.72	.48
23	–5.3	3.9	–1.37	.18
24	–5.8	3.9	–1.49	.14
25	–2.5	4.0	–0.63	.53
26	–2.3	3.9	–0.58	.56
27	–3.6	3.8	–0.96	.34
28	–6.0	5.2	–1.14	.26
29	–0.1	4.4	–0.01	.99
30	–3.5	3.9	–0.89	.38
31	–6.4	5.1	–1.25	.22
32	–6.5	6.1	–1.06	.30
33	1.9	4.0	0.47	.64
34	–4.5	4.1	–1.10	.28
35	–7.5	3.9	–1.91	.06
36	–19.9	8.4	–2.37	.02
37	–1.5	4.0	–0.37	.72
38	–8.4	4.0	–2.07	.05
39	–6.7	3.8	–1.76	.09
40	–7.6	3.8	–2.03	.05

Note: The degrees of freedom associated with each estimate is 39.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AK. Parameter estimates for teacher content and student engagement after one year

This appendix (tables AK1–AK9) presents various estimates of the effect of AMSTI on teacher content knowledge and student engagement in mathematics and science after one year. Table AK10 describes the analytic sample used to assess variation in the effects of AMSTI on achievement for subgroups of students after one year.

Table AK1 Estimates of fixed effects from the benchmark multilevel analysis of the t of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher content knowledge in mathematics after one year

Fixed effects model	Variance	Standard error	*t*–value	*p*–value
Adjusted grand school mean in control condition for reference pair	4.4	0.1	46.05	< .01
Adjusted average AMSTI effect across all schools	0.0	0.1	0.44	.66
Average effect associated with degree rank 0 on teacher outcome across all schools[a]	0.0	0.1	–0.18	.86
Average effect associated with degree rank 1 on teacher outcome across all schools[a]	–0.1	0.1	–0.77	.44
Average effect associated with years teaching experience on teacher outcome across all schools[a]	0.0	0.0	–1.67	.10
Average effect associated with years teaching experience with math on teacher outcome across all schools[a]	0.0	0.0	3.05	< .01
Effect associated with dummy variable indicating missing value for degree rank	0.1	0.2	0.49	.63
Effect associated with dummy variable indicating missing value for years teaching experience (also indicates missing value for years of teaching experience in subject area)	–0.4	0.3	–1.71	.09

Note: This table excludes effect estimates for matched pairs. The degrees of freedom associated with each estimate is 297.
a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant; these effects are estimated with missing values set to zero; therefore, the effect estimate should be interpreted accordingly.
Source: Teacher survey data.

Table AK2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher content knowledge in mathematics after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for teachers within schools	0.4	0.0	12.69	< .01
Variance component for schools	0.0	0.0	0.69	.25

Source: Teacher survey data.

Multilevel random effects model estimates for the assessment of the one-year effect of AMSTI on teacher content knowledge in science are not displayed. As explained in chapter 6, this effect was excluded because estimation led to a boundary constraint for the school-level random effect, which was assigned a value of 0, with no *p*-value. The estimate of impact was nonsignificant. Controlling for clustering through modeling a school-level random effect would reduce the precision of the effect estimate, leading to the prediction that the result would remain nonsignificant if the effects of clustering had been modeled.

Table AK3 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher content knowledge in science after one year

Fixed effects model	Variance	Standard error	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	4.8	0.2	27.11	< .01
Adjusted average AMSTI effect across all schools	0.1	0.1	1.58	.12
Average effect associated with degree rank 0 on teacher outcome across all schools[a]	0.0	0.1	–0.06	.95
Average effect associated with degree rank 1 on teacher outcome across all schools[a]	0.0	0.1	–0.15	.88
Average effect associated with years teaching experience on teacher outcome across all schools[a]	0.0	0.0	–2.48	.01
Average effect associated with years teaching experience with science on teacher outcome across all schools[a]	0.0	0.0	4.23	< .01
Effect associated with dummy variable indicating missing value for degree rank	–1.0	0.2	–4.31	< .01
Effect associated with dummy variable indicating missing value for years teaching experience (also indicates missing value for years of teaching experience in subject area)	1.4	0.3	4.55	< .01

Note: Table excludes effect estimates for matched pairs. The degrees of freedom associated with each estimate is 298.
a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant. These effects are estimated with missing values set to zero; the effect estimate should be interpreted accordingly.
Source: Teacher survey data.

Table AK4 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on teacher content knowledge in science after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41002)	Standard error	*t*-value	*p*-value
1001	–1.1	0.3	–3.73	< .01
2001	–1.1	0.3	–3.90	< .01
3001	–1.2	0.2	–4.98	< .01
4001	–0.9	0.3	–3.59	< .01
5001	–1.1	0.3	–3.33	< .01
6001	–0.7	0.2	–3.91	< .01
7001	–1.4	0.4	–3.87	< .01
8001	–1.0	0.2	–6.36	< .01
9001	0.1	0.2	0.65	.52
10001	–0.4	0.2	–2.21	.03
11001	–0.7	0.3	–2.21	.03
12001	–0.9	0.3	–3.58	< .01
13001	–1.4	0.2	–6.74	< .01
14001	–1.1	0.3	–4.12	< .01
15001	–1.2	0.2	–4.82	< .01
16001	–0.9	0.3	–3.53	< .01
17001	–1.4	0.2	–7.00	< .01
18001	–1.0	0.3	–3.53	< .01
19001	0.3	0.2	1.66	.10
20001	–1.2	0.4	–3.42	< .01
21002	–1.0	0.3	–2.94	< .01
22002	–0.9	0.3	–3.06	< .01
23002	–1.1	0.2	–5.09	< .01
24002	–1.3	0.3	–3.77	< .01
25002	–0.8	0.2	–3.34	< .01
26002	–0.7	0.3	–2.23	.03
27002	–0.9	0.3	–3.13	< .01
28002	–0.2	0.2	–1.30	.20
29002	–1.4	0.3	–5.83	< .01
30002	–1.0	0.3	–3.06	< .01
31002	–1.5	0.2	–6.71	< .01
32002	–1.2	0.3	–4.12	< .01
33002	–1.1	0.3	–4.03	< .01
34002	–1.3	0.5	–2.95	< .01
35002	–1.2	0.4	–2.91	.04
36002	–0.3	0.3	–0.87	.38
37002	–1.2	0.3	–3.97	< .01
38002	–1.0	0.4	–2.25	.03
39002	–1.4	0.2	–6.44	< .01
40002	–1.1	0.3	–3.58	< .01

Note: The degrees of freedom associated with each estimate is 298.
Source: Teacher survey data.

Table AK5 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student engagement in mathematics after one year

Fixed effects model	Variance estimates	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	2.5	0.1	39	21.09	< .01
Adjusted average AMSTI effect across all schools	0.2	0.1	39	2.60	.01
Average effect associated with degree rank 0 on teacher outcome across all schools[a]	–0.2	0.1	298	–1.52	.13
Average effect associated with degree rank 1 on teacher outcome across all schools[a]	–0.0	0.1	298	–0.11	.92
Average effect associated with years teaching experience on teacher outcome across all schools[a]	0.0	0.0	298	1.90	.06
Average effect associated with years teaching experience with math on teacher outcome across all schools[a]	0.0	0.0	298	–1.27	.21
Effect associated with dummy variable indicating missing value for degree rank	1.0	0.3	298	3.75	< .01
Effect associated with dummy variable indicating missing value for years teaching experience (also indicates missing value for years of teaching experience in subject area)	–0.2	0.3	298	–0.63	.53

Note: This table excludes effect estimates for matched pairs.

a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant. These effects are estimated with missing values set to zero; therefore, the effect estimate should be interpreted accordingly.

Source: Teacher survey data.

Table AK6 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student engagement in mathematics after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for teachers within schools	0.6	0.1	12.47	< .01
Variance component for schools	0.0	0.0	1.47	.07

Source: Teacher survey data.

Table AK7 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student engagement in science after one year

Fixed effects model	Variance	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	3.4	0.2	38	17.28	< .01
Adjusted average AMSTI effect across all schools	0.4	0.1	38	4.57	< .01
Average effect associated with degree rank 0 on teacher outcome across all schools[a]	–0.2	0.2	262	–1.39	.17
Average effect associated with degree rank 1 on teacher outcome across all schools[a]	0.0	0.1	262	–0.23	.82
Average effect associated with years teaching experience on teacher outcome across all schools[a]	0.0	0.0	262	–1.02	.31
Average effect associated with years teaching experience with science on teacher outcome across all schools[a]	0.0	0.0	262	0.86	.39
Effect associated with dummy variable indicating missing value for degree rank	–0.8	0.1	262	–5.66	< .01
Effect associated with dummy variable indicating missing value for years teaching experience (also indicates missing value for years of teaching experience in subject area)	1.6	0.2	262	7.14	< .01

Note: Table excludes effect estimates for matched pairs.
a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant. These effects are estimated with missing values set to zero; therefore, the effect estimate should be interpreted accordingly.
Source: Teacher survey data.

Table AK8 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student engagement in science after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for teachers within schools	0.5	0.1	11.48	< .01
Variance component for schools	0.1	0.1	2.02	.02

Source: Teacher survey data.

Table AK9 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on student engagement in science after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41002)	Standard error	*t*-value	*p*-value
1001	–0.6	0.4	–1.62	.11
2001	–0.5	0.2	–2.71	.01
3001	–0.4	0.2	–2.89	< .01
4001	–0.3	0.2	–1.7	.10
5001	–1.1	0.3	–3.39	< .01
6001	–0.5	0.1	–3.19	< .01
7001	–1.5	0.2	–6.66	< .01
8001	–0.6	0.2	–2.69	.01
9001	–0.1	0.3	–0.46	.65
10001	–0.8	0.3	–3.21	< .01
11001	–0.6	0.5	–1.38	.18
12001	–0.4	0.2	–2.48	.02
13001	–0.5	0.2	–2.42	.02
14001	–0.9	0.6	–1.47	.15
15001	–0.9	0.2	–4.82	< .01
16001	–1.0	0.2	–6.96	< .01
17001	–1.0	0.4	–2.24	.03
18001	–1.2	0.2	–6.92	< .01
19001	–0.3	0.2	–1.91	.06
20001	–1.1	0.2	–5.72	< .01
21002	–0.8	0.2	–4.71	< .01
22002	–0.4	0.2	–1.75	.09
23002	–0.6	0.2	–4.11	< .01
24002	–0.8	0.2	–3.74	.01
25002	–0.4	0.2	–1.97	.06
26002	–1.1	0.2	–7.06	< .01
27002	–1.0	0.4	–2.5	.02
28002	–1.1	0.3	–3.95	< .01
29002	–0.8	0.2	–4.69	< .01
30002	–1.3	0.4	–3.6	< .01
31002	–0.9	0.3	–3.07	< .01
32002	–1.2	0.2	–7.54	< .01
33002	–0.7	0.3	–2.09	.04
34002	–0.8	0.7	–1.13	.27
35002	–0.9	0.6	–1.47	.15
36002	0.3	0.1	1.76	.09
37002	–0.8	0.2	–4.64	< .01
38002	–1.7	0.2	–8.99	< .01
39002	–1.1	0.6	–2.02	.05
40002	–0.7	0.3	–2.33	.03

Note: The degrees of freedom associated with each estimate is 38.
Source: Teacher survey data

.

Table AK10 Analytic sample used to assess variation in effects of the Alabama Math, Science, and Technology Initiative (AMSTI) on achievement for subgroups of students after one year

	Mathematics problem solving		Science		Reading	
Covariate	**AMSTI**	**Control**	**AMSTI**	**Control**	**AMSTI**	**Control**
Racial/ethnic minority status						
Minority students	4,298	3,375	1,872	1,350	4,301	3,375
White students	5,437	4,930	2,087	1,939	5,437	4,933
Total	41	41	39	40	41	41
Socioeconomic status						
Students enrolled in the free or reduced-price lunch program	5,533	4,858	2,319	1,965	5,533	4,854
Students not enrolled in the free or reduced-price lunch program	4,487	3,831	1,761	1,481	4,484	3,834
Total	41	41	39	40	41	41
Gender						
Girls	5,058	4,435	2,079	1,729	5,053	4,429
Boys	4,964	4,256	2,003	1,717	4,966	4,262
Total	41	41	39	40	41	41
SAT 10 reading pretest						
Low (stanines 1–3)	2,128	1,699	920	655	2,130	1,702
Middle (stanines 4–6)	4,786	4,345	1,938	1,734	4,790	4,343
High (stanines 7–9)	2,389	1,913	951	770	2,389	1,914
Total	41	41	39	40	41	41
SAT 10 mathematics problem solving pretest						
Low (stanines 1–3)	2,113	1,629	na	na	na	na
Middle (stanines 4–6)	4,836	4,363	na	na	na	na
High (stanines 7–9)	2,371	1,965	na	na	na	na
Total	41	41	39	40	41	41

na is not applicable.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AL. Estimates of effects for terms involving the indicator of treatment status in the analysis of the moderating effect of the three-level pretest variable

The analysis of the moderating effect of the pretest resulted in three estimates pertaining to the moderating effect: the impact for the high category (the reference category), the additional impact associated with being in the middle category (relative to the high category), and the additional impact associated with being in the low category (relative to the high category). These estimates are displayed below (table AL1).

Table AL1 Estimates of effects for terms involving the indicator of treatment status in the analysis of the moderating effect of three-level pretest variable

	Mathematics problem solving			Science			Reading		
Covariate	Estimated effect	*p*-value	Effect size[a]	Estimated effect	*p*-value	Effect size[a]	Estimated effect	*p*-value	Effect size[a]
SAT 10 reading pretest[b]									
Impact on high pretest group (stanines 7–9)	1.3 (2.5)	.60	0.03	0.1 (1.9)	.94	0.00	2.6 (1.3)	.05	0.07
Additional impact on middle pretest group (in relation to high) (stanines 4–6)	0.3 (2.6)	.91	0.01	2.1 (1.9)	.27	0.07	–1.4 (1.3)	.30	–0.04
Additional impact on low pretest group (in relation to high) (stanines 1–3)	–0.5 (3.2)	.89	–0.01	–0.9 (2.8)	.76	–0.03	–1.4 (2.0)	.48	–0.04
SAT 10 mathematics problem solving pretest[b]									
Impact on high pretest group	1.1 (2.5)	.66	0.03	na	na	na	na	na	na
Additional impact on middle pretest group (in relation to high)	0.8 (2.9)	.78	0.02	na	na	na	na	na	na
Additional impact on low pretest group (in relation to high)	–1.3 (3.2)	.69	–0.03	na	na	na	na	na	na

na is not applicable.

Note: Numbers in parentheses are standard errors.

a. For each of the three main outcomes (mathematics problem solving, science, and reading), the estimated standard deviation for the control group from the analytic samples was used to estimate the average impacts of AMSTI (from the confirmatory analyses for impacts on mathematics and science and the corresponding exploratory analysis in reading) as the denominator in the standardized effect size estimate. This step was taken in order to express all estimates for a given scale in terms of the same standard deviation units, to facilitate comparison of results.

b. The pretest was divided into three categories: low, for scores in stanines 1–3; middle, for scores in stanines 4–6; and high, for scores in stanines 7–9. The cutpoints for the stanines were based on the pretest scale scores for the sample. As explained in appendix B, the study's technical working group advisors recommended that the study examine whether the effect of AMSTI on student achievement in mathematics varied depending on students' pretest scores on the SAT 10 reading exam. In the absence of a science pretest, the study team examined whether the effect of AMSTI on student achievement in science varied depending on pretest scores on the SAT 10 reading exam.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AM. Parameter estimates for the analysis of the moderating effect of racial/ethnic minority status on the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading after one year

This appendix (tables AM1 and AM2) presents estimates of the moderating effect of minority status on the impact of AMSTI on reading after one year. Presented below (table AM3) are the fixed effects estimates for matched pairs from the benchmark model used to estimate the moderating effect of racial/ethnic minority status on the effect of AMSTI on reading after one year.

Table AM1 Estimates of fixed effects from the benchmark multilevel analysis of the moderating effect of minority status on the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading after one year

Fixed effects model	Variance	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	242.5	33.5	39	7.23	< .01
Adjusted average AMSTI effect across all schools	3.6	0.6	39	6.15	< .01
Average effect associated with racial/ethnic status on student outcomes across all schools[a]	–2.6	0.6	17,951	–4.16	< .01
Additional impact associated with being a minority	–3.0	0.9	17,951	–3.59	< .01
Average effect associated with school-level average pretest on student outcome across schools	0.7	0.1	39	12.54	< .01
Average effect associated with student-level pretest deviation from school average pretest on student outcome across schools[a]	0.7	0.0	17,951	73.51	< .01
Average effect associated with male gender on student outcome across all schools	–2.1	0.3	17,951	–6.13	< .01
Average effect associated with eligibility for free or reduced-price lunch on student outcome across all schools[a]	–4.6	0.4	17,951	–10.77	< .01
Average effect associated with English proficiency on student outcomes across all schools[a]	7.1	2.1	17,951	3.41	< .01
Average effect of being in grade 4 (relative to grade 6) on student outcomes across all schools	–3.4	1.5	17,951	–2.30	.02
Average effect of being in grade 5 (relative to grade 6) on student outcomes across all schools	–5.2	1.2	17,951	–4.45	< .01
Average effect of being in grade 7 (relative to grade 6) on student outcomes across all schools	4.1	1.6	17,951	2.53	.01
Average effect of being in grade 8 (relative to grade 6) on student outcomes across all schools	–0.3	1.3	17,951	–0.20	.84
Effect associated with dummy variable indicating missing value for student pretest	–6.1	2.0	17,951	–3.00	< .01
Effect associated with dummy variable indicating missing value for indicator of proficiency in English	31.1	2.9	17,951	10.74	< .01
Effect associated with dummy variable indicating missing value for indicator of eligibility for free or reduced-price lunch	–15.9	6.9	17,951	–2.29	.02

Note: Table excludes effect estimates for matched pairs.

a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant; these effects are estimated with missing values set to zero; therefore, the effect estimate should be interpreted accordingly.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system

Table AM2 Estimates of random effects from the benchmark multilevel analysis of the moderating effect of minority status on the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading after one year

Random effects model	Variance	Standard error	*Z*-value	*p*-value
Variance component for students within schools	422.9	4.5	94.72	< .01
Variance component for schools	10.0	3.1	3.23	< .01

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AM3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the moderating effect of minority status on impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t*-value	*p*-value
1	–3.9	3.7	–1.08	.29
2	–5.3	3.7	–1.46	.15
3	2.6	3.5	0.75	.46
4	7.0	3.5	1.96	.06
5	–7.6	3.3	–2.30	.03
6	0.4	3.7	0.10	.92
7	–12.6	3.4	–3.68	< .01
8	–4.3	4.0	–1.07	.29
9	–1.9	3.2	–0.60	.55
10	–2.9	3.8	–0.76	.45
11	–6.8	3.6	–1.90	.06
12	–5.3	3.6	–1.49	.14
13	–1.8	4.0	–0.46	.65
14	–7.4	3.8	–1.96	.06
15	–6.4	3.3	–1.93	.06
16	–7.2	3.4	–2.14	.04
17	–4.6	3.4	–1.37	.18
18	–8.2	3.3	–2.46	.02
19	–5.0	4.5	–1.11	.28
20	–3.3	3.6	–0.92	.36
21	–6.3	3.5	–1.81	.08
22	–3.7	3.6	–1.02	.32
23	–5.9	3.4	–1.74	.09
24	–6.4	3.4	–1.87	.07
25	–3.0	3.5	–0.86	.39
26	–2.1	3.4	–0.61	.54
27	–3.3	3.3	–1.01	.32
28	–5.0	4.7	–1.07	.29
29	–2.2	4.0	–0.55	.58
30	–2.5	3.5	–0.71	.48
31	–7.5	5.1	–1.49	.14
32	–8.0	5.5	–1.44	.16
33	0.5	3.5	0.14	.89
34	–5.7	3.9	–1.48	.15
35	–8.3	3.5	–2.39	.02

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t*-value	*p*-value
36	–21.4	8.5	–2.53	.02
37	–1.5	3.6	–0.43	.67
38	–9.2	3.5	–2.61	.01
39	–6.9	3.2	–2.13	.04
40	–7.1	3.4	–2.10	.04

Note: The degrees of freedom associated with each estimate is 39.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AN. Parameter estimates for analysis of average effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by racial/ethnic minority students after one year

This appendix (tables AN1–AN3) presents various estimates of the effect of AMSTI on the reading achievement of racial/ethnic minority students after one year.

Table AN1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by racial/ethnic minority students after one year

Fixed effects model	Variance	Standard error	Degrees of freedom	*t*-value	*p*-value
Adjusted grand school mean in control condition for reference pair	269.4	40.0	39	6.74	< .01
Adjusted average AMSTI effect across all schools	0.7	0.6	39	1.06	.29
Average effect associated with school-level average pretest on student outcome across schools	0.6	0.1	39	9.85	< .01
Average effect associated with student-level pretest deviation from school average pretest on student outcome across schools[a]	0.7	0.0	7,584	35.77	< .01
Average effect associated with male gender on student outcome across all schools	–2.6	0.4	7,584	–5.93	< .01
Average effect associated with eligibility for free or reduced-price lunch on student outcome across all schools[a]	–5.2	0.8	7,584	–6.37	< .01
Average effect associated with English proficiency on student outcomes across all schools[a]	7.3	2.2	7,584	3.36	.01
Average effect of being in grade 4 (relative to grade 6) on student outcomes across all schools	–8.2	2.1	7,584	–4.00	< .01
Average effect of being in grade 5 (relative to grade 6) on student outcomes across all schools	–6.4	1.8	7,584	–3.49	< .01
Average effect of being in grade 7 (relative to grade 6) on student outcomes across all schools	1.9	1.7	7,584	1.12	.26
Average effect of being in grade 8 (relative to grade 6) on student outcomes across all schools	0.4	1.6	7,584	0.24	.81
Effect associated with dummy variable indicating missing value for student pretest	–10.4	2.7	7,584	–3.87	< .01
Effect associated with dummy variable indicating missing value for indicator of eligibility for free or reduced-price lunch	–16.1	8.2	7,584	–1.97	.05

Note: This table excludes effect estimates for matched pairs.
a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant; these effects are estimated with missing values set to zero; therefore, the effect estimate should be interpreted accordingly.
The estimate for the effect associated with the dummy variable of for proficiency English speaker is missing estimate.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AN2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading achievement by racial/ethnic minority students after one year

Random effects model	Variance estimates	Standard error	*Z*-value	*p*-value
Variance component for students within schools	424.8	6.9	61.59	< .01
Variance component for schools	15.2	5.0	3.01	< .01

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AN3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading achievement by racial/ethnic minority status after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t* -value	*p*-value
1	–8.1	4.7	–1.73	.09
2	–5.5	3.9	–1.40	.17
3	3.7	4.4	0.84	.40
4	10.1	3.9	2.56	.02
5	–5.5	3.6	–1.53	.13
6	–0.8	4.1	–0.18	.86
7	–12.7	3.7	–3.40	< .01
8	–4.2	5.0	–0.84	.41
9	–0.4	3.9	–0.11	.91
10	–3.1	3.6	–0.86	.40
11	–7.1	3.8	–1.86	.07
12	–9.0	4.4	–2.05	.05
13	–0.1	5.2	–0.02	.99
14	–8.4	4.0	–2.09	.04
15	–5.5	3.6	–1.51	.14
16	–6.0	3.6	–1.65	.11
17	–4.5	3.5	–1.27	.21
18	–9.2	3.7	–2.50	.02
19	–6.8	5.8	–1.17	.25
20	–0.3	3.9	–0.08	.94
21	–6.5	3.7	–1.78	.08
22	–2.3	4.0	–0.58	.56
23	–4.3	5.2	–0.83	.41
24	–6.9	3.8	–1.80	.08
25	–2.2	4.3	–0.51	.61
26	14.2	4.4	3.22	< .01
27	–4.2	3.8	–1.11	.27
28	–1.4	3.8	–0.38	.71
29	–0.2	5.1	–0.05	.96
30	–2.7	3.8	–0.71	.48
31	–7.3	5.0	–1.47	.15
32	–7.2	5.8	–1.24	.22
33	1.0	3.9	0.26	.80
34	–5.8	4.2	–1.37	.18

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t* -value	*p*-value
35	–7.7	4.0	–1.94	.06
36	–20.7	8.2	–2.53	.02
37	–1.4	4.0	–0.34	.74
38	–9.8	3.8	–2.58	.01
39	–7.4	3.5	–2.12	.04
40	–6.8	3.6	–1.92	.06

Note: The degrees of freedom associated with each estimate is 39.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Appendix AO. Parameter estimates for effect of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading for White students after one year

This appendix (tables AO1–AO3) presents various estimates of the effect of AMSTI on the reading achievement of White students after one year.

Table AO1 Estimates of fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by White students after one year

Fixed effects model	Variance	Standard error	Degrees of freedom	*t* -value	*p*-value
Adjusted grand school mean in control condition for reference pair	183.8	23.9	32	7.69	< .01
Adjusted average AMSTI effect across all schools	3.1	0.5	32	6.42	< .01
Average effect associated with school-level average pretest on student outcome across schools	0.8	0.0	32	20.14	< .01
Average effect associated with student-level pretest deviation from school average pretest on student outcome across schools[a]	0.8	0.0	10,285	89.01	< .01
Average effect associated with male gender on student outcome across all schools	–1.7	0.4	10,285	–3.93	< .01
Average effect associated with eligibility for free or reduced-price lunch on student outcome across all schools[a]	–4.3	0.5	10,285	–8.02	< .01
Average effect associated with English proficiency on student outcomes across all schools[a]	–1.1	6.3	10,285	–0.17	.87
Average effect of being in grade 4 (relative to grade 6) on student outcomes across all schools	0.0	1.5	10,285	–0.01	.99
Average effect of being in grade 5 (relative to grade 6) on student outcomes across all schools	–4.5	1.2	10,285	–3.91	< .01
Average effect of being in grade 7 (relative to grade 6) on student outcomes across all schools	5.9	1.9	10,285	3.14	< .01
Average effect of being in grade 8 (relative to grade 6) on student outcomes across all schools	–0.7	1.4	10,285	–0.52	.60
Effect associated with dummy variable indicating missing value for student pretest	–1.2	2.7	10,285	–0.44	.66
Effect associated with dummy variable indicating missing value for indicator of proficiency in English	19.4	5.4	10,285	3.63	< .01
Effect associated with dummy variable indicating missing value for indicator of eligibility for free or reduced-price lunch	–8.5	1.5	10,285	–5.76	< .01

Note: Table excludes effect estimates for matched pairs.

a. The dummy variable approach to handling missing data involves setting missing values for covariates to a constant; these effects are estimated with missing values set to zero; therefore, the effect estimate should be interpreted accordingly.

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AO2 Estimates of random effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by White students after one year

Random effects model	Variance	Standard error	*Z* -value	*p*-value
Variance component for students within schools	415.3	5.8	71.73	< .01
Variance component for schools	3.0	1.7	1.78	.04

Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

Table AO3 Estimates of matched-pair fixed effects from the benchmark multilevel analysis of the impact of the Alabama Math, Science, and Technology Initiative (AMSTI) on reading by White students after one year

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t* -value	*p*-value
1	–11.9	2.5	–4.86	< .01
2	–14.2	2.1	–6.91	< .01
3	–10.0	1.4	–7.23	< .01
4	–7.2	1.6	–4.50	< .01
5	–19.5	1.1	–17.27	< .01
6	–12.7	1.4	–8.85	< .01
8	–15.4	1.9	–8.01	< .01
9	–15.0	3.7	–4.02	< .01
10	–14.7	2.3	–6.47	< .01
11	–16.8	0.7	–25.83	< .01
12	–15.4	1.7	–9.13	< .01
13	–13.5	1.4	–10.00	< .01
14	–14.7	1.7	–8.73	< .01
15	–18.1	0.9	–21.37	< .01
16	–18.0	1.4	–12.65	< .01
17	–16.8	1.6	–10.46	< .01
18	–17.7	0.9	–19.91	< .01
19	–16.2	1.9	–8.37	< .01
20	–15.4	1.4	–11.27	< .01
21	–15.6	1.9	–8.00	< .01
22	–13.6	1.5	–9.15	< .01
23	–16.9	1.4	–11.91	< .01
24	–16.7	1.3	–12.58	< .01
25	–14.0	1.4	–9.87	< .01
	–15.1	1.1	–13.91	< .01
27	–14.9	0.9	–16.87	< .01
28	–18.3	3.4	–5.36	< .01
29	–16.2	3.7	–4.36	< .01
30	–15.5	1.3	–12.38	< .01
31	–12.9	3.5	–3.73	< .01
32	–18.6	9.1	–2.05	.05
33	–21.7	2.2	–9.68	< .01
34	–53.0	1.9	–27.96	< .01
35	–18.4	1.5	–12.36	< .01
36	–44.9	1.6	–28.19	< .01
37	–13.0	1.9	–6.99	< .01

Matched-pair identifier	Average effect associated with being member of matched pair (relative to reference pair number 41)	Standard error	*t* -value	*p*-value
38	–27.8	2.6	–10.63	< .01
39	–13.8	2.9	–4.78	< .01
40	–19.7	0.6	–34.16	< .01

Note: The degrees of freedom associated with each estimate is 32.
Source: Student achievement data from tests administered as part of the state's accountability system and student demographic data from state data system.

References

American College Testing. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading.* Retrieved July 16, 2008 from http://www.act.org/research/policymakers/pdf/reading_summary.pdf

Alabama Math, Science, and Technology Initiative. (n.d.). *History of AMSTI.* Retrieved January 20, 2008, from http://amsti.org/LinkClick.aspx?fileticket=an35hFwiBN8%3d&tabid=78

Alabama Math, Science, and Technology Initiative Committee. (2000). *Report on the review of literature*. Paper presented to the Alabama State Board of Education, Montgomery, AL.

Alabama State Department of Education. (n.d.). Accountability Reporting System. Retrieved June 10, 2010, from http://www.alsde.edu/Accountability/Accountability.asp

Alabama State Department of Education. (2000). *Survey analysis report: Alabama Math, Science, and Technology Initiative*. Montgomery, AL.

American Association for the Advancement of Science. (1990). *Science for all Americans, Project 2061*. New York: Oxford University Press.

Amrein-Beardsley, A. L. (2006). *Teacher research informing policy: An analysis of research on highly qualified teaching and NCLB.* Retrieved June 2, 2008, from http://www.usca.edu/essays/vol172006/Beardsley%20rev.pdf

Anderson, L. W., & Krathwohl, D. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Augustine, N. R., Vagelos, P. R., & Wulf, W. A. (2005, October). "Rising above the gathering storm: Energizing and employing America for a brighter economic future," National Academies Statements on Science, Technology, and Global Economic Competitiveness to the 109th Congress (First Session).

Battista, M. T. (1999). The mathematical miseducation of America's youth: Ignoring research and scientific study in education. *Phi Delta Kappan, 80,* 424–433.

Bell, S. H., & Bradley, M. C. (2008, March). *Calculating long-run impacts in RCTs that release the control group into the intervention prior to the end of follow-up.* Paper presented at the Annual Research Conference of the Society for Research on Educational Effectiveness, Crystal City, VA.

Bell, S. H., & Bradley, M. C. (2009). *Potential for using experimentally-based impact estimates after the control group in an educational RCT receives the intervention*. Report submitted to the Institute for Education Sciences, Washington, DC.

Bell, S. H., & Bradley, M. C. (in press). Calculating long-run impacts of interventions with staggered implementation once all research samples receive the treatment. *Empirical Economics.*

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments*. New York: Russell Sage Foundation.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, *29*(1), 30–59.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom.* ERIC Digest. ERIC Clearinghouse on Higher Education. Washington, DC: George Washington University. (ERIC Document Reproduction Service No. ED340272). Retrieved April 10, 2009, from http://www.ntlf.com/html/lib/bib/91-9dig.htm

Borman, G. D., Rachuba, L., Datnow, A., Alberg, M., MacIver, M., Stringfield, S., et al. (2000). *Four models of school improvement: Successes and challenges in reforming low-performing, high-poverty Title I schools.* CRESPAR Report No. 48. Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

The Center for Public Education. (2005). *Teacher quality and student achievement: Key lessons learned*. Retrieved January 15, 2008, from http://www.centerforpubliceducation.org/site/c.kjJXJ5 MPIwE/b.1510943/k.9EF1/ Teacher_quality_and_student_achievement_Key_lessons_learned.htm

Chang, C. Y., & Mao, S. L. (1999). Comparison of Taiwan science students' outcomes with inquiry-group versus traditional instruction. *Journal of Educational Research, 92*(6), 340–346.

Edmunds, J. A., & McColskey, W. (2007). *Levers for change: Southeast Region state initiatives to improve high schools*. (Issues & Answers Report, REL 2007-No. 24). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Educational Sciences, U.S. Department of Education. Retrieved October 5, 2009, from http://ies.ed.gov/ncee/edlabs

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109.

Gamoran, A., Anderson, C. W., Quiroz, P. A., Secada, W. G., Williams, T., & Ashmann, S. (2003). *Transforming teaching in math and science: How schools and districts can support change*. New York: Teachers College Press.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945.

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., et al. (2010). *Middle school mathematics professional development impact study: Findings after the first year of implementation* (NCEE 2010-4009). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Glennan, T. K., Jr., & Resnick, L. B. (2004). School districts as learning organizations: A strategy for scaling education reform. In T. K. Glennan, Jr., S. J. Bodilly, J. R. Gallagher, & K. A. Kerr (Eds.), *Expanding the reach of educational reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 517–563). Santa Monica, CA: RAND. Retrieved October 5, 2009, from http://www.rand.org/pubs/monographs/2004/RAND_MG248.pdf

Goldhaber, D. D., & Brewer, D. J. (1996). *Evaluating the effect of teacher degree level on educational performance*. Rockville, MD: Westat.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Haury, D. L. (1993). *Teaching science through inquiry*. ERIC/CSMEE Digest. (ERIC Document Reproduction Service No. ED359048)

Hedges, L., & Hedberg, E. C. (2007). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education, 22*(10). Retrieved January 16, 2008, from http://www.umaine.edu/jrre/archive/22-10.pdf.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406. Retrieved June 2, 2010, from http://sitemaker.umich.edu/lmt/files/hillrowanball.pdf.

James-Burdumy, S., Deke, J., Lugo-Gil, J., Carey, N., Hershey, A., Gersten, R., et al. (2010). *Effectiveness of selected supplemental reading comprehension interventions: Findings from two student cohorts* (NCEE 2010-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., et al. (2009). *Effectiveness of selected supplemental reading comprehension interventions: Impacts on a first cohort of fifth-grade students* (NCEE 2009-4032). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Kennedy, M. M. (2008). Contributions of qualitative research to research on teacher qualifications. *Educational Evaluation and Policy Analysis, 30*(4), 344–367.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., et al. (2004). *International outcomes of learning in mathematics literacy and problem solving: PISA 2003 results from the U.S. perspective.* (NCES 2005–003). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Lord, B., & Miller, B. (2000). *Teacher leadership: An appealing and inescapable force in school reform?* Newton, MA: Education Development Center.

Loucks-Horsley, S., Stiles, K., & Hewson, P. (1996, May). *Principles of effective professional development for mathematics and science education: A synthesis of standards.* (National Institute for Science Education Brief No. 1-1).

Miron, D., & Maxwell, M. (2004). *Impact of AMSTI on student performance on standardized tests: Analysis based on 2003 and 2004 standardized test data.* Tuscaloosa, AL: University of Alabama, Office of Community Affairs.

Miron, D., & Maxwell, M. (2005a). *Report on the performance of AMSTI schools versus controls on the 2005 standardized tests.* Tuscaloosa, AL: University of Alabama, Office of Community Affairs.

Miron, D., & Maxwell, M. (2005b). *Report on the performance of AMSTI schools versus controls on the 2004 and 2005 standardized tests.* Tuscaloosa, AL: University of Alabama, Office of Community Affairs.

Miron, D., & Maxwell, M. (2006a). *Report of* AMSTI *schools versus non-AMSTI schools on 2006 standardized tests*. Tuscaloosa, AL: University of Alabama, Office of Community Affairs.

Miron, D., & Maxwell, M. (2006b). *Report on the performance of the 2005 adopters of* AMSTI *versus controls on the 2006 standardized tests.* Tuscaloosa, AL: University of Alabama, Office of Community Affairs.

Miron, D., & Maxwell, M. (2007). *Report of AMSTI schools compared to non-*AMSTI *schools and the State of Alabama on 2007 standardized tests*. Tuscaloosa, AL: University of Alabama, Office of Community Affairs.

Mullens, J. E., Murnane, R. J., & Willett, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review, 40*, 139–157.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Science Foundation, Division of Science Resources Statistics. (2006a). *Science and engineering indicators 2006, Chapter 1: Elementary and secondary education.* Retrieved April 10, 2008, from http://www.nsf.gov/statistics/seind06/c1/c1s3.htm.

National Science Foundation, Division of Science Resources Statistics. (2006b). *Science and engineering indicators 2006, glossary*. Retrieved April 10, 2008, from http://www.nsf.gov/statistics/ seind06/append/c1/c1g.htm.

National Science Foundation, Division of Science Resources Statistics. (2006c). *Science and engineering indicators 2006, appendix table 1–22.* Retrieved April 10, 2008, from http://www.nsf.gov/statistics/seind06/append/c1/at01-22.xls.

Newmann, F., & Wehlage, G. (1993). Five standards of authentic instruction. *Educational Leadership, 51*(4), 8–12.

O'Sullivan, C. Y., Jerry, L., Ballator, N., & Herr, F. (1997). *NAEP 1996 science state report for Alabama* (NCES 97-499 AL). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

Patton, M. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). *Reformed teaching observation protocol (RTOP)* (Technical Report IN-003). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data Are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Martinez, M., & Spybrook, J. (2007). Strategies for improving precision in group-randomized trials. *Educational Evaluation and Policy Analysis, 29,* 5–29.

Raudenbush, S. W., Spybrook, J., Liu, X., & Congdon, R. (2006). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software* (through version 1.56). Version 2.0, available at http://www.wtgrantfoundation.org/resources/overview/research_tools.

Ricks, S. (2008). *External evaluation results for AMSTI.* Retrieved September 5, 2009, fromhttp://AMSTI.org/LinkClick.aspx?fileticket=heretyK0NAA%3d&tabid=78&mid=464.

Romberg, T. A., Carpenter, T. P., & Dremock, F. (Eds.). (2005). *Understanding mathematics and science matters*. Mahwah, NJ: Lawrence Erlbaum Associates.

SAS Institute. (2006). *SAS/STAT software: Changes and enhancements.* (Through release 9.1). Cary, NC: Author.

Sawyer, R., Koopmans, M., Detgen, A., Rose, L., Orellana, E., & Early, K. (2009). *AMSTI 2008 summer professional development institute report.* Washington, DC: Regional Educational Laboratory Southeast, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Sawyer, R., Koopmans, M., Simkin, L., Detgen, A., Futch, V., & Orellana, E. (2010). *AMSTI 2007 summer professional development institute evaluation report.* Washington, DC: Regional Educational Laboratory Southeast, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schochet, P. Z. (2008). *Technical methods report*: *Guidelines for multiple testing in impact evaluations* (NCEE 2008-4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schochet, P. Z. (2009). An approach for addressing the multiple testing problem in social policy impact evaluations. *Evaluation Review, 33,* 539–567.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57,* 1–22.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23,* 323–355.

Singh, K., Granville, M., & S. Dika. (2002) Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research, 95,* 323–332.

University of Alabama, Office of Community Affairs. *Summary of "Report of the Normative Data on Teacher Content Knowledge Pretests and Posttests from Five 2008 AMSTI Summer Institutes."* Retrieved November 30, 2011, from http://www.amsti.org/LinkClick.aspx?fileticket=B_G9eldn-YA%3d&tabid=133&mid=464