NCEE 2008-4021

U. S. DEPARTMENT OF EDUCATION

The Evaluation of Enhanced Academic Instruction in After-School Programs

Findings After the First Year of Implementation

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

The Evaluation of Enhanced Academic Instruction in After-School Programs

Findings After the First Year of Implementation

June 2008

Authors:

Alison Rebeck Black
Fred Doolittle
Pei Zhu
Rebecca Unterman
MDRC

Jean Baldwin Grossman
Public/Private Venture

Project Officer:

Elizabeth Warner
Institute of Education Sciences

NCEE 2008-4021
U.S. Department of Education

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

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover Whitehurst
Director

National Center for Education Evaluation and Regional Assistance
Phoebe Cottingham
Commissioner

June 2008

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract no. ED-01-CO-0060/0004 with MDRC.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244 or order online at www.edpubs.org.

This report is also available on the IES website at http://ncee.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, call the Alternate Format Center at 202-205-8113.

Contents

Appendixes

List of Exhibits

Box

Table

Figure

Acknowledgments

This study represents a collaborative effort among the authors and the staff from the participating school districts and schools, the curriculum developers — Harcourt School Publishers and Success for All (SFA), our colleagues at MDRC and Public/Private Ventures (P/PV), and Institute of Education Sciences (IES) staff. The study has benefited especially from the time, energy, and commitment put forth by staff in the participating school districts and community-based organizations to implement the two enhanced after-school programs used in the Evaluation of Enhanced Academic Instruction in After-School Programs, to allow access to after-school classrooms, and to respond to requests for data.

Susan Bloom and staff at Bloom Associates and on-site District Coordinators provided invaluable support to the school districts and community-based organizations in their efforts to implement the enhanced after-school programs and meet the demands of the evaluation.

The study's technical working group provided valuable insights on the evaluation's design, data analysis, and early versions of the report. We thank Megan Beckett, Thomas Dee, Carolyn Denton, Larry Hedges, Nancy Jordan, and Rebecca Maynard for their expertise and guidance.

The listed authors of this report represent only a small part of the team involved in this project. Linda Kuhn and the staff at Survey Research Management and local Data Collection Coordinators managed and conducted the baseline and follow-up testing and survey data collection effort.

At P/PV, Karen Walker coordinated the classroom observations and supported the observation staff. Laurie Kotloff and Nikki Johnson processed and managed the interview and observation data. These P/PV staff — along with Amy Arbreton, Julie Goldsmith, Jennifer Kim, Danijela Korom-Djakovic, Jennifer McMaken, Rebecca Raley, and Jessica Sheldon — also conducted classroom observations and interviews with school staff.

At MDRC, Rebecca Kleinman and Adam Wodka assisted with data collection and provided programming and analysis support. In addition, Dan Fallon, Glee Holton, Rebecca Kleinman, and Cynthia Willner conducted classroom observations and interviews with school staff. Gordon Berlin, James Kemple, William Corrin, John Hutchins, Chuck Michalopoulos, and Mary Visher provided substantive expertise through their thoughtful comments on, and reviews of, this report. Mario Flecha and Vivian Mateo assisted with report production. Robert Weber and John Hutchins edited the report, and Stephanie Cowell and Inna Kruglaya prepared it for publication.

The Authors

Disclosure of Potential Conflicts of Interest[1]

The research team for this evaluation consists of a prime contractor, MDRC, Inc., of New York City, NY, and three subcontractors, Public/Private Ventures (P/PV) of Philadelphia, PA, Survey Research Management (SRM) Corporation of Boulder, CO, and Bloom Associates, Inc. of New York. None of these organizations or their key staff has financial interests that could be affected by findings from the evaluation of the two enhanced after-school interventions considered in this report. No one on the Expert Advisory Panel, convened by the research team to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

[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 particular 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.

Executive Summary

This report presents findings, after one year of program implementation, from the Evaluation of Enhanced Academic Instruction in After-School Programs — a two-year intervention and random assignment evaluation of adapted models of regular-school-day math and reading instruction in after-school settings for students in grades 2 through 5. The study, which is being conducted by MDRC in collaboration with Public/Private Ventures and Survey Research Management, was commissioned by the National Center for Education Evaluation and Regional Assistance at the U.S. Department of Education's Institute of Education Sciences (IES).

Federal support for after-school programs is provided through the 21st Century Community Learning Centers (21st CCLC) program, established in 1999 and now a state-administered grant program. A primary purpose of the 21st CCLC program, as expressed in Title IV, Part B, is to "provide opportunities for academic enrichment" to help students meet state and local standards in core content areas. Findings from a previous National Evaluation of the 21st CCLC program indicate that, on average, the 21st CCLC program grants awarded between 1999 and 2002 had a limited academic impact on participating elementary school students' academic achievement.[1] A possible factor is the finding that most academic activities at the evaluation sites consisted of homework sessions in which students received limited additional academic assistance (such as reading instruction or assistance with math homework). In addition, participant attendance was limited and sporadic. However, analyses comparing the academic outcomes of frequent and infrequent participants suggest that increasing attendance alone is unlikely to improve the academic findings. Therefore, the limited academic effects in combination with the low levels of formal academic assistance offered in these programs highlight the need for improved academic programming.

In response, IES has supported the development and evaluation of instructional resources for core academic subjects that could be used in after-school programs. This study tests whether an intervention of structured approaches to academic instruction in after-school programs (for reading and math) produce better academic outcomes than regular after-school services that consist primarily of help with homework or locally assembled materials that do not follow a structured curriculum.[2]

---

[1]M. Dynarski et al., *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program, First-Year Findings.* Report submitted to the U.S. Department of Education. (Princeton, NJ: Mathematica Policy Research, Inc., 2003).

[2]The evaluation is not studying the impacts of the overall after-school program or the enrichment and youth development aspects of after-school services.

Overview of the Interventions

The two interventions being tested in this evaluation involve providing 45 minutes of formal academic instruction during after-school programs to students who need help meeting local academic standards. The model includes the use of research-based instructional material and teaching methods that were especially designed to work in a voluntary after-school setting. Two curriculum developers — Harcourt School Publishers and Success for All — were selected through a competitive process to adapt their school-day materials to develop a math model and a reading model, respectively. The developers were asked to create material that is engaging for students, challenging and tied to academic standards, appropriate for students from diverse economic and social backgrounds, and relatively easy for teachers to use with a small amount of preparation time.

- **Harcourt School Publishers** adapted and expanded its existing school-day materials to develop Harcourt *Mathletics*, in which students progress through material at their own rate, with pretests at the beginning of each topic to guide lesson planning and posttests to assess mastery or the need for supplemental instruction. The model also includes games to build math fluency; hands-on activities; projects; and computer activities for guided instruction, practice, or enrichment.

- **Success for All Foundation (SFA)** adapted its existing school-day reading programs to create *Adventure Island,* a structured reading model with daily lessons that involve switching quickly from one teacher-led activity to the next. It includes the key components of effective reading instruction identified by the National Reading Panel and builds cooperative learning into its daily classroom routines, which also include reading a variety of selected books and frequent assessments built into lessons to monitor progress.

As part of the intervention, these models were also supported by implementation strategies related to staffing, training and technical assistance, and attendance that were managed and supported by Bloom Associates, Inc.

- Sites hired certified teachers and operated the enhanced programs with the intended small groups of students, approximately 10 students per instructor.

- Instructors received upfront training, multiple on-site technical assistance visits, continued support by locally based staff, and daily paid preparation time.

- Efforts were made to support student attendance through close monitoring of attendance; follow-up with parents and students when absences occur, to en-

courage attendance and address issues preventing attendance; and attendance incentives to encourage and reward good attendance.

Research Questions

The primary research question that this evaluation examines is:

- Does the enhanced after-school instruction improve math or reading proficiency over what students would achieve in regular after-school programs, as measured by test scores?

In addition, the evaluation looks at two secondary questions:

- What are the impacts of the enhanced after-school instruction for subgroups of students based on their prior academic performance and grade level?
- Does the enhanced after-school instruction affect other in-school academic behavior outcomes, as measured by reports from regular-school-day teachers of student engagement, behavior, and homework completion?

Subgroup analysis can provide information that might allow for better targeting of the intervention. In particular, the research team hypothesized that the instructional strategies may impact students in the second and third grades (when basic reading and math skills are still being taught during the school day) differently than those in the fourth and fifth grades *and* that those entering the program with higher levels of achievement in the relevant subject may be impacted differently than those entering with lower preintervention achievement levels because of different educational needs.

The final question is important because the enhanced after-school program could change students' behavior in several ways. For example, because the regular after-school program focuses on homework help, one hypothesis is that substituting structured instruction for homework help in the after-school setting has a negative effect on homework completion. On the other hand, improved academic performance might help students in completing homework. There are also theories associating students' behavior in the classroom with their academic performance. One possible hypothesis is that if a student can better understand the academic subject, he or she might be more attentive or less disruptive in class.[3] Another competing hypothesis is that lengthening the academic instruction would introduce fatigue and induce students to act out during class.

[3]T. J. Kane, *The Impact of After-School Programs: Interpreting the Results of Four Recent Evaluations.* William T. Grant Foundation Working Paper. (New York: William T. Grant Foundation, January 16, 2004).

Study Design

This study employs a student-level random assignment design.[4] By randomly assigning students, by grade, within each after-school center to either the enhanced program group to receive 45 minutes of the formal academic instruction or the regular program group to receive the regular after-school services for those 45 minutes, researchers are able to eliminate systematic differences between the two groups of students. Though chance variation may still exist, differences between the groups on the outcomes measured can be attributed to the effect of the enhanced program.

This report presents findings for the first of two years of program operations (school year 2005-2006) on the two parallel studies (one of reading and one of math). The enhanced instruction was implemented in 50 after-school centers — 25 to test the reading program and 25 to test the math program. After-school centers were chosen based on their expressed interest and their ability to implement the program and research design. Assignment of centers to either the reading or the math enhanced program was based on a combination of local preferences, including knowledge of their student needs, sufficient contrast between current academic offerings in the subject area and the enhanced program, and their ability to meet the study sample needs. The centers had to affirm that they were not already providing academic support that involved a structured curriculum or that included diagnostic assessments of children to guide instruction in the subject that they would be implementing (that is, math or reading). The after-school centers are located in 16 sites within 13 states and include schools and community-based organizations in rural areas, in towns, and within the urban fringe of or in midsize to large cities across the country. Participating centers draw students from schools with an average of 78 percent of students receiving free or reduced-price lunches (a measure of low-income status).

The target population for the study is students in second through fifth grades who are behind grade level but not by more than two years. The study sample was recruited from students enrolled in after-school programs who were identified by local staff as in need of supplemental academic support to meet local academic standards. Given that instruction in these programs is provided in a small-group format and is not specifically developed for special needs, students with severe learning disabilities or behavioral problems were excluded from the sample selected. The sample students also had to be able to receive instruction in English. Students who applied to participate in the study were randomly assigned, by grade within their center, to re-

[4]Random assignment was conducted at the student, rather than a higher, level because random assignment at the student level provided more power with which to detect impacts for a given number of after-school programs. Additionally, implementation did not need to be the whole after-school center in order for it to operate. The study team was not concerned about control group contamination because use of the programs required specific training and materials not available to the control group teachers. Furthermore, treatment and control conditions were monitored throughout the study for potential cross-contamination.

ceive either the enhanced model of academic instruction or the services of the regular after-school program. The analysis sample for math includes 1,961 students, and the sample for reading comprises 1,828 students.

Impact findings from the first year are based on data collected from students, regular-school-day teachers, and school records. The Stanford Achievement Test, Tenth Edition (SAT 10), abbreviated battery for math or reading (depending on the intervention implemented), was administered to students at the beginning and end of the school year to measure the gains in achievement. For second- and third-grade students in the reading sample, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) was also administered to measure fluency. A survey of regular-school-day teachers was used to measure student academic behavior.

To help interpret the impact findings, this study also examines how well the special academic services were implemented and whether the enhanced program actually produced a service contrast with what the control group received in the regular after-school program. Thus, the study answers these two questions:

- **Implementation.** How are the after-school academic interventions implemented in the study centers?
- **Service contrast.** What are the measurable differences between services received by students in the enhanced program group and services received by students in the regular after-school program (or control) group?

In addition, the enhanced program was offered in a variety of types of schools. Because the effectiveness of after-school instruction may be related to what the students experience during the regular school day, a third issue is also examined:

- **Linking local school context to impacts.** Are factors related to local school-day context associated with program impacts?

Early Findings for Math

In the first year of the study, Mathletics, the math model put in place in 25 after-school centers, had the following findings:

- The enhanced math program was implemented as intended (in terms of staff characteristics, training, and usage of instructional materials).
- Students received an average of 179 minutes of math instruction per week.

- Math instructors reported that the intended pace of the daily lesson plan was easy to follow.
- The enhanced program provided students with 30 percent more hours of math instruction over the school year, compared with students in the regular after-school program group.
- There are positive and statistically significant impacts for the enhanced math program on student achievement, representing 8.5 percent more growth over the school year for students in the enhanced program group, as measured by the SAT 10 total math score.
- The math program did not produce statistically significant impacts (either positive or negative) on any of the three school-day academic behavior measures: student engagement, behavior, or homework completion.

Implementation of the Enhanced Math Program

Overall, the enhanced math program was implemented as intended in the 25 centers. Each center was expected to hire four certified teachers and to operate with 10 students per instructor. Across the 25 centers, 97 percent of staff were certified, and the programs operated with the intended small groups of students — on average, 9 students per instructor. Staff were trained by Harcourt staff at the beginning of the year and were provided ongoing assistance.[5] They also received paid preparation time. Structured protocol observations of implementation of after-school classes conducted by local district coordinators indicate that 93 percent of observed classes covered the intended content and used the recommended instructional strategies and kept pace with the daily lesson schedule.

The Service Contrast in Math

Students in the enhanced math program were offered and attended a different set of services during the after-school academic time slots than the regular program group.

The enhanced program offered its students academic instruction in math, whereas 15 percent of students in the regular after-school program group were offered academic instruction in math, and the other students received primarily homework help and/or tutoring on multiple subjects. Ninety-seven percent of staff members providing the instruction to the enhanced group

[5]Enhanced math program staff received two full days of upfront training on how to use the math materials, including feedback from the developers in practice sessions using the materials. Ongoing support given to the enhanced program staff consisted of multiple on-site technical assistance visits (an average of three), continued support by locally based staff, and daily paid preparation time of 30 minutes.

students were certified teachers, compared with 62 percent of the regular after-school program staff. Additionally, 94 percent of enhanced program staff received upfront training, and 95 percent received ongoing support, compared with 55 percent and 70 percent of the regular program staff, respectively. These differences are statistically significant at the 0.05 level.

Students in the enhanced program group attended, on average, 49 more hours of academic instruction in math over the course of the school year than the regular program group received (57.17 hours, compared with 8.57 hours). Given estimates of average school-day instruction, this represents an estimated 30 percent more hours of math instruction for students in the enhanced program. Students in the enhanced program group attended 20 percent more days than those in the regular after-school program group, and this difference is statistically significant (effect size = 0.38).

Impacts of the Enhanced Math Program

The main objective of the enhanced after-school math program is to improve student academic performance in math. The analysis looks at impacts on all students in the sample, as well as impacts on two sets of subgroups: students in the two lower grades (second and third) separately from those in the higher grades (fourth and fifth) *and* students who came to the program with higher levels of prior achievement in math separately from those with lower preintervention achievement levels as defined by SAT 10 performance standards of "below basic," "basic," and "proficient."[6]

Access to the enhanced academic after-school math program improved the math performance of students, on average, as measured by the SAT 10, and this finding is statistically significant. In the absence of the intervention, students would have improved their average total math test score by 33.0 scaled score points over the school year.[7] With the intervention, the enhanced program group was able to increase its average test score by 35.8 scaled score points. Therefore, the estimated difference between the enhanced and the regular after-school math

[6]The performance standards are available as part of the SAT 10 scoring. The cut points are criterion-referenced scores. The cuts are created by a panel of teachers based on what they feel a student should be able to do at a particular level of proficiency.

[7]A "scaled score" is a conversion of a student's raw score on a test to a common scale that allows for numerical comparison between students across different forms and levels of the test. The fall-to-spring growth in test scores for the control group (33 scaled score points, based on the abbreviated SAT 10 test) was bigger than the weighted average growth for students in grades 2 through 5 in a nationally representative sample (18 scaled score points, based on the full-length SAT 10 test). Compared with the national sample, both the enhanced program group and the regular program group in the study sample have a higher proportion of low-performing students. (In the math program sample, 78 percent of the students were performing below proficient in math at the beginning of the program.)

program groups is 2.8 scaled score points (effect size = 0.06),[8] which reflects an 8.5 percent difference in growth. Figure ES.1 illustrates this impact.

These statistically significant math impacts are also present across multiple subtests and subgroups. The average scores on the math subtests — problem-solving and procedures — for the enhanced program group are 2.5 scaled score points higher (effect size = 0.05) and 4.3 scaled score points higher (effect size = 0.08), respectively, than the average scores of the regular program group students.

The impact in total math scores from the program for the fourth- and fifth-grade subgroup is 3.9 scaled score points and is statistically significant. For second- and third-graders, the impact is not statistically significant (1.8 scaled score points), although the impacts for the higher and lower grades could not be statistically distinguished. Similarly, the impacts for the prior-achievement subgroups (below basic, basic, and proficient) could not be statistically distinguished. The program impacts on total math scores are 2.9 scaled score points (effect size = 0.06) for the below-basic group; 3.3 scaled score points (effect size = 0.07) for the basic group; and 3.0 scaled score points (effect size = 0.07) for the proficient group. All but the estimate for the basic group (which is approximately half the sample) are not statistically significant.

The analysis also looks at impacts on three measures of student academic behavior — How often do they not complete homework? How often are they attentive in class? How often are they disruptive in class? — for all students in the sample as well as for the two sets of subgroups. Contrary to concerns that the instruction could "overload" students with its academic focus, the findings suggest that enrollment in the enhanced math program did not adversely affect homework completion or the two classroom behavior measures for the full analysis sample or for any of the subgroups, nor did it lead to statistically significant differences in these measures for the enhanced versus the regular program group.

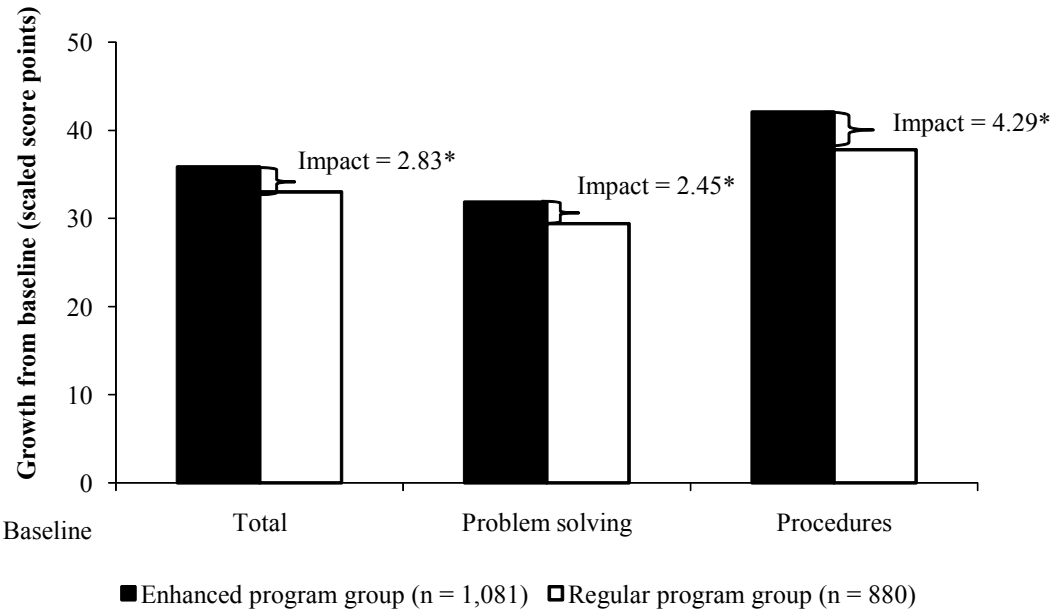
Linking Local School Context to Math Impacts

While the average impact on math test scores is 2.8 scaled score points, not all 25 centers in the study sample experienced this exact difference. Though the study was not designed with the power to detect impacts at the level of individual centers, 17 of the 25 centers did have positive point estimates of Mathletics impacts; 8 of 25 had negative point estimates. Thus, the analysis explored the possibility of variation in impacts for students who attended different types of schools and experienced different program implementation.

[8] "Effect size," which is used widely for measuring the impacts of educational programs, is defined as the impact estimate divided by the underlying population's standard deviation of the outcome measure; effect size is usually measured by the control group's standard deviation.

Figure ES.1

Student Growth on Test Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Math Program



SOURCES: MDRC calculations are from baseline and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The estimated impacts on follow-up results are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. Each dark bar illustrates the difference between the baseline and follow-up SAT 10 scaled scores for the enhanced program group, which is the actual growth of the enhanced group. Each light bar illustrates the difference between the baseline SAT 10 scaled score for the enhanced program group and the follow-up scaled score for the regular program group (calculated as the follow-up scaled score for the enhanced group minus the estimated impact). This represents the counterfactual growth of students in the enhanced group.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect sizes, which are calculated for each outcome as a proportion of the standard deviation of the regular program group, are 0.06, 0.05, and 0.08 for the math total, problem solving, and procedures scores, respectively.

Because the effectiveness of after-school instruction may be related to factors associated with program implementation or what the students experience during the regular school day, a correlational analysis examined the moderating effects of school characteristics and factors of program implementation. It is worth emphasizing that this analysis is nonexperimental and exploratory. Thus, the inference that a particular factor caused or did not cause the impact to be larger or smaller cannot be determined. For example, there could exist factors unaccounted for in the analysis that are correlated with both the program impact and certain school characteristics and that thus account for an observed relationship.

Nonetheless, this analysis uses a regression framework to link program impacts to the following school characteristics: the hours of in-school instruction in the relevant subject, the similarity of the in-school curriculum to the intervention materials, whether the school met its Adequate Yearly Progress (AYP) goals, the proportion of students receiving free or reduced-price lunch, and the in-school student-to-teacher ratio. The analysis also links impacts to two factors of program implementation: the number of days over the course of the school year that the enhanced math program was offered and whether a teacher from the enhanced program left during the school year. Specifically, a regression with interactions between the treatment indicator and each of these school characteristics and factors of program implementation is run to examine how the program impact is moderated by these variables. A chi-square test indicates that, overall, this set of school and implementation characteristics is associated with program impacts on the total math SAT 10 score (p-value = 0.05).

A t-test from the regression analysis shows that, controlling for all these characteristics, centers meeting AYP goals are associated with a higher program impact (p-value = 0.01). Centers serving schools that employ a direct instructional approach organized by lessons with a spiraled curriculum experience lower program impacts than centers that employ a curriculum similar to Mathletics (p-value = 0.03). With the available information, it is not possible to explain the reasons for these relationships.

Finally, individual t-tests from the regression analysis indicate that none of the other measures has a statistically significant relationship to the impacts of the enhanced math program.

Early Findings for Reading

Adventure Island, the reading model put in place in 25 after-school centers, had the following first-year findings:

- The enhanced reading program was implemented as intended (in terms of staff characteristics, training, and usage of instructional materials).

- Students received an average of 176 minutes of reading instruction per week in the reading centers.
- Reading instructors reported that it was difficult to include all aspects of the reading program and maintain the intended pace of the daily lesson plan.
- The enhanced program provided students with 20 percent more hours of reading instruction over the school year, compared with students in the regular after-school program group.
- The students in the enhanced reading program did not experience a statistically significant impact on their performance on the SAT 10 reading test; there are positive and statistically significant program impacts on one of the two measures in the DIBELS fluency test.
- The reading program did not produce statistically significant impacts (either positive or negative) on any of the three school-day academic behavior measures: student engagement, behavior, or homework completion.

Implementation of the Enhanced Reading Program

Overall, the strategies supporting the reading intervention were implemented as intended. Specifically, centers hired certified teachers (across the 25 centers, 99 percent of staff were certified) and operated the programs with the intended small groups of students — on average, 9 students per instructor. Instructors were trained by SFA at the beginning of the year and were provided ongoing assistance and paid preparation time.[9] The district coordinator reports from classroom observations of implementation indicate that 19 percent of Alphie's Lagoon and Captain's Cove classes included four or fewer of the six elements identified as key to intended implementation by the developer; 13 percent of Discovery Bay and Treasure Harbor classes included three or fewer of the five core elements. Observations that included fewer than 70 percent of the core elements indicate that teachers had difficulty delivering specific aspects of the program — in particular, the methods to improve fluency and the ability to cover all the intended lesson elements in the allotted time. In addition, enhanced after-school program staff indicated that the expected pacing of instruction was problematic for daily lessons.

[9]Enhanced reading program staff received two full days of upfront training on how to use the reading materials, including feedback from the developers in practice sessions using the materials. Ongoing support given to the enhanced program staff consisted of multiple on-site technical assistance visits (an average of three), continued support by locally based staff, and daily paid preparation time of 30 minutes.

The Service Contrast in Reading

Students in the enhanced reading program were offered and attended a different set of services during the after-school academic time slot than students in the regular after-school program.

The enhanced program offered its students academic instruction in reading, whereas 12 percent of those in the regular after-school program group were offered academic instruction in reading, and the other students received primarily homework help and/or tutoring on multiple subjects. Ninety-nine percent of staff members providing the instruction to the enhanced group students were certified teachers, compared with 60 percent of the regular after-school program staff. Additionally, 97 percent of enhanced staff received high-quality training to carry out their work, and 95 percent received ongoing support, compared with 58 percent and 55 percent of the regular program staff, respectively. These differences are statistically significant at the 0.05 level.

Students in the enhanced program group attended, on average, 48 more hours of academic instruction over the course of the school year than the regular program group received (55.0 hours compared with 6.54 hours). Given estimates of average school-day instruction, this statistically significant finding represents an estimated 20 percent more hours of reading instruction for students in the enhanced program. Students in the enhanced program group attended 10 percent more days than those in the regular after-school program group, and this difference is statistically significant (effect size = 0.19).

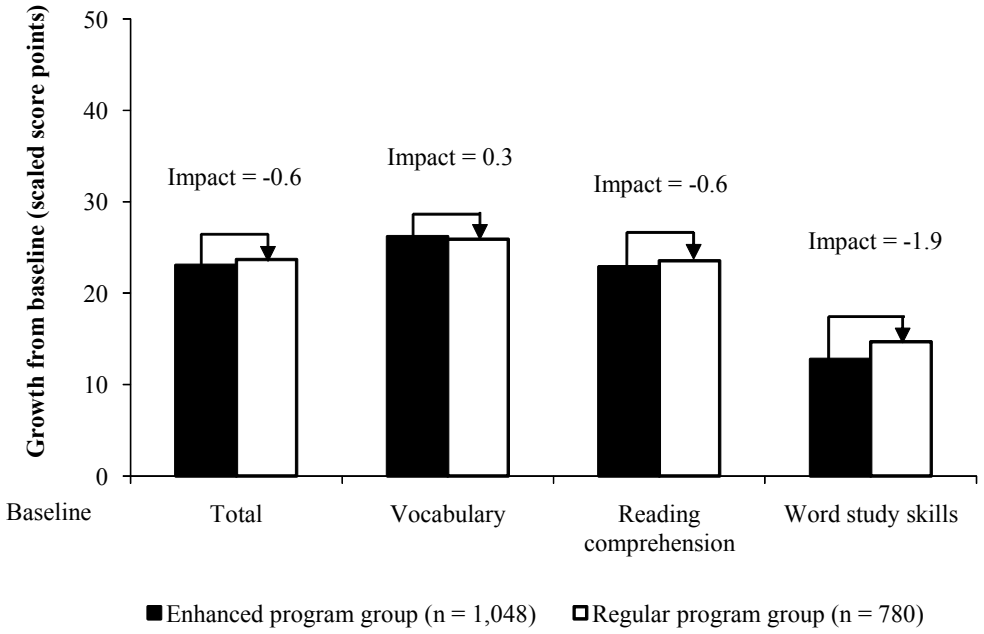
Impacts of the Enhanced Reading Program

Overall, the students in the first year of the enhanced reading program did not experience a statistically significant impact on their performance level on SAT 10 reading tests (total and subtests), above and beyond the level that they would have achieved had there been no enhanced program. This is true for both the full analysis sample and the subgroups defined by grade level and prior achievement. Figure ES.2 illustrates the amount of growth for both the enhanced and the regular program groups in reading over the school year and the lack of a statistically significant difference.

On the other hand, analysis shows that the enhanced reading program produced statistically positive gains in one of two measures of fluency for the younger students in the study sample. The enhanced program group scored 3.7 points higher (effect size = 0.12) in the nonsense word fluency subtest of DIBELS, which targets the alphabetic principle. However, after

Figure ES.2

Student Growth on Test Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Reading Program



SOURCES: MDRC calculations are from baseline and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The estimated impacts on follow-up results are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. Each dark bar illustrates the difference between the baseline and follow-up SAT 10 scaled scores for the enhanced program group, which is the actual growth of the enhanced group. Each light bar illustrates the difference between the baseline SAT 10 scaled score for the enhanced program group and the follow-up scaled score for the regular program group (calculated as the follow-up scaled score for the enhanced group minus the estimated impact). This represents the counterfactual growth of students in the enhanced group.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

Spring administration of the SAT 10 to fifth-graders does not include word study skills. Thus, the sample of students reporting follow-up scores on the word study skills subtest differs from the sample with baseline scores as well as from the sample with follow-up scores on the vocabulary and reading comprehension subtests, which do include fifth-graders.

accounting for multiple comparisons, the estimate is no longer statistically significant.[10] The estimated impact for the oral fluency measure is not statistically significant but is positive (effect size = 0.07).

The analysis also looks at impacts on three measures of student academic behavior — How often do they not complete homework? How often are they attentive in class? How often are they disruptive in class? — for all students in the sample as well as for the two sets of subgroups. Enrollment in the enhanced program did not produce statistically significant impacts on any of these measures for either the full analysis sample or the various subgroups.

Linking Local School Context to Reading Impacts

While there was no overall statistically significant impact on academic achievement for all students in the analysis sample in the first year of the enhanced reading program, not all 25 centers in the study sample experienced this exact impact. Though the study was not designed with the power to detect impacts of Adventure Island at the level of individual centers, 11 of the 25 centers did have positive point estimates; 14 of the 25 had negative point estimates. Thus, the analysis explored the possibility of variation in impacts for students who attended different types of schools and experienced different program implementation.

Because the effectiveness of after-school instruction may be related to factors associated with program implementation or what the students experience during the regular school day, a correlational analysis was conducted to shed light on the possible moderating effects of school characteristics and factors of program implementation. Note that this analysis is nonexperimental and, thus, not causal; inferences drawn from it need to be interpreted with caution. The school characteristics included in the analysis are the hours of in-school instruction in the relevant subject, whether the school met its AYP goals, the proportion of students receiving free or reduced-price lunch, and the in-school student-to-teacher ratio. The analysis also links impacts to two factors of program implementation: the number of days over the course of the school year that the enhanced reading program was offered and whether a teacher from the enhanced program left during the school year.[11] No evidence was found linking the program impact on total reading scores to any of these school environment or implementation characteris-

[10]The DIBELS nonsense word fluency subtest is one of six reading measures estimated for second- and third-grade students. When accounting for multiple test corrections using the Benjamini-Hochberg procedure, this estimate is no longer statistically significant. See Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A New and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B*(57): 1289-1300 (1995).

[11]The types of reading curricula in use in schools during the regular school day were not available in a form that allowed the grouping of centers into categories.

tics. Additionally, the full set of characteristics is not correlated with the program impacts on the total reading SAT 10 score (p-value = 0.71).

Next Steps

The original design of the Evaluation of Enhanced Academic Instruction in After-School Programs called for studying one year of program implementation. However, the study was expanded to include a second year of implementation and data collection using a sample of 15 math centers and 12 reading centers. This sample includes students who were part of the study in the first year and students who were new to the study in the second year, allowing the new wave of data collection to shed light both on the cumulative impact of the enhanced after-school program on returning students and on the impact of a more mature program on new students. Those results will be presented in the final report of the evaluation.

Chapter 1

Overview of the Study

This report summarizes results from the first year of implementation and evaluation as part of a two-year Evaluation of Enhanced Academic Instruction in After-School Programs being conducted by MDRC in collaboration with Public/Private Ventures and Survey Research Management. The study has been commissioned by the National Center for Education Evaluation and Regional Assistance at the U.S. Department of Education's Institute of Education Sciences (IES). This project studies adapted models of regular-school-day reading and math instruction in the after-school setting, examining their impact on student academic outcomes in a sample of 50 after-school centers serving students in grades 2 through 5. A separate team at Bloom Associates, Inc., organized the process of selecting the math and reading model developers for the project and is supporting the implementation of the reading and math interventions in the after-school setting.

The primary purpose of this study is to determine the effects on student academic performance of structured approaches to teaching math and reading in after-school programs. The study includes two school years of implementation of the special instruction and of data collection and analysis. This report presents findings after the first year of implementation, and a final report will present impacts after two years of implementation.

This chapter begins with a description of the origins of the study, the theory of action (or logic model) underlying the instructional approach developed and tested as part of this project, and the key research questions. The chapter then presents the process for selecting developers for reading and math and an overview of the intervention being tested. Chapter 2 describes the study sample selection process, the study design, and the analytic approach. Chapter 3 discusses the implementation of the special math instruction and how it differs from the usual after-school services in the math sites. Chapter 4 presents the impacts of this math intervention. Chapters 5 and 6 present similar findings for the reading intervention.

Origins of the Study

As the pressure for students to meet challenging academic standards grows, parents, principals, and policymakers are increasingly turning their attention to the out-of-school hours as a critical opportunity to help prepare students academically (Bodilly and Beckett 2005; Ferrandino 2007; Miller 2003). Indeed, the federal government has been making a substantial investment toward this goal through its 21st Century Community Learning Centers (21st CCLC) funding. The 21st CCLC program is a state-administered discretionary grant program in which states hold

a competition to fund academically focused after-school programs. Under the No Child Left Behind Act of 2001, the program funds a broad array of before- and after-school activities (for example, remedial education, academic enrichment, tutoring, recreation, and drug and violence prevention), particularly focusing on services to students who attend low-performing schools, to help meet state and local student academic achievement standards in core academic subjects (U.S. Department of Education 2007). A distinguishing feature of after-school programs supported by 21st CCLC funds has been the inclusion of an academic component.

Findings from the earlier National Evaluation of the 21st CCLC program (Dynarski et al. 2003) indicate that, on average, the 21st CCLC program grants awarded between 1999 and 2002 had a limited effect on participating elementary school students' academic achievement. A possible factor is that most academic activities at the evaluation sites consisted of homework sessions in which students received limited additional academic assistance (such as reading instruction or assistance with math homework). In addition, participants' attendance was limited and sporadic. Among the centers examined in the IES study, the average enrollee attended about two days a week. Also, attendance was more frequent at the beginning of the school year but declined as the school year progressed. However, analyses comparing the academic outcomes of frequent and infrequent participants suggest that increasing attendance alone is unlikely to improve the academic findings. Therefore, the limited academic effects in combination with low levels of formal academic assistance offered in these programs highlight the need for improved academic programming.

In response, IES has supported identification and development of instructional resources for core academic subjects that could be used in after-school programs. This study is a test of the effectiveness in improving academic outcomes of two newly developed programs (for reading and math). The evaluation addresses the question of whether structured approaches to academic instruction in after-school programs produce better academic outcomes than the after-school academic support currently used in the sampled centers, which often consists primarily of help with homework or locally assembled materials that do not follow a structured curriculum. The instructional approaches include adaptations of reading and math materials created for the regular school day, which include diagnostic assessments to provide instruction on the topics with which students need the most help and are supported by implementation strategies related to staffing, training, and technical assistance.

The Theory of Action for the Intervention

Low-achieving students lack the fundamental skills needed to advance academically. Though students may attend after-school programs, these often provide some homework help or locally assembled activity but not structured instruction. The theory of action hypothesizes that formal instruction that is focused on key skills, is engaging, and is guided by ongoing assess-

ment of skills — coupled with training and support for teachers and incentives to encourage student attendance — will increase student interactions with adults about academics, increase instructional hours focused on topics students most need help with, and lead to improvements in academic outcomes as measured by achievement tests. An open question posited by the logic model involves effects on academic behavior and homework completion. Extending instruction time after school might positively affect student behavior if students improve academically to grade level and are therefore able to be more attentive and more engaged during the school day, or it might negatively affect behavior if students begin to feel overwhelmed by the amount of time spent on formal academic instruction.[1] Similarly, devoting the first 45 minutes of an after-school program to formal instruction, which might otherwise be focused on homework help, could either reduce homework completion or help students improve academically so they are more able to complete their homework.

Key Research Questions

This evaluation design examines whether the enhanced after-school academic instruction in math and reading makes a difference for students, and it addresses the following key research question:

- Does the enhanced after-school instruction in math or reading improve proficiency in that subject as measured by test scores?

Additionally, the evaluation looks at secondary effects to answer the following questions:

- What are the impacts of the enhanced after-school instruction for subgroups of students based on prior academic performance and grade level?
- Does the enhanced after-school instruction affect other in-school academic behavior outcomes, such as reports by regular-school-day teachers of student engagement, behavior, and homework completion?

Subgroup analysis can provide information that might allow for better targeting of the intervention. In particular, the research team hypothesized that the instructional strategies may impact students in the second and third grades (when basic reading and math skills are still being taught during the school day) differently than those in the fourth and fifth grades *and* that those entering the program with higher levels of achievement in the relevant subject may be impacted differently than those entering with lower preintervention achievement levels, because of different educational needs.

[1]For an example of this latter perspective, see Britsch et al. (2005), which calls for after-school programs to avoid duplicating school-day instruction.

The final question is important because the enhanced after-school program could change students' behavior in several ways. For example, because the regular after-school program focuses on homework help, one hypothesis is that substituting structured instruction for homework help in the after-school setting has a negative effect on homework completion. On the other hand, improved academic performance might help students in completing homework. There are also theories associating students' behavior in classroom with their academic performance. One possible hypothesis is that if a student can better understand the academic subject, he or she might be more attentive or less disruptive in class (Kane 2004). Another competing hypothesis is that lengthening the academic instruction would introduce fatigue and induce students to act out during class.

The Selection of the Instructional Models

Prior to the start of the study, organizations were selected through a competitive process to develop enhanced instruction models in math and reading. In the fall of 2003, a request for proposals (RFP) led to the selection of organizations by the end of January 2004, development of new reading and math models by August 2004, implementation of the models in a small number of pilot sites during the 2004-2005 school year,[2] refinement of the model, and operation of the model in the evaluation sites during the 2005-2006 school year. The project staff placed ads announcing the RFP in key education periodicals, posted it on the Web sites of both the project team and the Department of Education, and sent copies to more than 50 curriculum publishers.

Nine proposals (five in reading and four in math) were received that the project team judged responsive to the requirements of the RFP, and the technical proposals were then all rated by two panels of outside experts (one for the reading program and one for the math program) who have published widely in peer-reviewed journals in the relevant subject.

In February 2004, Harcourt School Publishers was selected to adapt its math materials for use in after-school programs, and Success for All was selected to adapt its reading materials for after-school use.

Overview of the Intervention

The intervention being tested involves providing structured academic support during an initial period of the typical two- to three-hour after-school program schedule. The after-school programs in this study begin with attendance-taking and a snack, followed by some academic

[2]Of the 10 schools that piloted the programs, three are part of the study testing the same program that they implemented in the pilot year, and one school is testing the alternate program. For the three testing the same program, students who participated during the pilot year are not in this study.

support, usually in the form of homework help or tutoring, then enrichment and/or recreational activities. The enhanced after-school math or reading models provide 45 minutes of focused daily instruction that substitutes, by design, for all or a portion of the time devoted to homework completion or other less intensive academic support.

The design of the instructional model included the use of research-based instructional material and teaching methods that were especially designed to work in a voluntary after-school setting. In particular, the planned activities included the following elements:

- Materials consistent with evidence-based research on effective models for reading/math improvement
- Student diagnostic assessment integral to the model[3]
- Content geared to struggling students at multiple levels (Although the enhanced programs span skill levels from kindergarten through grade 5, grades 2 through 5 are the focus of the study.)
- Instruction in a small-group format (a ratio of 10 students to 1 teacher)
- Lessons of 45 minutes duration
- Lessons and exercises that are self-contained within each after-school session
- Materials that can stand alone and be used in settings whether the in-school instruction is similar or different

Recognizing the special circumstances of after-school programs (which come at the end of the school day and are voluntary) and the likely variety of study sites (situated across the entire county), the developers attempted to make the material engaging for students, challenging and tied to academic standards, appropriate for students from diverse economic and social backgrounds, and relatively easy for teachers to use with a small amount of preparation time.

The study includes evaluation of two sets of material that were put in place in the local settings, Harcourt School Publishers for math and Success for All for reading. Below are brief descriptions of the basic structure of each of these models.

Harcourt School Publishers adapted existing school-day materials to develop Harcourt *Mathletics*, a new math model for after-school programs built around five mathematical themes or strands: numbers and operations, measurement, geometry, algebra and functions, and data analysis and probability. Daily 45-minute periods are constructed to mirror a gym exercise

[3]Shepard (2001, pp. 1066-1101).

session, with a short group activity ("the warm-up") followed by 30 minutes focused on skill-building ("the workout") and a final small-group activity to complete the session ("the cool-down"). Students progress through material at their own rate, with pretests at the beginning of each topic to guide lesson planning and posttests to assess mastery or the need for supplemental instruction. The model also includes games to build math fluency; hands-on activities; projects; and computer activities for guided instruction, practice, or enrichment. A key challenge for teachers using this math model is providing differentiated instruction to the students who are working on a variety of skills and activities, depending on their individualized education plan.

Success for All Foundation (SFA) adapted its existing school-day reading programs to create *Adventure Island*, a new reading model for after-school programs built around the theme of a tropical island. Adventure Island is a structured reading model, with prescribed daily activities in each 45-minute lesson that involve switching quickly from one activity to the next. It includes key elements identified by the National Reading Panel (2000): phonemic awareness, phonics, fluency, vocabulary, comprehension, and strategic reading. It builds cooperative learning into its daily classroom routines, which also include reading a variety of selected books and frequent assessments built into lessons to monitor progress. A key component of the reading model is its assessment strategy, which is used to group students by their initial reading level (not by grade), identify skills in need of emphasis in instruction, and reassess students and regroup them depending on student progress. A key challenge for teachers using this reading model is to master the sequence and timing of activities, allowing them to provide a fast-paced daily lesson with the desired mixture of instructional strategies and topic coverage.

To create a strong test of these instructional models, the following implementation strategies were implemented.

Staffing Strategy and Enhanced Program Duration

The staffing strategy calls for instruction by certified teachers and a 10:1 ratio of students to teacher. The instructional model is designed to be used with groups of approximately 10 students four days a week for 45 minutes (a total of 180 minutes per week) and to operate throughout the school year during weeks when the after-school program is in session.[4] Sites hired certified teachers and operated the enhanced programs with the intended small groups of students, approximately 10 students per instructor.

Three-quarters of the after-school enhanced program staff were teachers who taught during regular hours in the same school; others were retired teachers or other school staff, such

[4]Centers in one site implemented the program three days a week for 60 minutes each day, still totaling 180 minutes per week.

as a Special Education teacher, guidance counselor, or staff from a different school within the district. For this reason there was overlap in which students in the enhanced after-school program group were taught by the same teacher during the school day. Among those who did teach in that same school during the school day, up to 55 percent taught grades 2 through 5 during the day and may have taught one or more of the students in the enhanced after-school program during the regular school day.[5]

Support for Instructors

The intended support for instructors included upfront training, multiple on-site technical assistance visits, continued support by locally based staff, and daily paid preparation time. Enhanced group instructors received this training and support in a variety of ways throughout the school year:

- **Local district coordinators.** The project funded a part-time district coordinator for 10 hours per week per school; the district coordinators served up to two centers in each site in the study. These individuals needed to have experience with elementary grade reading or math instruction; some coaching or administrative experience; and familiarity with district policies, personnel, and the population served. As part of their role, they observed instruction, coached teachers, monitored student attendance, recorded and analyzed student data on progress through the curricula, substitute-taught when necessary, and served as a key contact for teachers and Bloom Associates.

- **Initial training.** Prior to the start of the school year, all teachers, district coordinators, and district point people — the lead staff person familiar with the structure and operation of the existing after-school program and the school district housing it — attended a two-day training session organized by Bloom Associates, the operations and technical assistance organization from the project team. The training sessions included an orientation to the project and training on the academic model, conducted by representatives of the developers. The training covered the instructional approaches used in the academic models, the schedule for using the 45-minute blocks of time, an overview of the materials provided to each teacher, examples of instructional approaches and classroom management techniques specific to the developed academic model, guidance on how to use the assessment tools embedded in the model, and opportunities to practice instruction and the use of materials.

[5]Because some second- through fifth-grade staff did not teach the same level after school as they teach during the school day, this 55 percent serves as an upper bound for the amount of overlap.

The point person and local district coordinators received an extra day of training on their role in the project, management aspects of implementing the academic model, and coaching techniques. All but one of the instructors was hired in time for the training, and all of those who had been hired attended.

- **Provision of all materials needed to implement the academic model.** Bloom Associates worked with the developers to provide each teacher with all the materials and supplies needed to use the academic model. These materials were sent to sites in a storage cart on wheels for each instructor so that materials could be moved easily if the teacher was not using her own classroom to teach the enhanced after-school program.

- **Paid daily preparation time.** The study design called for 30 minutes of paid time each day the program met for instructors to prepare.

- **On-site visits from representatives of the developers.** Representatives of Harcourt School Publishers or Success for All visited each site twice during the school year. The first visit occurred four to six weeks after program implementation began, and the second visit was about four months later. These visits lasted one day per school and were usually done in conjunction with visits from Bloom Associates staff. They included observation of instruction, follow-up and specialized training sessions for instructors, review of records on the pace and coverage of instruction, and meetings with the on-site district coordinators and point person.

- **Technical assistance visits by Bloom Associates of the project team.** As part of the visits by the developers (or separately, in some cases), Bloom Associates staff visited the sites twice during the first school year, four to six weeks after program implementation began and then again about four months later. In these visits, Bloom Associates staff met with district coordinators and point people for each site, as well as with the lead teachers (in some centers, a teacher was selected to help with administrative responsibilities), and they attended one of the weekly staff meetings conducted to discuss the implementation of the intervention and any other issues that arose.

- **Weekly phone calls between Bloom Associates and the district coordinators.** These phone calls covered particular problems arising in the sites as well as general issues, like the use of student assessments to guide instruction, the desired pacing of instruction through the materials, differentiated instruction techniques, coaching techniques to improve instruction, and strategies to improve student attendance.

- **Weekly teacher meetings.** District coordinators and a lead teacher in each center organized weekly meetings for instructors to discuss problems they were encountering in instruction, to convey information from the weekly phone calls with Bloom Associates, to address logistical and administrative issues related to scheduling and materials, to identify students with poor attendance, and to discuss upcoming training and technical assistance events.

- **Midyear training.** In January 2006, Bloom Associates organized follow-up training for district coordinators, lead teachers, and point people from each site on special topics arising during the first part of the year. Topics included use of diagnostic tests, pacing of instruction, and coaching techniques. Representatives of the developers also trained any new teachers brought into the project midyear.

Efforts to Support Student Attendance

Unlike in elementary school, attendance in after-school programs is not mandatory. National statistics for the federal 21st Century Community Learning Center (21st CCLC) program, which funds after-school programs, show that attendance rates vary across after-school programs (Naftzger et al. 2006). According to N. Naftzger, in the 2004-2005 school year, for example, 65 percent of all students enrolled in 21st CCLC-funded programs that exclusively served elementary students attended for 30 days or more during that school year (which is the 21st CCLC definition of a "regular attendee").[6]

Given the voluntary nature of participation in after-school programming, the project called for efforts to make the instruction engaging and to support student attendance through close monitoring of attendance; follow-up with parents and students when absences occur, to encourage attendance and address issues preventing attendance; and attendance incentives to encourage and reward good attendance.

In order to do this, sites adopted policies to support attendance in the enhanced instruction. The project team and sites put the following features in place:

- **Monitoring of attendance.** The project collected weekly attendance reports on students in the enhanced program group and provided these reports to Bloom Associates. The reports were discussed with sites in the weekly phone calls between Bloom Associates and the district coordinators, and follow-up

[6]Based on data from the 21st CCLC Profile and Performance Information Collection System, maintained by Learning Points Associates, under the auspices of the Learning Points Associates contract with the U.S. Department of Education to provide analytic support for the 21st CCLC program.

activities — such as phone calls to parents to encourage consistent attendance — were planned.

- **Continued efforts to encourage attendance until a formal withdrawal decision.** Even when a student remained absent from the enhanced program for an extended period, site staff continued to encourage a return to the program. Staff would make periodic contacts with parents to see whether a return were possible and would make sure that parents and students understood that the students could return to the enhanced program even though they had been absent. When there was evidence that a return was not possible — because of circumstances like moving away from the school, a change in child care arrangements that made participation impossible, or health issues — then the site and project staff made a formal determination that a child "withdrew" from the program.

- **Incentive plans.** Each after-school center developed an incentives plan over the summer of 2005 and announced it to families and students in the fall of 2005. The local district coordinator, lead teachers, and district point person were responsible for the operation of the incentive policy. The details of the incentive plans were tailored to local circumstances, but each site plan included:

  - Monthly prize drawings in each class for students with excellent attendance during the month

  - Monthly rewards (for example, a trophy and a party) for the class with the best attendance

  - Weekly prizes and treats that teachers could distribute to students with good attendance and to students who made progress in class (A system of points and rewards is built into Adventure Island, and points earned each week can be spent at the "Ships Store" to buy small prizes or candy. Students in the Mathletics program received points for good attendance and completion of skill packs.)

  - An end-of-year celebration for participating students

Overview of the Regular After-School Services

During the 45 minutes when students in the enhanced after-school math or reading program were receiving structured academic support, regular program group staff who were sur-

veyed reported that students in the regular program group were offered academic activities in the form of homework help, tutoring, or, in some instances, the use of other academic materials. While after-school centers that did provide formal instruction in their regular after-school program were not selected to be part of this evaluation, 27 percent of regular program group staff surveyed indicated that the usual services did include academic instruction in math or reading. However, further analysis of after-school staff survey data indicate that this instruction involved the use of the school-day curricula materials, teacher-made materials, or games and activities — not the use of formal materials designed for the after-school setting.[7]

Survey responses also indicate that about 40 percent of staff (38 percent in the math program sites and 40 percent in the reading program sites) were not certified in elementary education and that, of those, 19 percent in the math program sites and 28 percent in the reading program sites had no prior elementary school teaching experience.

Finally, after-school staff who were surveyed indicated that training and support were provided on an ongoing basis to a little more than half of staff in the regular after-school program.[8] In addition, 64 percent of the math centers and 37 percent of the reading centers did not provide the staff of the regular after-school program paid daily preparation time.

The Structure of the Report

Within the intervention, the math and reading models consist of different elements, leading to somewhat different descriptions and analyses of their implementation. Chapter 2 describes the common measures and methods used to study the intervention, including the process for selecting sites and students into the study, the study design, and the analytic approach. Chapter 3 focuses on the math analysis sample, the implementation of Harcourt Mathletics, and the contrast in services received by students in the enhanced and regular after-school program groups. Chapter 4 presents the impact findings for the math analysis sample and key subgroups and identifies possible factors associated with the impacts. Chapters 5 and 6 present the implementation and service contrast and the findings for the reading analysis sample.

[7]See Chapters 3 and 5 for the details of services offered in the regular after-school program. Specifically, see Figure 3.2 for math and Figure 5.3 for reading.

[8]In the regular after-school program in math centers, 55 percent and 70 percent of staff indicated receiving high-quality training and ongoing support, respectively. In the regular after-school program in reading centers, 58 percent and 55 percent of staff indicated receiving high-quality training and ongoing support, respectively.

Chapter 2

Sample Selection, Study Design, and the Analytic Approach

The Evaluation of Enhanced Academic Instruction in After-School Programs is comparing regular after-school services with after-school services that include a daily 45-minute period of enhanced academic instruction. The key question for the study is whether students given access to these enhanced instructional models have better academic outcomes compared with students who are provided regular after-school services.[9] To answer this question, the project team conducted a purposive selection of after-school programs to conduct an efficacy test of the enhanced math and reading after-school instruction developed for the project. This chapter describes the process for how sites and students were chosen to participate in this study and the methodological details for estimating impacts.

Sample Selection

Within 16 sites (in 13 states), 50 after-school centers were chosen based on their expressed interest and their ability to implement the program and research design as outlined below. Assignment to the reading or math enhanced program was based on a combination of local preferences (including knowledge of their student needs, sufficient contrast between current academic offerings in the subject area and the enhanced program, and their ability to meet the study sample needs). Students attending the after-school program and identified by the schools as needing extra academic support in the subject area assigned to the program were then encouraged to enroll in the study. This section describes the process for recruiting sites and students into the study. For the purposes of this study, a "site" is defined as the organization managing the after-school program, which in 12 sites is a school district and in 4 sites is a community-based organization. Within each site is a minimum of two after-school centers where the after-school study is implemented.[10] Each center is housed in a school.

Criteria and Process for Site Selection

The project team sought after-school programs that were able and willing to implement the math and reading models with reasonable fidelity of intended content and instructional strat-

[9]The study is not examining the impacts of the overall after-school program or of the enrichment and youth development aspects of after-school services.

[10]Within the New York City school district, the program was implemented in two centers, each managed by a different community-based organization.

egies and that were able to meet the research requirements of the project. To identify such sites, the process for site recruitment and selection involved several steps. All 21st Century Community Learning Center (21st CCLC) grantees operating elementary school programs that met the size and maturity criteria — as well as similar after-school programs identified through other contacts (including national organizations and other research networks as well as states) — were notified of the study opportunity. The project team was contacted by more than 300 program operators inquiring about participating in the study and engaged in discussions with additional organizations representing networks of after-school service providers.

Sites were selected that met the following criteria:

- **Serve the desired students.** The target group for the evaluation is students who are from low-income families, attend low-performing schools, and do not currently meet locally defined academic standards.

- **Operate with reasonable administrative stability.** After-school programs had to have been in operation for at least one year (to avoid start-up problems), have committed funding for the upcoming school year, and have the ability to assign a point person and hire district coordinators to work with Bloom Associates, Inc., and provide support to the program staff.

- **Have appropriate facilities.** Sites needed to have access to classrooms, video players, and computers to ensure a physical setting conducive to academic instruction and the use of the math or reading materials.

- **Include staff able to deliver instruction.** The after-school programs were required to have or hire staff members with experience and ability to deliver academic instruction using structured math or reading materials, with a preference for certified elementary school teachers.[11]

- **Have adequate student attendance.** To increase the opportunity for regular and sustained student participation, sites needed to have formal attendance rules creating an expectation of regular student attendance in prior years of operation with after-school programs operating at least four days per week.

- **Operate with needed staffing ratios and schedule.** Sites needed to be able to provide the enhanced academic instruction with a student-to-teacher ratio of approximately 10:1 as well as provide teachers with paid time to prepare lessons and review student work on a daily basis.

[11]This project used certified teachers as a proxy for people with experience using a structured curriculum.

- **Provide the desired service contrast.** Sites were asked about the services provided in their regular after-school program. If they did not use structured materials and provide direct instruction, then it was believed that there would be sufficient contrast between their "business as usual" after-school programs and the enhanced program.

- **Be able to meet research requirements.** Sites must be willing and able to follow the research procedures as to random assignment and data collection and must contribute at least 60 to 80 students, roughly equally distributed across the second through fifth grades, for the research sample.

To economize on project resources for implementation support and data collection, sites were recruited if they could contribute at least two after-school centers serving children in grades 2 through 5. The 16 sites selected for the study, shown on Table 2.1, are geographically dispersed across the country.

Table 2.1

Sites Selected to Implement Mathletics and Adventure Island

Site Name	Location	Number of Centers	
		Mathletics	Adventure Island
Perry County Schools	Marion, AL	2	–
Mt. Diablo Unified School District	Concord, CA	1	2
The Lighthouse Program	Bridgeport, CT	3	–
School District of Palm Beach County	Palm Beach, FL	1	3
McDuffie County Schools	Thomson, GA	2	–
Atlanta Public Schools	Atlanta, GA	4	4
Geary County Schools	Junction City, KS	4	–
Bossier Parish Schools	Bossier City, LA	–	2
Detroit Public Schools	Detroit, MI	–	4
Hands Across Cultures	Espanola, NM	–	2
Builders for the Family and Youth	Brooklyn, NY	–	1
Crown Heights Beacon	Brooklyn, NY	–	1
Norristown Area School District	Norristown, PA	2	2
Hempstead Independent School District	Hempstead, TX	2	–
Bryan Independent School District	Bryan, TX	–	2
West Allis-West Milwaukee School District	West Allis, WI	4	2
Sample size		25	25

Characteristics of Centers

As stated above, the goal of this evaluation is to conduct an efficacy study of the effects of a structured approach to after-school academic assistance. After-school centers were selected because they appeared to have the capacity to provide a "fair test" of the new math or reading enhanced program — that is, it appeared that there would be a clear service contrast between what students in the enhanced classes received and "business as usual." Therefore, because centers were not selected to be a random sample of a larger population of centers, the math analysis does not attempt to generalize statistically beyond the observed sample of 25 centers implementing the new enhanced math program, and similarly for the 25 centers testing the reading program, during the 2005-2006 school year.

Each individual after-school center in the study implemented the enhanced program using *either* the special math model or the special reading model. All but seven centers were operated by school district staff (seven were run by community-based organizations), and all but those in one district received 21st CCLC funding, and all were housed in elementary schools and could arrange the desired staffing for the intervention (certified teachers, small classes, and paid preparation time). The centers selected reported that they were not providing academic support that involved a structured curriculum or that included diagnostic assessments of children to guide instruction in the subject they would be implementing (that is, math or reading). About half the centers (13 of the 25 math centers and 9 of 23 reading centers)[12] provided students receiving the special intervention additional time after the 45-minute instruction period for help with homework. In all but four centers, students receiving the special intervention then participated in other after-school activities, such as recreation, for the remainder of the afternoon schedule.[13] (The actual service contrast in the study centers is discussed in detail for math in Chapter 3 and for reading in Chapter 5.)

The Process for Recruiting Students into the Study

The target population for the study is students in second through fifth grade who are not on grade level but are not more than two years behind grade level. The study sample was recruited from students identified by local staff as in need of supplemental academic support to

[12]Information about whether students received additional time for homework help was not available for two of the reading centers.

[13]In four of the 50 centers (two math and two reading), students go home after the 45-minute instruction period because no other after-school activities are offered.

meet local academic standards,[14] who were signing up for the existing after-school program in each of the study sites and were likely to attend for the full school year.[15]

Among students already applying to the after-school program, local after-school center staff (that is, the district coordinator and teachers) identified those who were in need of additional support to meet local academic standards and asked parents whether they would enroll their children in the study. If fewer than 60 to 80 students were enrolling in the after-school program and were identified as in need of additional support, local after-school center staff worked with regular-school-day teachers and the principal to identify and recruit additional students who met the eligibility criteria for the program.[16] Local data collection staff, who were part of the research team, then worked with eligible students and their parents to complete the study intake process. After completion by the parents of an informed consent form, enrollment form, and contact sheet, the data collection staff administered to the students a baseline achievement test consisting of either the math or the reading portion of the Stanford Achievement Test Series, Tenth Edition (SAT 10) abbreviated battery (depending on the enhanced program being implemented in that center).[17] Once students had completed these steps, they were eligible for the random assignment lottery. The data collection staff then submitted a roster of eligible students to MDRC staff, and MDRC conducted the random assignment lottery using its computerized system and then informed the local after-school center staff of the results.

Study Design

The random assignment design is the best-known way to create comparable groups for program evaluation. By randomly assigning students to either the enhanced program or the regular program group, researchers are able to ensure that there are no systematic differences be-

[14]Local staff used a variety of measures (classroom performance, performance on state or local administered tests) to recommend students for the program. Because the properties and performance standards for these measures may differ from those of the study-administered baseline test, some students identified by local staff as in need of supplemental support tested at levels indicating proficiency on the study-administered baseline test.

[15]Given that instruction in these programs is provided in a small-group format, students selected for the study were required to not have serious learning disabilities or behavioral problems and are able to be instructed in English.

[16]How students were identified varied by center. After-school staff looked at test scores or relied on feedback from the students' regular-school-day teacher to determine whether a student needed additional academic support.

[17]In one site, the school district was already administering the SAT 10 in its schools in the spring as part of a state testing program, and the use of the SAT 10 was prohibited. Thus, at baseline, students took the Ninth Edition of the Stanford Achievement Test Series, and these SAT 9-normed scores were converted to SAT 10-normed scores so that they are comparable with scores for other students in the study.

tween the two groups of students, meaning that any subsequent differences between the groups on the outcomes measured can be fairly attributed to the effect of the enhanced program.

The project was designed to provide each center with materials and training for four enhanced classes, with approximately 10 students attending in each. With limited slots, students were randomly assigned to either the *enhanced after-school program* group, where they were offered the enriched instruction in either reading or math during an initial 45-minute block of time, or to the *regular after-school program* group, where they received the existing academic support services in the participating programs (usually, help with homework).[18] Random assignment was conducted separately by center and grade level ("blocking" by center and grade). In other words, the second-graders at Center A who enrolled in the study were randomly assigned to the enhanced program and regular program groups separately from applicants from other grade levels at that center. This ensured that each grade got the needed ratio to operate the program at capacity. Additionally, the program was designed to test the interventions operating with a student-to-teacher ratio of approximately 10:1. In order to assure attendance of approximately 10 students in the enhanced class, up to 13 students per grade were randomly assigned to the enhanced class. (Appendix A provides additional details about the random assignment process.)

Local district coordinators worked with the enhanced program teachers to enroll students assigned to the enhanced program group in the special math or reading instruction and to assure that members assigned to the regular after-school program group were not included. Throughout the school year, they also monitored program operations to ensure that students in the enhanced program group were not attending the recreational portions of the after-school program while the enhanced classes met and that students in the regular after-school program group were not attending the enhanced academic classes.

Statistical Precision

The statistical precision of an impact estimator is its ability to detect true intervention effects when they exist. A common way to represent statistical precision is through the "minimum detectable effect size" (MDES).[19] An important goal for the design of the study was to ensure that the sample size would be sufficient to allow for estimates of reasonable minimum detectable effect sizes for the math and the reading study samples as well as for subgroups. This study — with a math sample of 1,961 and a reading sample of 1,828 — is equipped to detect impacts as small as a 0.06 standard deviation for each of the study samples. Chapters 3 (for math) and 5 (for

[18]See Chapters 3 and 5 for details on the academic services provided to students in the regular after-school program.

[19]Minimum detectable effect sizes are discussed in greater detail in Appendix B.

reading) present minimum detectable effect sizes given the subgroup sample sizes, and Appendix C (Figures C.1 and C.2) presents more detailed information about the analysis sample.

Subgroup Definition

Subgroup analysis can provide information that might allow for better targeting of the intervention. The research design of this study allows for detecting effects among relevant subgroups of students defined prior to random assignment.[20] The research team defined two sets of subgroups a priori that were believed likely to be differentially impacted by the intervention.

In particular, the research team hypothesized that the instructional strategies may impact students in the second and third grades (when basic reading and math skills are still being taught during the school day) differently than those in the fourth and fifth grades because the students and materials differ[21] *and* that those entering the program with higher levels of achievement in the relevant subject may be impacted differently than those entering with lower preintervention achievement levels, because those at different skill levels have different needs. Given this, impacts were examined for subgroups based on:

- **Grade level.** Combining the younger grades (grades 2 and 3) into one group and the older grades (grades 4 and 5) into another.
- **Prior achievement level.** Using SAT 10 performance standards[22] to define subgroups based on students' prior achievement. This approach divides the sample into four performance groups: below basic, basic, proficient, and advanced. Because few students in the study sample are in the advanced group,[23] this category was dropped from all the achievement-based subgroup analyses, and impact estimates for the other three subgroups are reported.[24]

[20]See Appendix B for detailed power calculation for full-sample and subgroup analyses.

[21]It is possible that the instructional strategies may impact students differently by grade; however, to ensure maximum precision for the subgroup analysis, lower-grade students and upper-grade students are combined.

[22]The performance standards are available as part of the SAT 10 scoring. The cut points are criterion-referenced scores. The cuts are created by a panel of teachers based on what they feel a student should be able to do at a particular level of proficiency.

[23]At baseline, 59 students from the math analysis sample (28 treatments and 31 controls) and 14 students from the reading analysis sample (7 treatments and 7 controls) performed at the advanced level.

[24]This study was initially designed to detect policy-relevant impacts for subgroups that consisted of approximately half the full sample. This three-way split based on achievement levels seems more useful and policy relevant than a two-way split. However, the sample size of these three groups may not be adequate to detect effects. This is discussed further in Chapters 4 and 6.

Overview of the Analytic Approach

As mentioned in Chapter 1, the key question for the study is whether students who are given access to this enriched instruction have better academic outcomes than students who are provided regular after-school services. However, to interpret the impact findings, one must understand how well the actual special academic services received by the enhanced after-school program group during this 45-minute period were implemented and whether the offer of the enhanced program actually produced a service contrast. Thus, the study first answers the questions below:

- **Implementation.** How are the after-school academic interventions implemented in the study centers?
- **Service contrast.** What are the measurable differences between services received by students randomly assigned to the enhanced program group and services received by students assigned to the regular after-school program group? (That is, what is the service contrast?)

A third issue is also examined:

- **Linking local school characteristics to impacts.** Are factors that are related to local school context associated with program impacts?

The enhanced program was offered in a variety of types of schools. Understanding how variation in the regular-school-day context is linked to impacts on achievement can help one interpret the meaning of the overall findings.

The following sections lay out the methodological details for examining implementation and estimating impacts. Included are measures used to gauge implementation, service contrast, academic outcomes of achievement and student behaviors, and links between local school context and impacts, as well as the estimation methods.

Measures

The evaluation draws on multiple data sources — some used exclusively for the analysis of program implementation, some exclusively for the impact analysis, and some for both aspects of the study. Table 2.2 describes the available data, listing each source, the sample and time it was collected, and the information it provides. (See Appendix C for outcome measure response rates.)

The Evaluation of Academic Instruction in After-School Programs

Table 2.2

Data Collected for the Evaluation

Data Source	Sample and Time Collected	Description of Data
After-school program attendance	Data are available for members of the research sample for the 2005-2006 school year.	Daily attendance was collected for all days when the special instruction was offered.
After-school staff surveys	Data are available for all after-school staff providing academic support in the study sites; includes data for approximately 200 staff serving the enhanced program group and 160 staff serving the regular program group; data were collected from February to April 2006.	Surveys cover topics consisting of, but not limited to, staff characteristics (years of education, teaching experience, credentials), the nature of activities they lead or participate in, their experience with the materials they use, and the support they received to implement the services they provide.
Structured protocol observations of reading and math instructional practices	Research staff observed the 45-minute modules of instruction for half the instructors serving the enhanced program group (randomly sampled). Observations were conducted from February to April 2006.	Observers rated the use of instructional practices, which include, but are not limited to, organization, clarity of presentation, modeling of concepts, monitoring student progress, cooperative peer learning, and classroom management practices on a scale of 1 ("definitely needs improvement") to 4 ("outstanding").
Structured interviews with after-school staff	Following the observation of a session (mentioned above), research staff interviewed the instructor. Interviews were conducted from February to April 2006.	Open-ended questions to enhanced staff included, but were not limited to, their perspectives on the strengths and weaknesses of the enhanced program, how their implementation of it has evolved over time, challenges in implementing the enhanced program, how these challenges were addressed, and suggestions for improvement. Staff were systematically asked whether they were able to cover topics at the intended pace during a class period. If not, then a follow-up question was asked, and responses were categorized as follows: consistently a problem, sometimes a challenge, rarely a problem, was a problem initially but is no longer a problem.

(continued)

Table 2.2 (continued)

Data Source	Sample and Time Collected	Description of Data
Structured protocol observations of the implementation of Mathletics and Adventure Island	Data are available for all instructors serving the enhanced program group. Multiple observations were conducted by the local district coordinators to systematically assess whether important aspects of the curriculum occurred during a class period. District coordinators typically observed each enhanced program group instructor three times during school year 2005-2006.	For Mathletics, the observations of implementation protocol includes a checklist of six core instructional elements: sole use of the curricular materials throughout the instructional period, establishment of routines that allow for smooth transitions between the parts of the instructional session and maximizing time-on-task, provision of direct and differentiated instruction during the workout, inclusion of teacher-led warm-ups and cool-downs for all students, use of other work-out components (such as skill packs) appropriately, and inclusions of all the components in the allocated times. For Adventure Island, the observations of implementation protocol includes a checklist of core instructional elements, which are a mixture of procedural factors (use of curricular materials, implementation of cooperative learning strategies, awarding of points to reward cooperative learning and the use of fluency techniques, and completion of lesson plan in the allotted time) and indicators for whether key topics were covered (phonics, fluency, and comprehension).
Student surveys	Data are available for enhanced program and regular program group students. Fielded in late fall 2005 and spring 2006.	Questions cover such issues as receipt of academic support outside regular school hours from sources other than the after-school program, sources of help with homework, sense of adult support and expectations from after-school program staff.
Regular-school-day teacher survey	Data are available for the primary regular-school-day teacher for students in the enhanced program and regular program groups. Fielded in spring 2006.	Regular-school-day teachers answered such questions as: Did students receive individual academic help during the regular school day in reading or math? Did they complete their homework? And how was their behavior in class?
Student achievement test: Stanford Achievement Test Series, 10th ed. (SAT 10), abbreviated battery	Data are available for enhanced program and regular program group students. Fielded in fall 2005 (pre-random assignment) and in spring 2006 (follow-up).	For math sites, total math score and subscales for problem solving and procedures are used in the analyses. For reading sites, total reading score and subscales for vocabulary/word reading, reading comprehension, and word study skills (this last subscale is not available for 5th-graders in the spring) are used in the analyses.

(continued)

Table 2.2 (continued)

Data Source	Sample and Time Collected	Description of Data
Student achievement test: Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	Data are available for enhanced program and regular program 2nd- and 3rd-grade students at centers implementing Adventure Island. Fielded in spring 2006.	Data include measures of oral reading fluency and nonsense word fluency.
Student achievement test: state-administered tests, from regular-school-day student records	Data are available for enhanced program and regular program group students for the 2005-2006 school year.	Data include test scores on local or state standardized tests.
School or district employees	In spring 2007, phone calls were made to school or district employees at schools housing the after-school centers.	Employees were asked the name of the reading and math curricula, and the duration of reading and math instruction, during the 2005-2006 school year.
Common Core of Data	Data for the 2005-2006 school year are available for the schools housing the after-school centers.	Data include characteristics of the school, such as the school setting, student body demographics, and student-to-teacher ratio.
State Department of Education Web sites	Data for the 2005-2006 school year are available for the schools housing the after-school centers.	Data include Adequate Yearly Progress (AYP) status of the schools housing after-school centers.

Implementation Measures

To understand how the interventions were implemented, the project team collected data on the use of the instructional models and on the three strategies used to implement the models.

Use of Special Instructional Models

- **Use of instructional elements.** Did teachers use all the various materials and instructional methods?
- **Pacing.** Were teachers able to keep up with the intended pace of the enhanced program model?
- **Instructional characteristics.** Did teachers deliver the material clearly, interact appropriately with students, and manage their classrooms to focus attention on learning?

Strategies Used to Implement the Models

- **Staffing.** Did sites hire certified teachers and operate the programs with the intended small groups of students, approximately 10 students per instructor?
- **Support for instructors.** Did instructors receive upfront training and continued support and daily paid preparation time?
- **Efforts to support student attendance.** Did staff closely monitor attendance; follow up when absences occurred, to encourage attendance and address issues preventing attendance; and provide attendance incentives to encourage and reward good attendance?

The use of special instructional models includes descriptive measures for three different aspects of teachers' implementation. Information on the use of instructional elements comes from structured protocol observations of implementation. Under the guidance of Bloom Associates, local district coordinators formally observed instructors in each center three times, on average, over the school year.[25] Factors recorded on a check-off list indicate to what extent teachers covered specific core content and instructional strategies of the enhanced program. For Mathletics, core content and instructional strategies include sole use of the curricular materials throughout the instructional period, establishment of routines that allow for smooth transitions

[25]Bloom Associates trained district coordinators to use the structured protocol of instructional practice. The protocol consists of core elements identified by each of the developers as key to implementation. (See Appendix D, Boxes D.1 and D.2.) Each formal observation was conducted by the district coordinator and was based on an observation of the full 45-minute class of academic instruction.

between the parts of the instructional session and maximizing time on task, inclusion of a teacher-led warm-up and cool-down for all students, provision of direct and differentiated instruction during the workout, use of other workout components (such as skill packs) appropriately, and inclusions of all the components in the allocated times. For Adventure Island, core elements are a mixture of procedural factors (use of curricular materials, implementation of cooperative learning strategies, awarding of points to reward cooperative learning and the use of fluency techniques, and completion of the lesson plan in the allotted time) and indicators for whether key topics were covered (phonics, fluency, and comprehension).

Measures of pacing and instructional characteristics were created by the research staff, and data on these measures were collected by the research team using a structured protocol to look at classroom observations of instructional practices and to conduct structured interviews of teachers in the after-school enhanced program. Two randomly selected teachers in each center (half of all instructors in the evaluation) were observed and interviewed between January 30 and April 10, 2006. As part of these structured interviews with after-school staff, each teacher was asked, "Can you get through all the material you need to in each session?" As part of the observations, teachers were rated on different features of their instructional practice (specifically, measures of instructional delivery, classroom management, cooperative learning, and quality of meeting space/material/time) using 4-point scales, with "4" indicating the strongest instructional practices.[26]

As for the implementation strategies, the staffing strategy and support for instructors are measured from data drawn from the survey of the staff teaching the enhanced program classes. Efforts to support student attendance, as mentioned in Chapter 1, involved attendance policies that were put in place. Program attendance is measured with attendance records, for both the enhanced and the regular program groups.

Service Contrast Measures

To measure the differences between services received by students randomly assigned to the enhanced program group and services received by students assigned to the regular after-school program group, the project team collected data that answers the following four questions.

- **Service offerings.** Were there differences in the service offerings?
- **Overall attendance in the after-school programs.** How did the enhanced program affect overall attendance in the after-school program?

[26]The scales used in this study, which have been used in prior research by Public/Private Ventures (P/PV), are discussed in more detail in Appendix D.

- **Hours of academic instruction.** How many more hours of after-school academic instruction did students in the enhanced program group attend, compared with students in the regular program group?

- **Other sources of academic support.** Did regular program students receive academic support from other sources that affected the service contrast in academic instruction produced by the enhanced program?

As a result of the instructional strategies mentioned above, the academic supports offered to students in the regular program group were different from those for students in the enhanced program group, in various ways, including the qualifications of the staff, support provided to the staff, attendance policies, and the nature of the services offered (such as whether the program provided homework help, tutoring, or structured academic support and whether the program focused on math, reading, or mixed subjects). Survey responses of enhanced program staff and regular program staff are used to describe the services received by students in the two programs.[27]

While the designers of the enhanced instruction programs explicitly wanted to make their programs "fun," attendance in both the program and the academic portion are voluntary. To focus on the question of whether the enhanced services affected attendance in the after-school program, the research team collected attendance data for days when the enhanced program met.

The difference in hours of academic instruction addresses the extent to which the offer of the enhanced program actually produced a service contrast in instructional hours, the heart of the designed strategy. This is the key aspect of the service contrast that is important in interpreting the impact findings and is measured by combining two data sources: (1) the attendance of students on the days that academic support was provided and (2) responses from the after-school program staff survey about whether they provided academic instruction in the subject being tested, rather than homework help, or tutoring, or some other approach.[28] For the enhanced program group, all the hours attended by students were academic instruction. For the regular after-school program group, hours were counted as "instructional hours" if staff said that they were providing academic instruction in the subject being tested.

[27]Percentages presented in the following chapters are based on the number of staff who responded to each survey item.

[28]Staff reports of academic instruction are subject to recall and other biases. In addition, given that the primary purpose of this measure is for use in informing implementation support, there are no validity and reliability statistics for it.

If students and parents in the regular after-school program group sought out academic support from alternative out-of-school sources to compensate for not receiving the enhanced instruction, this could dilute the service contrast. In addition, if teachers in the regular school day provided more special academic help to students in the regular after-school program group, this would also lessen the service contrast. Thus, these findings are also important in interpreting the impact findings. Surveys of students and regular-school-day teachers provide information on additional sources of academic support that students might receive outside after-school programs.

Students were surveyed in late fall 2005 and spring 2006 about whether they attended a math or reading class or an activity outside the regular school day that was not part of the after-school program. (The students were not asked to provide details about the class or activity.) They were also asked how many days a week they attended this class or activity.[29] Additionally, in a survey administered once during the spring of the first-year program period, the school-day teachers of sample members were asked whether students received "any special support in reading/math during the school day." They also reported in the survey the number of minutes of individualized instruction that they or an aide provided each sample member in math or reading during the prior week.

Key Outcome Measures

Table 2.3 lists the outcome measures used in the impact analysis. Note that all the listed outcomes are measured at the level of individual students.

The primary tool for gauging student achievement is the SAT 10 abbreviated test.[30] The outcome measure used as the principal measure is the "total" score for the subject that was implemented in the center, but the impacts on the subcomponents of the total — vocabulary or word reading, comprehension, and word study skills for reading; and problem-solving and procedure skills for math — were also examined in case the curricula differentially affect subskills. All SAT 10 test scores are scaled scores so that the scores can be compared across grades.[31]

---

[29]These data are student self-reports of academic support received and are subject to bias inherent in such a method of data collection. It is unclear whether such bias would differ, however, for enhanced program students versus regular program students.

[30]In one site, the school district was already administering the SAT 10 in its schools as part of a state reading program. Thus, at follow-up, the students in this site took the SAT 10 full battery given by their district, and those scores are used in the analysis.

[31]A secondary measure of academic achievement used in sensitivity testing is the student performance on district-administered tests. Not all districts in the study test second-grade students, so results for this measure are based on a subset of the analysis sample. Additionally, because each district uses a different test, scores are rescaled. Appendix E describes the scaling of this measure.

Table 2.3

Key Outcome Measures for the Impact Analysis

Outcome Domain	Reading Outcome	Math Outcome

Student achievement	Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery
	• Reading total scaled scores
	  • Vocabulary (all grades)
	  • Reading comprehension (all grades)
	  • Word study skills (grades 2-4)
	Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (grades 2-3)
	• Oral reading fluency
	• Nonsense word fluency
		Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery
		• Math total scaled scores
		  • Problem solving (all grades)
		  • Procedures (all grades)

Student academic behavior	Regular-school-day teacher survey • Homework completion • Disruptive behavior in regular-school-day class • Attentiveness in regular-school-day class	Regular-school-day teacher survey • Homework completion • Disruptive behavior in regular-school-day class • Attentiveness in regular-school-day class

The measures of student academic behavior — homework completion, attentiveness and nondisruptiveness in class — are drawn from the survey of the sites' regular-school-day teachers. These three measures are included to see whether the enhanced after-school program changes students' behavior in any way.[32] All three measures in this domain are on a scale ranging from 1 to 4, with "1" indicating that the specific behavior never occurred and "4" indicating that it occurred often.[33]

[32]The regular after-school program focuses on homework help. One hypothesis is that substituting structured instruction for homework help in the after-school setting has a negative effect on homework completion. On the other hand, improved academic performance might help students in completing homework. There are also theories associating students' behavior in classroom with their academic performance. One possible hypothesis is that if a student can better understand the academic subject, he or she might be more attentive or less disruptive in class (Kane 2004). Another competing hypothesis is that lengthening the academic instruction would introduce fatigue and induce a student to act out during class.

[33]For a detailed description of outcome measures, see Appendix E.

The outcomes used for subgroup analyses based on characteristics of the students (grade and prior achievement level) are identical to those used for the full sample analysis, with one exception. Because reading fluency — an important reading skill mastered in the early grades — is critical for subsequent reading gains, fluency was also measured for grades 2 and 3 in the reading centers, using two subscales of a standard fluency test, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Thus, for this subgroup, test scores from DIBELS were also used to measure student achievement. A further description of the key outcome measures can be found in Appendix E.

Measures of School Characteristics Potentially Linked to Impacts

Understanding which aspects of local school context or program implementation are associated with program impacts will help readers know what to make of the findings. Thus, the following school characteristic variables were examined: the hours of in-school instruction in the relevant subject, the instructional approach of the curriculum used during the school day, whether the school met its Adequate Yearly Progress (AYP) goals, the proportion of students receiving free or reduced-price lunch, and the in-school student-to-teacher ratio. In addition, the following program implementation variables were examined: the number of days over the course of the school year that the enhanced program was offered and whether a teacher from the enhanced program left during the school year. These characteristics will help demonstrate whether or not school characteristics and program implementation, in general, are associated with impacts.

Analytical Approaches

To examine how well the actual special academic services received by the enhanced after-school program group were implemented, means are calculated from after-school staff survey responses and after-school classroom observation measures. Additionally, the interview data were analyzed to understand what after-school staff teaching the enhanced program thought of the program. All the responses to a particular interview question were examined, and categories (codes) were created that describe the range of responses. Answers were then assigned the appropriate code, and the proportion of respondents in each code was counted.

In order to determine the net effect of the enhanced after-school programs on both the amount of enhanced instruction received by students and the academic outcomes, it is desirable to compare the experiences of a group of students who were exposed to the enhanced program with the experiences of a similar group of students who also applied but were not selected to enroll in the enhanced programs. Since the enhanced program and regular program groups in this study were decided through a random assignment process, on average, the regular program

group students resemble the enhanced program group students in every dimension except that they did not receive the enhanced instruction. Therefore, they represent how students in the enhanced program group would have performed if they had not been selected. (For a detailed explanation of how outcome levels for these two groups are calculated and presented throughout the report, see Box 2.1.) As a result, by calculating the regression-adjusted difference between the enhanced program group and the regular program group, the effects of the enhanced after-school program — above and beyond what the regular after-school program generated for comparable students — were estimated.

All impact results reported in the following chapters come from an Ordinary Least Squares (OLS) regression model that takes into account the characteristics of the random assignment block (grade within center), the students' prior achievement levels, and other student characteristics. The estimated effect reflects the impact of being randomized to the enhanced program instead of the regular after-school program for an average student in the sample.[34] A detailed description of the impact model can be found in Appendix F.

In addition to examining the program impacts on various academic and behavioral outcomes for the analysis sample and for different student subgroups, a hierarchical linear model was used to investigate whether the sizes of the impacts are associated with particular characteristics of the program implementation, the schools that house the after-school centers, and what occurs during the regular-school day. Note that this analysis is not based on the experimental design of the study and is exploratory in nature. Appendix G presents the analytical details of the methodology.

[34]Randomization ensures that the enhanced program students and the regular program students start out similar to each other in terms of baseline test scores and other characteristics. However, there may still be small differences between the groups that are attributable to chance. The model described here adjusts for the small differences that may exist between the groups. The model controls for individual-level pretest measures as well as a student's gender, race/ethnicity, free/reduced-price lunch status, age, whether a student is from a single-adult household, whether a student is overage for grade, and the mother's education level.

Box 2.1

Description of the Calculation and Presentation of Outcome Levels

Throughout the report, when a table is presented to report estimated effects of the after-school programs, the mean outcome levels for the enhanced and the regular program groups are reported, to provide context for interpreting the estimated differences. The program impacts are estimated using an impact regression model that utilizes all available observations from both the enhanced program group and the regular program group, and the mean outcome levels are calculated by using the same impact regression model.

When calculating the regression-adjusted mean outcome levels for the enhanced and regular after-school program groups, the adjustment is made using the observed mean covariate values for the enhanced program group in the estimated impact model. In other words, means for *both* groups are "regression-adjusted" using this common set of baseline covariate values: the *enhanced program group's observed means.*

By adjusting based on the observed mean covariate values for the enhanced program group, the tables report:

- Observed mean outcome levels for students randomly assigned to the enhanced program group
- Regression-adjusted mean outcome levels for students randomly assigned to the regular program group, using the observed mean covariate values for the enhanced program group as the basis for the adjustment

By presenting the observed mean outcome values for the enhanced program group, the discussion is based on the actual mean outcomes for the enhanced program group, and one can compare these levels with those for other reference groups or for the same group of sample members over time. The reported mean outcome level for the regular after-school program group also has a straightforward interpretation: it represents how the enhanced program group students would have performed had they not been selected into the enhanced program. In other words, it represents the "counterfactual."

Throughout the text of this report, when presenting these outcome levels, the discussion refers to the *observed* mean level for the enhanced program group as the ''enhanced program group." The mean value for the counterfactual, or the *regression-adjusted* mean for the regular program group, is referred to as the "regular program group." In addition, tables that present observed means (adjusted only for randomization strata) for both the enhanced program group and the regular program group are included in Appendix F, Tables F.4 and F.8.

Chapter 3

The Implementation of Enhanced After-School Math Instruction and the Contrast with Regular After-School Services

This chapter begins by describing the sample of after-school centers and students for the evaluation of enhanced math instruction. It briefly discusses the characteristics of the centers and then describes in detail the student sample for the evaluation. It continues by describing the design of the enhanced math instruction and then presents findings on how the instruction was implemented. The chapter concludes by comparing the services provided students randomly assigned to the enhanced math program with the services for students randomly assigned to the regular after-school program.

The Math Analysis Sample

Sites in the Math Study Sample

Table 3.1 shows that, out of the 25 schools that house the after-school centers implementing the enhanced math instruction, 10 are located in a large or midsize city, eight are within the urban fringe of a large or midsize city, four are in a large or small town, and three are in a rural area. Slightly less than half the students in the schools are black (44 percent), and approximately one-third (32 percent) are white. While the types of communities surrounding these centers vary, 75 percent of all students in these schools come from low-income families.[35] The average student-to-teacher ratio in these schools is 15:1. Five of the 25 schools did not meet the Adequate Yearly Progress (AYP) goals set by their state under the federal No Child Left Behind Act in school year 2005-2006.[36]

During the regular school day, students in 14 of the schools received 50 to 60 minutes of math instruction, with 11 schools offering more than 60 minutes. (See Table 3.2.) In these

[35]This information comes from the 2005-2006 National Center for Education Statistics' Common Core of Data (CCD), which compiles school-level demographic data, including school locale, ethnicity, and free or reduced-price lunch status. The proportion of low-income families is defined as the proportion of students in a school who are eligible for free or reduced-price lunch. School locale designations fall into one of eight categories: large city, midsize city, urban fringe of a large city, urban fringe of a midsize city, large town, small town, rural (outside core-based statistical area), and rural (inside core-based statistical area).

[36]Data on whether a school met its AYP goals were obtained from each state's Department of Education Web site.

Table 3.1

Characteristics of Schools Housing After-School Centers Implementing Mathletics

School Characteristic	
Number of schools	
School setting[a]	
Large or midsize city	10
Urban fringe of a large or midsize city	8
Large or small town	4
Rural area	3
Schools not making Adequate Yearly Progress (AYP)	5
Composition of student body	
Race/ethnicity of students (%)	
Black	44.33
White	32.02
Hispanic	19.23
Asian	1.87
American Indian	0.41
Low-income students[b] (%)	74.98
Average student-to-teacher ratio	15:1
Sample size (total = 25)	

SOURCES: AYP status was collected from each state's Department of Education Web site. All other school-level characteristics were collected from the Common Core of Data Web site, http://nces.ed.gov/ccd/. All data reflect the 2005-2006 school year.

NOTES: Composition of the student body is calculated by averaging the proportion of students within each school (collected from the CCD) across all schools.

[a]National Center for Education Statistics category designations, retrieved August 8, 2007.

[b]A student is defined as low-income if the student is eligible for free/reduced-price lunch.

schools, the school-day instructional approach varies. Thirteen schools in the study sample use an approach during the day that has a format with math topic sections within chapters in which each section contains guided practice problems, numerous computational problems, and a few application problems (word problems) and a mixed/cumulative review section at the end of each section and chapter (for example, Scott Foresman-Addison Wesley, Harcourt, McGraw-Hill, Houghton Mifflin). Another seven schools use an approach that is unit based (units are longer than chapters) and are investigation driven with comparatively fewer practice problems and involving interconnected subproblems (for example, Every Day Math, Move-It-Math, Real Math). And four schools use a curriculum that employs a direct instructional approach organized by lessons with spiraled curriculum (for example, Saxon).

The Evaluation of Academic Instruction in After-School Programs

Table 3.2

Characteristics of the Regular School Day in Schools Housing After-School Centers Implementing Mathletics

Regular-School-Day Characteristic	Number of Schools
Minutes of math instruction offered	
Number of schools with 60 minutes or less	14
Number of schools with more than 60 minutes	11
Math materials/curricula[a]	
Everyday Mathematics (Wright Group/McGraw-Hill)	
MOVE IT Math	
Real Math (SRA/McGraw-Hill)	
Harcourt	
Houghton Mifflin Math	
McGraw-Hill	
Saxon	
Scott Foresman-Addison Wesley Mathematics	
Sample size (total = 25)	

SOURCES: Data were collected from research staff interviews with point persons and phone calls made to schools and districts in spring 2007.

NOTES: Data reflect grades 2 through 5 only. School and district staff were asked for the names and publishers of the math curricula and the amount of time spent on math instruction in each of grades 2 through 5 during the regular school day in the 2005-2006 school year. Responses regarding curricula varied in specificity and include both curricula names, such as MOVE IT Math, and publishers of curricula, such as McGraw-Hill.

[a]The number of schools using the listed curricula is not presented because some schools use different curricula for different grades.

Characteristics of Students in the Math Study Sample

The process of sample intake and random assignment produced a full-study sample of 2,108 students for the math centers (with 55 percent in the enhanced program group and 45 percent in the regular program group). Collection of follow-up data on student outcomes produced response rates for all data sources above 90 percent, exceeding the target rate of 85 percent. Two-tailed t-tests show that response rates are equivalent for the enhanced and the regular after-school program groups across centers for all outcome measures. (See Appendix C for response rate analysis.) The sample used in the analysis is limited to students with follow-up data from both the evaluation-administered achievement test and the regular-school-day teacher survey.

This results in an analysis sample for math that is 93 percent of the full math study sample.[37] The final analysis sample used throughout this report consists of 1,961 students for the math centers, divided into 1,081 enhanced program students (55 percent) and 880 regular after-school program students (45 percent).[38]

Given these analysis sample sizes, the study is equipped to detect impacts as small as a 0.06 standard deviation. This translates into 2.68 scaled score points on the Stanford Achievement Test Series, Tenth Edition (SAT 10) total math test. The weighted average annual growth for students in grades 2 through 5 in a nationally representative sample is 18 scaled score points, based on the full-length SAT 10 test. Therefore, a 2.68 scaled score point impact is equivalent to 15 percent of the expected improvement of students in grades 2 through 5 nationally.[39] In addition, the minimum detectable difference in effects for a subgroup comprising approximately half the students in the sample is 0.08 standard deviation in math, and the minimum detectable effect size (MDES) for a subgroup of a quarter the size of the full analysis sample is 0.12. Details on MDES calculations, given this sample size, are discussed fully in Appendix B.

Using the demographic data received from the applications, as well as the baseline test scores, Table 3.3 presents the baseline characteristics for those students assigned to the enhanced program and receiving Harcourt School Publishers Mathletics and for those students assigned to the regular after-school program group. It also shows the characteristics of students in key subgroups defined by grade level and by baseline math achievement test score. The information in this table can be used to describe the analysis sample of students and to compare the enhanced and regular program research groups used in the impact analysis.

The math analysis sample is made up of approximately equal numbers of students in the second through fifth grades (sample sizes: 971 for grades 2 and 3; 990 for grades 4 and 5). Like the student body in the schools linked to the after-school centers in the study, most of the sample members are black (46 percent) or Hispanic (26 percent). About half the sample members (47 percent) are male; 19 percent are overage for grade; and 81 percent were eligible for free or reduced-price lunch. About one-third of the students in the math sample (34 percent) lived in a

[37]The sample used in the impact analysis is defined as students who had both a follow-up achievement test score and a teacher survey. Nineteen students are excluded because they have a SAT 10 score but no teacher survey; 110 students are excluded because they have a teacher survey but no SAT 10 score; and 18 students are excluded because they have neither source of follow-up data.

[38]Statistical tests were conducted to determine whether the analysis sample is different from the full sample. While the analysis sample reflects the general characteristics of the full sample, students were less likely to be included in the analysis sample if their families had moved in the two years prior to the start of this study. Since the analysis sample contains about 93 percent of students in the full sample, the results are reflective of the behavior of most of the targeted students. See Appendix C for details.

[39]Note that since the study targets low-performing students, the actual growth in the sample is different from the national average level.

The Evaluation of Academic Instruction in After-School Programs

Table 3.3

Baseline Characteristics of Students in the Math Analysis Sample

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Full analysis sample						
Enrollment						
2nd grade	468	262	206			
3rd grade	503	271	232			
4th grade	510	275	235			
5th grade	480	273	207			
Total	1,961	1,081	880			
Race/ethnicity (%)						
Hispanic		26.11	23.73	2.38	0.05	0.18
Black, non-Hispanic		46.67	45.52	1.14	0.02	0.52
White, non-Hispanic		22.13	25.57	-3.44 *	-0.08	0.03
Asian		0.93	1.24	-0.32	-0.03	0.49
Other		4.17	3.93	0.24	0.01	0.79
Gender (%)						
Male		46.81	46.49	0.32	0.01	0.89
Average age (years)		8.65	8.70	-0.04	-0.03	0.07
Overage for grade[a] (%)		17.58	20.05	-2.48	-0.06	0.15
Free/reduced-price lunch (%)						
Eligible (among information providers)		80.06	78.92	1.13	0.03	0.49
No information provided		3.52	2.27	1.25	0.08	0.11
Average household size		1.92	1.90	0.02	0.02	0.73
Single-adult household (%)		33.46	33.73	-0.27	-0.01	0.90
Mother's education level (%)						
Did not finish high school		17.76	17.60	0.16	0.00	0.93
High school diploma or GED certificate		33.86	32.09	1.77	0.04	0.41
Some postsecondary study		41.63	44.98	-3.35	-0.07	0.13
No information provided		6.75	5.32	1.43	0.06	0.18
SAT 10 math total scaled scores		569.25	569.11	0.14	0.00	0.92
Problem solving		574.30	573.72	0.58	0.01	0.69
Procedures		563.22	563.59	-0.38	-0.01	0.83
Sample size (total = 1,961)		1,081	880			

(continued)

Table 3.3 (continued)

Characteristic	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Grade subgroups					
Grades 2 and 3					
Overage for grade[a] (%)	13.32	15.95	-2.63	-0.07	0.24
Mother's education level (%)					
Did not finish high school	19.14	17.29	1.85	0.05	0.46
High school diploma or GED certificate	33.21	32.10	1.11	0.02	0.71
Some postsecondary study	42.03	45.60	-3.58	-0.07	0.26
No information provided	5.63	5.01	0.62	0.03	0.67
SAT 10 math total scaled scores	538.48	537.74	0.73	0.02	0.69
Problem solving	543.67	543.84	-0.17	0.00	0.93
Procedures	533.10	530.76	2.34	0.04	0.35
Sample size (total = 971)	533	438			
Grades 4 and 5					
Overage for grade[a] (%)	21.72	24.04	-2.33	-0.06	0.38
Mother's education level (%)					
Did not finish high school	16.42	17.93	-1.50	-0.04	0.53
High school diploma or GED certificate	34.49	32.08	2.41	0.05	0.42
Some postsecondary study	41.24	44.37	-3.13	-0.06	0.31
No information provided	7.85	5.62	2.22	0.10	0.16
SAT 10 math total scaled scores	599.18	599.62	-0.44	-0.01	0.82
Problem solving	604.03	602.72	1.31	0.03	0.54
Procedures	592.51	595.56	-3.05	-0.05	0.22
Sample size (total = 990)	548	442			
Prior-achievement subgroups					
Students scoring at below basic level					
Overage for grade[a] (%)	26.78	27.89	-1.11	-0.03	0.80
Mother's education level (%)					
Did not finish high school	21.76	24.53	-2.77	-0.07	0.51
High school diploma or GED certificate	38.49	29.72	8.77	0.19	0.05
Some postsecondary study	33.05	38.69	-5.63	-0.11	0.23
No information provided	6.69	7.06	-0.37	-0.02	0.88
SAT 10 math total scaled scores	542.19	540.48	1.72	0.04	0.21
Problem solving	548.54	546.57	1.97	0.04	0.29
Procedures	531.50	529.39	2.11	0.04	0.34
Sample size (total = 467)	239	228			

(continued)

Table 3.3 (continued)

Characteristic	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Students scoring at basic level					
Overage for grade[a] (%)	16.99	19.69	-2.69	-0.07	0.26
Mother's education level (%)					
Did not finish high school	16.99	18.70	-1.71	-0.04	0.48
High school diploma or GED certificate	34.15	34.63	-0.48	-0.01	0.87
Some postsecondary study	41.50	41.51	-0.01	0.00	1.00
No information provided	7.35	5.15	2.20	0.10	0.15
SAT 10 math total scaled scores	564.37	564.61	-0.24	-0.01	0.76
Problem solving	569.84	569.69	0.16	0.00	0.89
Procedures	557.41	558.83	-1.42	-0.02	0.36
Sample size (total = 1,055)	612	443			
Students scoring at proficient level					
Overage for grade[a] (%)	9.90	12.73	-2.82	-0.07	0.45
Mother's education level (%)					
Did not finish high school	13.37	7.09	6.28	0.16	0.09
High school diploma or GED certificate	29.70	27.70	2.00	0.04	0.70
Some postsecondary study	51.98	60.95	-8.97	-0.18	0.10
No information provided	4.95	4.26	0.69	0.03	0.77
SAT 10 math total scaled scores	602.34	602.53	-0.19	0.00	0.88
Problem solving	605.79	604.97	0.81	0.02	0.69
Procedures	603.21	605.17	-1.96	-0.03	0.49
Sample size (total = 380)	202	178			

(continued)

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs application packet and baseline results on the Stanford Achievement Test Series, 10th ed (SAT 10) abbreviated battery.

NOTES: The estimated differences are regression-adjusted using ordinary least squares, controlling for indicators of random assignment strata. The values in the column labeled "Enhanced Program" are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in the next column are the regression-adjusted means using the observed distribution of the enhanced program group across random assignment strata as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated difference effect size for each characteristic is calculated as a proportion of the standard deviation of the regular program group.

F-tests were calculated for the full analysis sample and each subgroup sample in a regression model containing the following variables: indicators of random assignment strata, math total scaled score, race/ethnicity, gender, free-lunch status, overage for grade, mother's education, mobility, and family size. The F-values are not significant for any of the samples analyzed.

Table 3.3 (continued)

There are 28 enhanced program group students and 31 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]A student is defined as overage for grade at the time of random assignment if a student turned 8 before the start of the second grade, 9 before the start of the third grade, 10 before the start of the fourth grade, or 11 before the start of the fifth grade. This indicates that the student was likely to have been held back in a previous grade.

household with a single adult, and 18 percent of students had a mother who did not finish high school, with 33 percent of students' mothers having a high school diploma or a General Educational Development (GED) certificate. Twenty-four percent of the sample scored at a level defined by the publisher of the achievement test used in this study as "below basic" proficiency; 54 percent scored at the "basic" proficiency level; and 19 percent scored at "proficient."[40]

Students assigned to receive the enhanced math instruction and those assigned to the regular after-school program look similar across all characteristics. The one difference that is statistically significant is that more students assigned to the regular after-school program are white, non-Hispanic, compared with the enhanced program students (26 percent versus 22 percent).

The Implementation of the Enhanced Math Instruction

Students randomly assigned to the *enhanced after-school program* group were offered special math instruction during an initial 45-minute block of time, while students randomly assigned to the *regular after-school program* group received the existing academic support services in the participating programs (usually, help with homework). Both groups received similar services for the remainder of the afternoon schedule. The enhanced math instruction involved use of Harcourt School Publisher's Mathletics, supported by implementation strategies related to staffing, support for instructional staff, and efforts to support student attendance. This section describes how these elements were put in place for the enhanced program group, the implementation challenges encountered, and the response to these challenges.

Harcourt School Publisher's Harcourt Mathletics Math Program

Harcourt School Publishers was selected to adapt its existing *Intervention* materials for an after-school program titled *Mathletics,* built around five mathematical themes, or strands:

[40]These percentages are calculated by dividing the sample size of the three achievement test subgroups in the table by the analysis sample. These three groups sum to 97 percent because 59 students performed at the advanced level on the baseline SAT 10.

numbers and operations, measurement, geometry, algebra and functions, and data analysis and probability. The program is designed to teach prerequisite skills that should have been learned in prior school years but were not mastered by the students needing help in math. The Harcourt math program provides a combination of development of math concepts and of specific math computational skills.

Students are grouped by grade, with separate materials for grades 2 through 5. Daily 45-minute periods are modeled after a gym exercise session. Each class period includes a short *warm-up* problem for all students, followed by two 15-minute *workout rotations* focused on individual skill-building, and a final whole group *cool-down* activity that is directly related to the topic of the warm-up activity, to complete the session.

Students are expected to progress through material during the workout at their own rate, with pretests at the beginning of each topic to guide lesson planning and posttests to assess mastery or the need for supplemental instruction. Four-page, paper-and-pencil instruction and practice packets (called "skill packs") are a part of the program. Pages 1 and 2 of each pack provide instruction on the skill (done with the teacher), alternative instructional methods to convey the concept if a student does not grasp key concepts, guided practice, independent practice, and a quick assessment to determine whether a student is ready to continue working independently. Page 3 includes sections for problem-solving, vocabulary development, conceptual understanding, and a review (including concepts covered earlier), with page 4 presenting an activity for reasoning, problem-solving, and the application of the skill. The program also includes board games; a math card game to build math fluency; hands-on activities; projects; and computer activities for guided instruction, practice, or enrichment. Teachers are trained to use a Planning Guide to diagnose a student's performance on the pretests and to determine which program activities are appropriate for the student. Students chart their daily progress with a "My Math Fitness Plan" chart, which lists assignments and their completion.

In classrooms using the Harcourt Mathletics program, all students participate in the initial warm-up exercise with the teacher. The teacher presents the students with one math problem. Students work independently to solve the problem, and then the teacher goes over the solution to the problem, walking the students through each step and allowing students to volunteer answers. Students then break into small groups or do individual work during the workout section of the class, with two 15-minute rotations. The teacher works in a small group with two to three students on a specific math topic or skill to begin a skill pack in each 15-minute workout rotation, while the remaining students are working on their own on pre- or posttests or completing skill packs or computer math activities; some students work in pairs on math games as well. Over the course of a week, the teacher tries to meet with each student at least twice, with the goal of having students' complete work on at least one or two skill packs per week. After the workout section, students return to the larger group for the cool-down, which again involves the

students independently working on one problem and then reviewing the answer together. Given the structure described, this program requires teachers to set up their classrooms with work stations for the various types of activities and to help students handle the transitions between the activities. Teachers using this math program provide differentiated instruction to the students who are working on a variety of skills and activities, depending on their individualized education plan.

Use of Assessments to Guide Instruction

The program in each grade covers all five math strands, with sections for specific skills within each strand. For example, the second-grade curriculum covers four specific skills under "Place Value: Counting to 100," another five specific skills related to "Place Value: Two-Digit Numbers," and so forth, up to a total of 65 skills across the five math strands. Each small cluster of skills begins with a pretest to determine whether the student should skip the cluster or undertake it and ends with a posttest to determine whether a student has mastered the material or needs additional help. Because students' math skills and learning vary at the outset and some students progress more rapidly than others, this leads to a "spread" in the topics under study in a class of students.

Implementation Findings

This section reports on how Mathletics was implemented in the study centers, drawing on surveys and structured interviews of after-school program staff involved in its operation, conducted by the research staff; structured protocol observations of instructional practice of after-school instructors, conducted by the research staff; structured protocol observations of implementation of Mathletics, conducted by district coordinators; and attendance records.

The Amount of Instruction Offered

Ninety percent of the after-school program staff teaching Mathletics reported on the staff survey that they offered an average of 179 minutes of instruction per week, either in four 45-minute lessons or in three 60-minute lessons. (Ten percent of staff did not respond to the survey; see Appendix C for information about response rates.) The intended amount of instruction was 180 minutes. Table 3.4 provides information on the duration of the Mathletics program. It shows the number of centers offering various numbers of days of Mathletics. All the math centers offered Mathletics for a minimum of 70 days during the school year, with four centers offering 70 to 79 days of instruction, four centers offering 80 to 89 days of instruction, four centers offering 90 to 99 days, 10 centers offering 100 to 109 days, and the remaining three centers offering 110 days or more.

Table 3.4

Duration of Mathletics

Duration	Number of Centers
70 to 79 days	4
80 to 89 days	4
90 to 99 days	4
100 to 109 days[a]	10
110 to 119 days	3
Sample size (total = 25)	

SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs attendance records.

NOTES: The duration of Mathletics may have varied between classes in a center if, for example, an instructor was not present and a substitute was unavailable or due to other after-school logistics unrelated to Mathletics. Mathletics classes that met for a different number of days than the specified duration for their center are noted.

[a]In one of the centers, a class of 8 students (21.62 percent) met for 99 days. A class of 13 students (27.08 percent) from another center met for 94 days.

Teachers' Reactions to the Content of the Program

Instructors were surveyed about Mathletics near the midpoint of the school year. Ninety percent of staff (103 of 115) responded to the survey, and, among those, not all staff answered every question. Staff were asked whether the materials were appropriate for their students. Seventy-four percent of 102 staff responding to the question reported that it was "true," and 19 percent reported that it was "sort of true" that the "materials address the topics students need help on." Eighty-eight percent of 93 staff responding to the question reported that the materials and exercises were at "about the right level of difficulty," with 9 percent of staff saying that the materials were "too easy," and 3 percent saying "too challenging."

Measures of Implementation of Mathletics

The project team collected data on three different aspects of teachers' implementation of the Mathletics program: use of instructional elements, pacing of instruction, and characteristics of instruction.

USE OF INSTRUCTIONAL ELEMENTS

As discussed in Chapter 2, under the guidance of Bloom Associates staff, local district coordinators conducted structured protocol observations of the implementation of Mathletics classes in each center three times, on average, over the school year. The protocol included a checklist of the following: use of the Mathletics materials throughout the instructional period, establishment of routines that allow for smooth transitions between the parts of the instructional session, inclusion of a teacher-led warm-up and cool-down for all students, use of workout components (such as skill packs) appropriately, provision of direct and differentiated instruction during the workout, and inclusions of all the components in the allocated times. Based on these observations, 93 percent of all observed classes used the materials and organized the transitions between the parts of the daily sessions as intended.

PACING OF INSTRUCTION

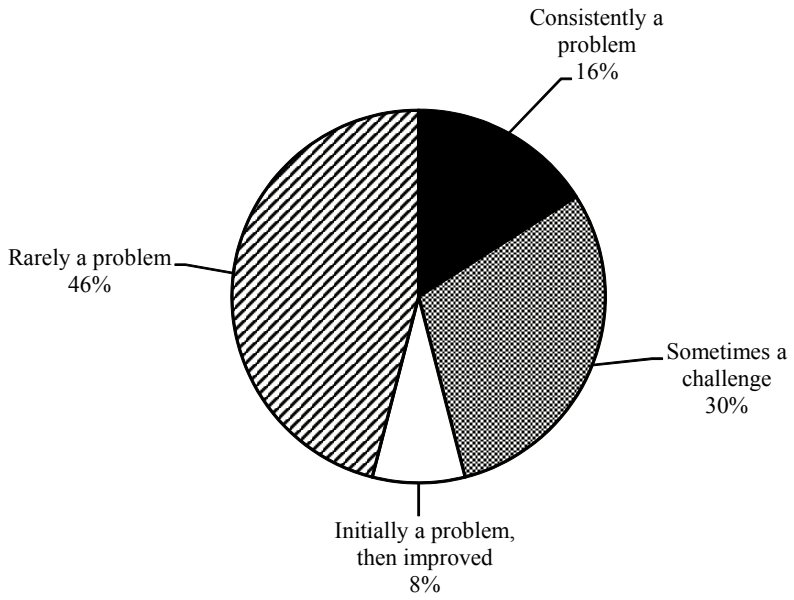
To cover the materials in individual lessons and during the overall school year, teachers needed to maintain the intended pace of instruction. Thus, a second dimension of implementation was whether teachers were able to cover topics at the intended pace during a class period. As part of the field research, two randomly selected teachers in each center (half of all math teachers in the evaluation) were observed teaching a Mathletics class and then were interviewed immediately afterward. As part of the interview, each teacher was asked, "Can you get through all the material you need to in each session?" Fifty of the 51 teachers indicated experiencing some challenges related to pacing. Their responses were categorized as follows: 16 percent described pacing as a "consistent problem" and said that, as a rule, they had trouble completing the daily lesson in the allotted time. Thirty percent indicated that pacing was "sometimes a challenge," whereas 8 percent indicated that they had difficulties with pacing at the beginning of the year but that it was "no longer a problem" for them as they and the students became more familiar with the program. And 46 percent of teachers indicated that they were able to cover the material in the allotted time and that pacing was "rarely a problem" for them. Figure 3.1 reports the answers of the 50 teachers responding to this question.[41]

When teachers who reported that pacing was a challenge at least sometimes were asked to identify what, in particular, they found challenging, 21 of the 23 teachers reported that the 15 minutes allotted for the instructional rotation were not always enough time for all students to master the skill or concept. Five of the 23 teachers pointed out that, for "struggling" students (that is, students who were characterized by teachers as lower performers), the rotation time was especially insufficient.

[41]Fifty-one Harcourt math teachers were interviewed, but one did not respond to this question.

Figure 3.1

Staff Reports of Ability to Complete Material Within Each Session of Mathletics



SOURCE: MDRC calculations are from structured interviews with enhanced program group staff conducted by the research team.

NOTES: A total of 51 enhanced program group staff were randomly sampled. Percentages are based on 50 staff who responded to the interview question "Can you get through all the material you need to in each session?"

CHARACTERISTICS OF INSTRUCTION

Members of the research team observed Mathletics teachers and rated different features of their instructional practice, using 4-point scales, with "4" indicating the strongest instructional practices. Table 3.5 reports the results of these observations of instructional practice.[42]

Eighty-eight percent of teachers using the Harcourt math program were rated 3 or higher on the 4-point scale in presenting an organized sequence of instruction and in the use of the materials, and 78 percent were rated the same in presenting the material clearly (providing clear

[42]The scales used in this study, which have been used in prior research by Public/Private Ventures (P/PV), are discussed in more detail in Appendix D.

Table 3.5

Ratings of Instructional and Classroom Management Practices for Sampled Staff Who Implemented Mathletics

	Percentage of Mathletics Staff	
Classroom Practice	Rated 3	Rated 4
Organizes instruction and use of materials in a logical sequence	23.53	64.71
Presents content clearly	21.57	56.86
Uses modeling to explain material	39.22	29.41
Monitors student progress during direct instruction	45.10	43.14
Monitors student progress during independent work	19.61	3.92
Connects new content to content students already know	31.37	13.73
Includes all students in activities	68.63	7.84
Manages classroom behavior effectively	41.18	41.18
Is responsive to students	45.10	29.41
Sample size (total = 51)		

SOURCE: MDRC calculations are from observations of randomly selected enhanced program classes conducted by the research team.

NOTES: Two staff members from each center were randomly chosen to be observed; the sample reported represents 51 out of 96 staff teaching at any given time. Researchers rated enhanced program staff on a 4-point scale. As a general guide, staff received a score of 4 on a classroom practice if the practice was outstanding, 3 if it was good or very good, 2 if it could use improvement, and 1 if it definitely needed improvement.

directions to students, dividing the material into manageable pieces, and presenting topics in a clear way). Sixty-nine percent were rated 3 or higher in using modeling to explain material (by modeling the day's activity, modeling how to solve math problems, and modeling how to think out key steps).

A high proportion of the math teachers (88 percent) were rated 3 or better on monitoring student progress during their direct instruction in a small group of two or three students during the workout or in whole-group instruction during the warm-up and cool-down, whereas 24 percent were rated 3 or better on monitoring student progress when students were working in-

dependently while the teacher was instructing other students.[43] Seventy-six percent of instructors were rated 3 or 4 in including all students in activities; 82 percent were so rated in managing classroom behavior; and 75 percent were rated 3 or 4 in being responsive to students. However, 45 percent were rated 3 or higher in connecting new content to content already known.

Implementation Strategies Used for Mathletics[44]

STAFFING

There are two key staffing strategies: (1) hiring certified teachers as instructors, with a preference for experienced teachers, and (2) establishing 10:1 student-to-teacher ratios for instruction. Based on responses to the after-school staff survey, 97 percent of Mathletics instructors were certified teachers; 78 percent had more than four years of elementary school teaching experience; 12 percent had three to four years of such experience; and 11 percent had two or fewer years of experience. None of the instructors had no prior elementary school teaching experience.[45]

Random assignment was conducted in a way to produce enhanced program groups of 10 to 13 students per grade, to allow for some attrition and absences and still maintain an average class size of 10 students. When surveyed near the midpoint of the school year, Mathletics instructors reported an average of nine students enrolled in their classes per staff member. When asked, "How many students actually attend this activity on a typical day?" instructors reported that an average of approximately eight students per staff member were present.

Of the 101 teachers hired at the beginning of the school year, there were 11 instances of teachers leaving, spread across six centers. Of the 11 who left, 2 taught second grade; 4 taught third grade; 3 taught fourth grade; and 2 taught fifth grade.[46]

[43]In the Mathletics class, independent student work occurred while the teacher was providing direct instruction to a small group (and was unable to move throughout the room); thus, the program model does not allow for the teacher to easily monitor independent work.

[44]Findings in this section are largely drawn from the after-school staff survey completed in early 2006 by all staff providing academic support to students in the participating after-school centers. Percentages are based on the number of staff who responded to each survey item.

[45]In addition, sites trained a substitute teacher to teach Mathletics, but these individuals are not included in the findings of this chapter unless they replaced a regular teacher prior to the time that the after-school staff survey was fielded.

[46]Four left for professional reasons; for example, they needed to take more courses in order to renew their certificate. Two left for personal reasons, such as needing to take care of a sick family member. There were three instances of a teacher leaving the program due to a conflict with their supervisor and two instances of teachers leaving the program because they specifically did not work well with the after-school math curriculum.

SUPPORT FOR STAFF

Enhanced program group instructors received training and support in a variety of ways throughout the school year. All the instructors were hired in time to attend the summer training on Mathletics prior to the start of the school year. In addition, the training on Mathletics was repeated in January 2006 for teachers on board at that point who had not been trained prior to the summer, and 14 math teachers were trained during the midyear conference (8 replacements for teachers who left and 6 new substitute teachers). Ninety-four percent of Mathletics instructors responding to a staff survey question in early 2006 stated that it was "very true" or "sort of true" that they received high-quality training to carry out their activities.[47]

Another implementation strategy was to provide all materials needed to teach Mathletics so staff would not be burdened by purchasing supplies. Sixty-nine percent of instructors reported that they had enough materials and equipment to carry out their work, with another 21 percent reporting that this was "sort of true." The implementation plan also called for 30 minutes of paid daily preparation time, and 91 percent of instructors reported that they had 30 minutes or more of paid preparation time each day.

The project also provided ongoing, on-site technical assistance, with Harcourt School Publisher representatives visiting each math site twice during the school year; a project-funded, part-time district coordinator to support implementation; and frequent technical assistance from Bloom Associates (one or two on-site visits during the first intervention year and weekly conversations by phone). Ninety-five percent of instructors said that it was "very true" or "sort of true" that they received ongoing support for how to teach children in their activity.[48]

ATTENDANCE

Enhanced program group staff followed up with their after-school students who were absent and provided incentives for students to continue attending. The enhanced math program was offered to students, on average, 95 days over the course of the school year. Students attended, on average, 73 days (or 77 percent of the time of the enhanced math program). Attendance of students in the enhanced program could have been influenced by the special efforts of staff to monitor absences and follow up to encourage attendance, by incentives for good attendance, as well as by Mathletics. Because these are offered together as a package for the enhanced group, it is not possible to disentangle the influence of each factor on attendance; the factors could be offsetting or reinforcing.

[47]Specifically, 65 percent reported "very true," and 29 percent reported "sort of true."
[48]This was made up of 73 percent "very true" and 23 percent "sort of true."

To put these findings on overall attendance in context, they can be compared with the amount of attendance in the previous random assignment impact study of 21st CCLC elementary school programs commissioned by the U.S. Department of Education and conducted by Mathematica Policy Research (Dynarski et al. 2003, 2004).[49] In this earlier study (mentioned in Chapter 1) that examined whether after-school programs led to improved academic achievement, students in the treatment group who participated in the after-school program attended, on average, 58 days during the course of the school year — 15 fewer days than the students in the enhanced math program. More than half (57 percent) of students in the enhanced math program attended more than 75 days, and 25 percent attended 51 to 75 days over the course of the school year, while in the earlier study 39 percent and 15 percent of those participating attended that often, respectively. Nine percent of students in the enhanced math program attended for 25 days or fewer, compared with 27 percent of treatment group participants in the earlier study.[50]

Challenges in Implementing the Mathletics Math Program

In structured interviews with two randomly selected teachers from each math center (half of all math instructors in the evaluation), teachers were asked to identify the challenges that they encountered in implementing the Mathletics math program. They indicated the following concerns.

TIME REQUIRED FOR COMPLETING PAPERWORK, PLANNING, AND PREPARATION

Mathletics provides a "differentiated" program tailored to the learning needs of individual students. Each child's progression through the curriculum is guided by a series of pre- and posttests, which determine his or her instructional level and skill-pack assignments. Teachers have to develop individual plans for each child and decide which children should be grouped together for the following day's 15-minute rotations. Because assignments are determined daily, just about half of the teachers interviewed (26 of the 51) reported that it was difficult for them to accomplish the necessary preparation within the 30-minute paid preparation period during the

[49]Because of differing research questions in the two studies, "attendance" was defined slightly differently. In the current study, attendance was collected on the days when the special instruction met because that was the service contrast being tested in the impact study, not the impact of attendance in the overall after-school program. This means that the "total days attended" count in this study excludes attendances on days that the after-school program operated but that the special instruction was not offered. The Mathematica report collected attendance data for all days that the after-school program operated. This difference in definition means that the difference in attendance in the two studies is somewhat underestimated. In addition, the Mathematica study was designed with a point of random assignment earlier in the intake process for the after-school program, and, therefore, 19.5 percent of the treatment group did not attend the after-school program at all. In the current study, 22 students (or 2 percent) of the enhanced program group did not attend the enhanced program.

[50]See Dynarski et al. (2004, p. 14).

afternoon prior to instruction.[51] Thirty-five percent of the teachers (18 of the 51 sampled) reported that they used more than the allotted time to complete their paperwork and prepare for the following day some of the time (12 reported that they generally used more time than allotted, and 6 reported that they did if, for example, they had tests to score). These teachers reported finishing their preparations at home in the evening, the next morning before school, or during their school-day prep or lunch period.

DIFFICULTY USING COMPUTER-BASED ACTIVITIES

Sixteen (of 51) teachers from six sites reported difficulties using the computer-based activities. The most common problem, reported by 9 of the 16, had to do with some aspect of the *design* of the computer program, such as accessing computer-based activities that matched the student's level. Eight teachers reported other technical problems initially, most of which were subsequently resolved or occurred intermittently. These included problems with malfunctioning computers (reported by six teachers) and lack of compatibility between Mathletics software and local computers (according to two teachers).

OCCASIONAL LACK OF CONSISTENCY BETWEEN AFTER-SCHOOL AND SCHOOL-DAY MATH INSTRUCTION

Thirteen of the 51 teachers reported occasional inconsistencies between math instruction in the school-day and after-school programs, such as how a concept or skill was taught (for example, the detailed procedures for subtraction), the vocabulary used to explain concepts, or difficulties when math topics not yet covered in the school day were introduced in the after-school program.

The Difference in After-School Academic Services Received by the Enhanced Program Group and the Regular Program Group

Math program impacts, which are reported in Chapter 4, are produced by the *difference* between the after-school academic services received by the enhanced program group and those received by the regular, "business as usual" program group. This section describes the academic support services offered to and received by the regular after-school program group and compares these services with those received by students in the enhanced program group.

The service contrast for which impacts are estimated is described through five interrelated findings. First, the *service offerings* differ: 15 percent of the regular after-school group

[51]The program requires daily tasks of scoring tests, documenting the results, determining each child's instructional level, and planning the next session's rotations.

staff offered some form of academic instruction in math. Overall for the regular program group, homework help and/or tutoring on multiple subjects were the most common academic support offered. Second, *staff members* providing the instruction to the enhanced group students were also more likely to be experienced certified teachers and received more training and support for their instruction than staff for the regular program group. Third, *overall attendance* in the after-school program was greater for students in the enhanced program group. Fourth, students in the enhanced program group received more hours of *academic instruction in math,* with the average service difference being 49 hours, or about 30 percent more total math instruction over the course of the school year than the students in the regular program group received. Finally, *academic support from other sources* (during the regular school day or other out-of-school activities) did not lessen the service contrast produced in the after-school program.

The section now focuses on differences in attendance in these services between the two groups, and it concludes with analysis of differences in special academic support received from other sources — during the regular school day and outside school.

Differences in Service Offerings

The academic support offered to students in the regular program group was different from the support for students in the enhanced program group, in various ways, including the nature of the services offered and the staffing strategy, support provided to the staff, and attendance policies. Because sites that provided formal math instruction in their regular after-school program were not selected for the evaluation, the regular or "business as usual" programs described in this chapter are not necessarily indicative of the state of after-school programming in the United States in general but, rather, are a reflection of what comparison group members received in this study.

The previously mentioned survey of after-school staff covered both staff providing the enhanced math instruction and staff providing academically oriented services to students in the regular after-school program. The findings for the regular after-school program group in this section are based on the latter staff's responses to the survey.[52]

Academic Support Services

Regular after-school program staff were surveyed about the nature of the services offered in the regular after-school program. In the math sites, the majority of regular program staff (66 percent) reported focusing on mixed subjects, depending on student needs, by providing help with

[52]Percentages are based on the number of staff who responded to each survey item.

homework or individual or small-group tutoring. Of the remaining staff, 4 percent reported focusing on a single subject other than math, and 30 percent reported a "main focus" on math.

Figure 3.2 presents more detailed information about the services provided by regular program staff. In the math sites, 30 percent of the regular program staff reported a focus on math, with 15 percent (13 instructors) reporting that they provided academic instruction in math (as opposed to tutoring, homework help, or a response of some other type of support). Of the 13 instructors reporting that they provide instruction in math, eight instructors (or 9 percent of all regular program staff) reported that they formally assessed student progress monthly, and nine instructors used student assessments to guide their instruction.[53] Six instructors (or 7 percent of all regular program staff) provided math instruction using a daily lesson plan and supporting materials. Detailed responses to the staff survey provide additional information about the activities/materials of these six after-school staff: two regular after-school program staff reported that they use school-day math curricula; another one uses math games and activities; one mentioned use of unnamed math books; and two mentioned use of math materials created for the after-school setting.

Staff Providing Academic Support Services

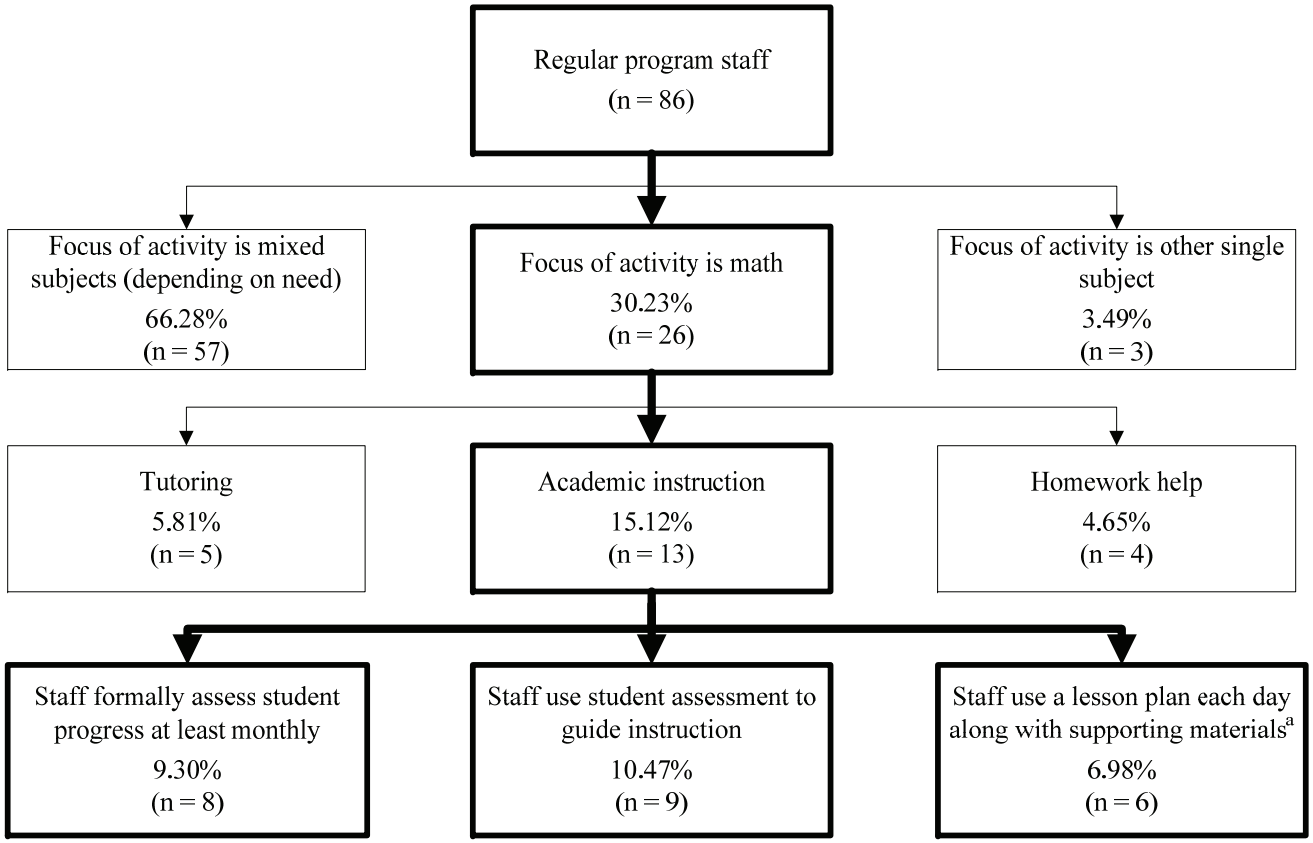
In the regular after-school program, certain staff members were involved in providing academic support to students, while other staff members were primarily involved in enrichment or recreational activities. This and the following sections focus on the staff providing academic support within the after-school program. The findings are based on responses to the after-school staff survey. As shown in the top panel of Table 3.6, 62 percent of regular program staff members were certified teachers (compared with 97 percent of the enhanced program staff), and 64 percent had more than four years of elementary teaching experience (compared with 78 percent of the enhanced program staff), while 11 percent had no prior elementary school teaching experience (compared with none of the enhanced program staff). As the table shows, these differences between the enhanced and regular program staff are statistically significant. Additionally, the enhanced program averaged a student-to-staff ratio of 9:1, while the regular after-school program averaged a student-to-staff ratio of 11:1, with the difference being statistically significant.

The difference in staffing between the enhanced and the regular program groups, which occurred coincident with the implementation of Mathletics, could contribute to program impacts. However, the effect of having more certified experienced teachers and a lower student-to-teacher ratio after school cannot be disentangled from the effect of the implementation of Mathletics.

[53]Frequent assessment to guide instruction and a daily lesson plan are key elements of the *enhanced* instruction curricula.

Figure 3.2

Academic Services Offered by Regular "Business as Usual" After-School Program Staff at Centers Implementing Mathletics



SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs after-school staff survey.

NOTES: Percentages are based on the regular program group sample, comprising 86 staff. Staff for whom values are missing are included in the calculations. Of the 26 staff who focus on math but do not provide academic instruction, tutoring, or homework help, three staff (3.49 percent) responded that they use another method of helping students, and one staff (1.16 percent) did not respond to the question.

[a]Staff responded "sort of true" or "very true" to the question "I have a lesson plan to follow each day, along with supporting materials."

The Evaluation of Academic Instruction in After-School Programs

Table 3.6

Characteristics of After-School Staff and Support for Staff at Centers Implementing Mathletics

Service Offering	Enhanced Program	Regular Program	Estimated Difference		P-Value for the Estimated Difference
Staffing strategy					
Certified in elementary education (%)	97.09	62.35	34.73	*	0.00
Years of elementary school teaching experience (%)					
No experience	0.00	10.59	-10.59		
1-2 years	10.68	20.00	-9.32		
3-4 years	11.65	5.88	5.77		
More than 4 years	77.67	63.53	14.14		
			chi-square[a]	*	0.00
Staff-youth ratio (youth enrolled)	1:9	1:11	-1.85	*	0.02
Sample size (total = 189)	103	86			
Support for staff					
High-quality training to carry out activity[b] (%)	94.06	54.76	39.30	*	0.00
Ongoing support from district for how to teach children in activity[b] (%)	95.10	69.51	25.59	*	0.00
Amount of paid preparation time to carry out activity (%)					
None	0.00	64.29	-64.29		
Less than 15 minutes per day	2.00	11.90	-9.90		
15 minutes to less than 30 minutes per day	7.00	11.90	-4.90		
30 or more minutes per day	91.00	11.90	79.10		
			chi-square	*	0.00
Sample size (total = 189)	103	86			

SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs after-school staff survey.

NOTES: All findings are based on staff self-reports. The values reported for the enhanced program group and the regular program group are the unadjusted means for the staff in each group. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. For service offerings where the table presents the distributions across more than two responses, chi-square tests were used to test whether the distributions for the enhanced program group and the regular program group were the same. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The sample size reported represents the number of staff who filled out a survey. The sample size for each service offering varies by as much as 3 for the enhanced program group and 4 for the regular program group due to nonresponse on particular survey items. Staff for whom values are missing are not included in the calculations.

[a]This chi-square test may not be valid due to small sample sizes within the cross-tabulation.

[b]This presents percentages of after-school staff who responded "sort of true" or "very true" when surveyed.

Support for Staff

As the lower panel of Table 3.6 shows, staff providing academic support in the regular after-school programs were less likely than staff for the enhanced programs to report having received high-quality training to carry out their work (a statistically significant difference of 39 percentage points) or to report receiving ongoing support for how to teach children in their activity (a statistically significant difference of 26 percentage points). In addition, they were less likely to report receiving paid daily preparation time. In the math sites, 64 percent reported getting no paid preparation time at all, and 12 percent reported getting 30 minutes or more. In comparison, 91 percent of the enhanced math program staff received 30 minutes or more of paid preparation time — for a difference of 79 percentage points. A chi-square test found that the differences in the paid preparation time are statistically significant.

Differences in Attendance in the After-School Program

As mentioned above, the sites that were selected for the project all expected enrolled students to attend regularly, and none operated as a drop-in program with sporadic attendance. All regular programs took daily attendance (as required for the 21st CCLC program), but no special staff were assigned to follow up with regular after-school program students who were absent (as the district coordinators did for the enhanced program group).

The first panel in Table 3.7 presents attendance on the days that Mathletics operated. The first row of data shows the number of days attended, and the second row reports average hours of attendance in math instruction offered by the after-school program. The following discussion presents findings for the analysis sample and then for subgroups based on school grade.

Attendance in the After-School Program When Mathletics Operated

STUDENTS IN THE ENHANCED PROGRAM GROUP ATTENDED THE AFTER-SCHOOL PROGRAM MORE THAN STUDENTS IN THE REGULAR PROGRAM GROUP ON DAYS WHEN MATHLETICS OPERATED

In the math sites, students in the enhanced program attended 12 more days over the school year than those in the regular program, a statistically significant difference. For subgroups based on student grade level and baseline achievement, the same pattern of greater attendance among the enhanced program group is present. Findings for subgroups are presented in Appendix H. In math sites for all subgroups, the enhanced program group attended more days than the regular program group, and the differences are statistically significant.

The Evaluation of Academic Instruction in After-School Programs

Table 3.7

Attendance of Students in the Math Analysis Sample

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Attendance in after-school program[a]					
Number of days attended	73.46	61.19	12.26 *	0.38	0.00
Total hours of math instruction received[b]	57.17	8.57	48.60 *	2.76	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	28.68	20.85	7.83 *	0.19	0.00
Number of days per week[d]	0.97	0.59	0.37 *	0.27	0.00
Regular school day[e]					
Students receiving special support (%)	2.24	2.25	-0.01	-0.02	0.69
Minutes per week of individualized help	49.77	48.89	0.88	0.01	0.90
Sample size (total = 1,961)	1,081	880			

(continued)

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs attendance records, student survey responses, and regular-school-day teacher survey responses.

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size for each measure is calculated as a proportion of the standard deviation of the regular program group.

[a]Attendance in the after-school program is based on the days the enhanced program operated.

[b]Students in the enhanced classes received 45 minutes of instruction (and 60 minutes in one site that met only three days a week) on the days they were present. Total hours is calculated for these students by multiplying each student's total days of attendance by 45 (or 60 in the one site).

Students in the regular program group were not supposed to receive any structured instruction. However, some regular program staff indicated on the survey that they provide structured academic instruction. Total hours is calculated for these students by multiplying the total number of days attended by 45, then by the proportion of regular program staff within the center who reported providing structured instruction. If no regular program staff in a center indicated that they provide structured instruction, then total hours for these students in that center is zero. If no regular program staff in a center answered this question, this calculation could not be performed for these students. Calculated as such, the sample size for the regular program group is 770.

Table 3.7 (continued)

[c]This information comes from student survey responses to questions for each day of the week that ask, "Do you go somewhere else for a math class or to be tutored in math?" These calculations are based on a smaller sample than the reported analysis sample by one student in the regular program who did not complete a survey.

[d]Students who responded that they do not receive math support from other out-of-school sources are included in these averages.

[e]This information comes from regular-school-day teacher survey responses. "Special support" refers to special support in math during the school day (that is, pull-out tutoring, remedial math assistance, assigned to a computer assisted lab, and so on). "Individualized help" refers to individual help from the teacher or an aide with a task or answering a question. Teachers who responded that they did not provide support may or may not have responded that they provided minutes of individualized help. Thus, average minutes includes responses for all students, not just those who received special support.

Amount of Academic Instruction Received in the After-School Program

STUDENTS IN THE ENHANCED PROGRAM GROUP ATTENDED MORE HOURS OF ACADEMIC INSTRUCTION IN MATH THAN THOSE IN THE REGULAR PROGRAM GROUP

The average hours of attendance for students in the regular after-school program group reflects the upper-bound estimate that 15 percent of regular after-school program staff reported providing academic instruction in math.

In the math sites, students in the enhanced program group averaged 49 more hours of math instruction than the regular program group, or approximately sixty-five 45-minute sessions, over the course of the school year. This difference is statistically significant. This impact on math instruction is an estimated 30 percent more math instruction when taking into account regular-school-day math instruction. This percentage increase is estimated based upon information on the number of minutes of school-day math instruction reported above in this chapter. More specifically, if students receive 60 minutes per day of instruction (as is common for math) and attend 90 percent of 180 scheduled school days, then they would receive 162 hours of instruction. The 49 hours of extra math instruction is 30 percent more instructional time.

Academic Support in Math from Other Sources

Surveys of students and regular-school-day teachers provide information on two additional sources of academic support that students might receive outside after-school programs. The bottom panel of Table 3.7 contains the findings for academic support from other nonschool sources and during the regular school day.

Support Outside School

STUDENTS IN THE ENHANCED PROGRAM GROUP PARTICIPATED IN MORE CLASSES OR ACTIVITIES IN MATH OUTSIDE SCHOOL THAN STUDENTS ASSIGNED TO THE REGULAR AFTER-SCHOOL PROGRAM GROUP

Students were surveyed in late fall 2005 and spring 2006 about whether they attended a math class or activity outside the regular school day that was not part of the after-school program. (The students were not asked to provide details about the class or activity.) They were also asked how many days a week they attended this class or activity. Results presented here are from the spring survey.[54]

A higher percentage of the enhanced program group reported participating in outside math classes or activities. Twenty-nine percent of the enhanced program group, compared with 21 percent of the regular program group, said that they participated in an outside math class or activity, and the enhanced program group averaged 0.97 day per week of participation, compared with 0.59 day for the regular program group, with differences on both measures being statistically significant.[55]

Support During the Regular School Day

A second way in which the difference in "after-school academic instructional hours" could be diluted was if the school provided extra instruction during the school day to the children who did not get into the enhanced after-school program. To understand whether this occurred, the research team fielded a year-end survey of the school-day teachers of sample members and asked each teacher whether each sample member received "any special support in math during the school day, such as pull-out tutoring, a computer lab, or a special class." They were also asked to report the number of minutes of individualized instruction that they or an aide provided each sample member in math or reading during the prior week.

THERE ARE NO STATISTICAL DIFFERENCES IN THE AMOUNTS OF ACADEMIC SUPPORT DURING THE REGULAR SCHOOL DAY BETWEEN STUDENTS IN THE ENHANCED AND THE REGULAR PROGRAM GROUPS

Enhanced program group students received 50 minutes of individualized instruction per week (10 minutes per day), compared with 49 minutes for the students in the regular program group, but this difference in minutes is not statistically significant. Finally, there is no statistically significant difference in the percentages of students in the enhanced and the regular program groups who received special in-school support in math.

---

[54]There are no statistically significant differences between the findings for the fall and spring student surveys. For simplicity of presentation, this chapter reports only the spring survey responses.

[55]Findings for the grade-level and prior-achievement subgroups are similar. See Appendix H.

Chapter 4

The Impact of Enhanced After-School Math Instruction

The main objective of the enhanced after-school programs was to improve student academic performance in the targeted subject. Based on the theory of action laid out in Chapter 1, this chapter provides impact analysis findings for the math analysis sample and focuses on answering the primary research question: "What are the impacts of the enhanced after-school math instruction (Mathletics) on student achievement?" In addition, secondary program effects on certain student academic behaviors — such as homework completion, attentiveness, and disruptiveness in class — are also analyzed. The chapter then presents exploratory analysis on the associations between the math program impacts and the characteristics of the school.

Program Impacts on Student Academic Achievements and Behaviors

Impacts on Student Academic Achievement

The Stanford Achievement Test, Tenth Edition (SAT 10), abbreviated battery math test was administered to all students in the math analysis sample. Individual test scores on the total test and two subscales — problem-solving and procedures — were collected and used to measure individual student's academic achievement in math.

Table 4.1 shows that enrollment in the enhanced academic after-school math program improved the math performance of students, on average. The average total math scaled score for the enhanced program group is 2.8 points higher than the average scores of those who were not in the enhanced group. This impact translates into an effect size of a 0.06 standard deviation upward shift of the regular program group test scores.[56]

The first pair of bars in Figure 4.1 helps to demonstrate this result. Before the program started, the average total test score among the enhanced program group was 569.3 scaled score

[56]Effect size is used widely for measuring the impacts of educational programs. Here, effect size is defined in terms of the standard deviation of student achievement for the underlying population (the regular program group, in this case).

The Evaluation of Academic Instruction in After-School Programs

Table 4.1

Impact of the Enhanced Math Program on Student Achievement

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Full analysis sample					
SAT 10 math total scaled scores	605.10	602.27	2.83 *	0.06	0.01
Problem solving	606.15	603.71	2.45 *	0.05	0.04
Procedures	605.30	601.01	4.29 *	0.08	0.01
Sample size (total = 1,961)	1,081	880			
Grade subgroups					
Grades 2 and 3					
SAT 10 math total scaled scores	583.23	581.43	1.81	0.04	0.28
Problem solving	584.82	584.04	0.78	0.02	0.64
Procedures	583.55	579.33	4.22	0.08	0.07
Sample size (total = 971)	533	438			
Grades 4 and 5					
SAT 10 math total scaled scores	626.37	622.52	3.85 *	0.09	0.01
Problem solving	626.91	622.73	4.17 *	0.09	0.01
Procedures	626.46	622.19	4.27 *	0.08	0.04
Sample size (total = 990)	548	442			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 math total scaled scores	584.29	581.41	2.87	0.06	0.21
Problem solving	586.30	583.39	2.90	0.06	0.24
Procedures	580.17	577.43	2.74	0.05	0.39
Sample size (total = 467)	239	228			
Students scoring at basic level					
SAT 10 math total scaled scores	600.52	597.23	3.30 *	0.07	0.03
Problem solving	601.74	598.28	3.46 *	0.08	0.04
Procedures	600.63	595.67	4.96 *	0.09	0.02
Sample size (total = 1,055)	612	443			
Students scoring at proficient level					
SAT 10 math total scaled scores	634.67	631.67	3.00	0.07	0.31
Problem solving	634.02	630.40	3.62	0.08	0.22
Procedures	640.08	637.93	2.14	0.04	0.63
Sample size (total = 380)	202	178			

(continued)

Table 4.1 (continued)

SOURCE: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible: for the full analysis sample, scores range from 389 to 796, 414 to 776, and 413 to 768; for the second- and third-grade subgroup, scores range from 389 to 741, 414 to 719, and 413 to 715; and for the fourth- and fifth-grade subgroup, scores range from 450 to 796, 468 to 776, and 485 to 768.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 math total scaled score is calculated as a proportion of the standard deviation of the regular program group, which is 44.64 based on the analysis sample. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.00. For each subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group.

There are 28 enhanced program group students and 31 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

points.[57] In the absence of the intervention, this group of students would have improved their average score over the school year by 33.0 points, to 602.3 scaled score points (as indicated by the light bar in the graph).[58] With the intervention, the enhanced program group was able to increase its average test score over the school year by 35.8 points, to 605.1 scaled score points. Therefore, the estimated difference between the enhanced program group and the regular program group (whose performance stands for what the enhanced program group would have achieved had there

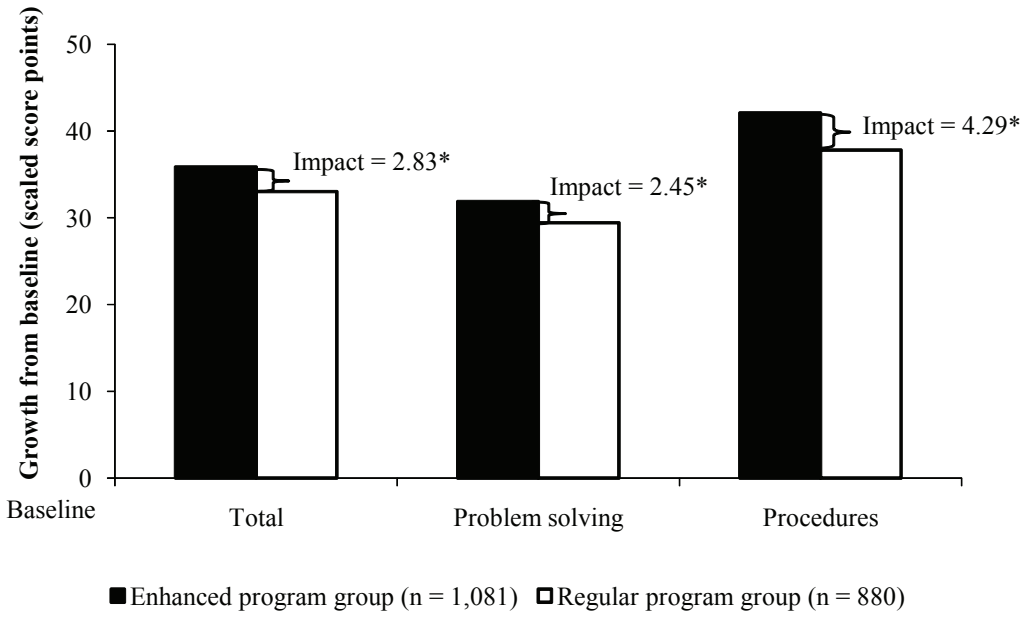
[57]See Chapter 3, Table 3.3: "Baseline Characteristics of Students in the Math Analysis Sample."

[58]The fall-to-spring growth in test scores for the sample (33 scaled score points, based on the abbreviated SAT 10 test) was greater than the weighted average growth for students in grades 2 through 5 in a nationally representative sample (18 scaled score points, based on the full-length SAT 10 test). However, note that the study sample has a higher proportion of low-performing students than the national sample. (At the beginning of the program, 78 percent of the students in the math program sample were performing "below proficient" in math.)

The Evaluation of Academic Instruction in After-School Programs

Figure 4.1

Student Growth on Test Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Math Program



SOURCES: MDRC calculations are from baseline and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The estimated impacts on follow-up results are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. Each dark bar illustrates the difference between the baseline and follow-up SAT 10 scaled scores for the enhanced program group, which is the actual growth of the enhanced group. Each light bar illustrates the difference between the baseline SAT 10 scaled score for the enhanced program group and the follow-up scaled score for the regular program group (calculated as the follow-up scaled score for the enhanced group minus the estimated impact). This represents the counterfactual growth of students in the enhanced group.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect sizes, which are calculated for each outcome as a proportion of the standard deviation of the regular program group, are 0.06, 0.05, and 0.08 for the math total, problem solving, and procedures scores, respectively.

been no intervention at all) is 2.8 scaled score points, which reflects an 8.5 percent difference in growth (2.8 points divided by 33 points), or about three-quarters of an additional month's worth of learning. This estimated difference is statistically significant at the 0.05 level.[59]

To investigate more deeply what types of math knowledge the program affected, the two subtests embedded in the SAT 10 were examined. The program positively affected both of the subtests measured by the abbreviated math SAT 10 test: problem-solving and procedures. The remaining two pairs of bars in Figure 4.1 show that the average scores in problem-solving and procedures for the enhanced program group are 2.5 scaled score points higher (effect size = 0.05) and 4.3 scaled score points (effect size = 0.08) higher, respectively, than those of the regular program group students, and these differences are statistically significant.

To determine whether the program was effective for both older and younger students, the analysis examined the impacts separately for second- and third-graders and for fourth- and fifth-graders. The second panel of Table 4.1 shows that the estimated difference between the enhanced and regular after-school program groups in total math scores for the older students is positive and statistically significant (3.9 scaled score points). For younger students, none of the impacts on test scores is statistically significant.[60] In addition, the differences between the impacts for these two subgroups are not statistically different from zero.[61]

To test whether students with different prior achievement levels benefit differently from the enhanced math program, students were divided into three subgroups according to their preintervention achievement levels: below basic, basic, and proficient. There were 467 students whose scores were "below basic"; 1,055 students scored at the "basic" level; and 380 students had "proficient" scores.[62] The bottom panel of Table 4.1 shows the results for these subgroups. The program impacts on total math scores are 2.9 scaled score points (effect size = 0.06) for the "below basic" group; 3.3 scaled score points (effect size = 0.07) for the "basic" group; and 3.0 scaled score points (effect size = 0.07) for the "proficient" group. The "basic" group's estimate

[59]Assuming that learning is equally distributed across a school year, 8.5 percent of a 9-month school year (0.085*9) is 0.765 month of additional learning.

[60]Students from different grade levels were grouped into younger and older groups (rather than examined separately by grade) to increase the power of the subgroup analysis. Sensitivity checks, though, reveal that while there is no differential program impact between the fourth- and fifth-graders, the program impact is significantly bigger for second-graders (effect size = 0.15) than for third-graders (effect size = –0.04).

[61]The p-values for the differences between these two groups are 0.347, 0.147, 0.987 for the total, problem-solving, and procedures test scores, respectively. (The p-value for this test is a statistical measure of probability that a difference between groups happened by chance. For example, a p-value of 0.01 means there is a 1 in 100 likelihood that the result occurred by chance. The lower the p-value, the more likely that the effect on the two differences is not the same.)

[62]At baseline, 59 students (28 treatments and 31controls) from the math analysis sample performed at the advanced level. The program impact on student total math scores is not significant for this group.

is statistically significant.[63] Furthermore, an F-test for the differential impacts across these three groups demonstrated that the estimated impacts for each subgroup are not statistically different from each other.[64] For both the problem-solving and the procedures subtest, students who performed at the "basic" level in the baseline test experienced positive and significant program impacts, even though the impacts are not statistically different across subgroups for these two subtests. In addition, a two-tailed t-test shows that the program impact on procedures (4.3 scaled score points) is not significantly different from the program impact on problem-solving (2.5 scaled score points).

To summarize, the enhanced after-school program produced positive and statistically significant impacts on math SAT 10 test scores for students participating in the enhanced program. Although the impacts for the higher grades are positive and statistically significant, impacts for the higher and lower grades could not be distinguished statistically. Similarly, F-tests indicate that the impacts on students coming to the program with higher or lower prior math achievement are not significantly different. The robustness of these findings was checked by using the full sample instead of the analysis sample and by using two alternative estimation models, one of which includes prior achievement and the random assignment block indicators as covariates and another that includes the random assignment block indicators as covariates. (In other words, the impact estimates are unadjusted except for the randomization strata.) These checks yield similar results to those reported here. For more details of the robustness check methods and results, see Appendix F.

In each of the 25 after-school centers using the enhanced math program, the local school district's standardized tests — which are tied to local accountability measures — are another achievement measure of policy interest. Hence, student scores on locally administered tests were collected and analyzed, and the results were compared with those from the study's test, the SAT 10. Note, first, that because the locally administered tests were not available for second-graders in 10 of the 25 schools, the sample on which this analysis was conducted is a subset of the analysis sample.[65] Second, because the locally administered tests differ by site, all test scores were standardized within each study site, and all estimated impacts on this measure are in effect size. (See Appendix E for details.) Appendix Table F.1 presents the results of this

[63] As sample size decreases, the smallest program impact that can be estimated with confidence increases — that is, the minimum detectable effect size (MDES) is larger for a smaller sample than for a bigger sample, everything else being equal. In this case, the MDES for the "below basic" group is 1.5 times as big as the MDES for the "basic" group, and the MDES for the "proficient" group is 1.7 times as big as the one for the "basic" group. For a more detailed discussion of the MDES, see Appendix B.

[64] A linear interaction model was also used to test whether the program impacts on the total score and subtests vary linearly with baseline test scores. It turns out that the linear relationship is not statistically significant for any of the three achievement outcome measures.

[65] Additionally, three schools did not have scores for third-graders.

analysis. The pattern of the first-year impacts on the locally administered math test is the same as the one shown in Table 4.1 for the analysis sample, and the impacts are not statistically significant.[66]

Impacts on Student Academic Behaviors

The expected effects of the enhanced math program on student academic behaviors are uncertain: on the one hand, if students felt better able to do their schoolwork, their classroom behavior may have improved; on the other hand, the additional formal instruction that students received in the after-school program may cause "fatigue" and, therefore, negatively affect their behavior during the regular school day. To assess this issue, three measures of student academic behavior — How often do they not complete homework? How often are they attentive in class? How often are they disruptive in class? — were examined. The measures are drawn from the survey of the sites' regular-school-day teachers and are included to see whether the enhanced after-school program changes students' behavior in any way. All three measures in this domain are on a scale ranging from 1 to 4, with "1" indicating that the specific behavior *never* occurred and "4" indicating that it occurred *often.* Table 4.2 shows that enrollment in the enhanced program did not interfere with homework completion and had no statistically significant impacts on the two classroom behavior measures for the full analysis sample or for any of the subgroups.

Variation in Impacts

While the average impact on math test scores was 2.8 scaled score points (or 0.06 standard deviation in effect size), not all 25 math centers in the study sample experienced this exact gain. The study design, which randomly assigned students within centers, enables the evaluation to explore the variation in impacts across centers, and a composite F-test does indicate statistically significant variation in impact across centers (p-value = 0.05). Figure 4.2 presents the average impact for the full analysis sample and the distribution of impacts, by center.[67]

The figure shows that 17 of the 25 center-level impact estimates (solid boxes in the figure) are above zero, and 8 of the 25 are negative. The positive estimates range from 0.1 to 18.0 scaled score points. In addition, all but one of the negative estimates (at –10.7) are between –1.2 and –4.7 scaled scores in magnitude.

[66]Note that, out of the 10 states for which state test results are available for the study sample students, two were using norm-referenced tests similar to SAT 10. The other eight states used criterion-referenced tests, which are often closely linked to specific content in the curriculum used during the regular school day. (See Appendix E for detailed descriptions of the state tests.)

[67]Center-level impacts were estimated by replacing the treatment indicator in the impact model with 25 center-level dummies (interacted with the treatment indicator).

The Evaluation of Academic Instruction in After-School Programs

Table 4.2

Impact of the Enhanced Math Program on Student Academic Behavior

Student Academic Behavior Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Full analysis sample					
Student does not complete homework	2.22	2.25	-0.03	-0.03	0.43
Student is disruptive	2.17	2.16	0.01	0.01	0.81
Student is attentive	3.33	3.31	0.02	0.02	0.61
Sample size (total = 1,961)	1,081	880			
Grade subgroups					
Grades 2 and 3					
Student does not complete homework	2.17	2.27	-0.10	-0.11	0.09
Student is disruptive	2.20	2.21	-0.01	-0.01	0.91
Student is attentive	3.31	3.33	-0.02	-0.03	0.62
Sample size (total = 971)	533	438			
Grades 4 and 5					
Student does not complete homework	2.27	2.24	0.03	0.03	0.56
Student is disruptive	2.14	2.11	0.03	0.03	0.64
Student is attentive	3.35	3.29	0.06	0.08	0.19
Sample size (total = 990)	548	442			
Prior-achievement subgroups					
Students scoring at below basic level					
Student does not complete homework	2.62	2.53	0.08	0.08	0.38
Student is disruptive	2.24	2.33	-0.10	-0.09	0.30
Student is attentive	3.06	2.97	0.09	0.12	0.25
Sample size (total = 467)	239	228			
Students scoring at basic level					
Student does not complete homework	2.19	2.30	-0.11	-0.12	0.05
Student is disruptive	2.25	2.24	0.01	0.01	0.93
Student is attentive	3.33	3.30	0.03	0.04	0.56
Sample size (total = 1,055)	612	443			
Students scoring at proficient level					
Student does not complete homework	1.92	1.90	0.03	0.03	0.79
Student is disruptive	1.93	1.83	0.10	0.10	0.35
Student is attentive	3.58	3.68	-0.10	-0.13	0.13
Sample size (total = 380)	202	178			

(continued)

Table 4.2 (continued)

SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs regular-school-day teacher survey.

NOTES: All survey responses are on a scale of 1 to 4, where 1 equals "Never" and 4 equals "Often."

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size for each outcome is calculated as a proportion of the standard deviation of the regular program group.

There are 28 enhanced program group students and 31 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

The sample sizes reported represent the number of students from the analysis sample in the given subgroup. The sample size for each outcome varies by the number of regular-school-day teachers who did not respond to the question. Across the analysis sample, the variation ranges from 8 to 18 for the enhanced program group and from 3 to 11 for the regular program group.

The next section examines to what degree the variation in impacts across centers is related to variation in the regular-school-day characteristics in which the program was operated.[68]

Linking Impact on Total Math Scores with School Characteristics

Because the effectiveness of after-school instruction may be associated with factors related to program implementation or what the students experience during the regular school day, measures of school characteristics and program implementation may help explain the variability of effects presented in Figure 4.2. Correlational analysis was conducted to shed light on such possible relationships. A multi-level hierarchical model with students nested within centers[69] was utilized to estimate the program impact, and, at the center level of the model, treatment effect was specified as a function of school characteristics as well as of program implementation measures.[70] This analysis is nonexperimentally based; thus these results should be viewed cautiously and as hypothesis-generating rather than as establishing causal inferences. Though a more complete analysis of these relationships will be done when the second year of data are collected, first-year findings allow the first step of this analysis.

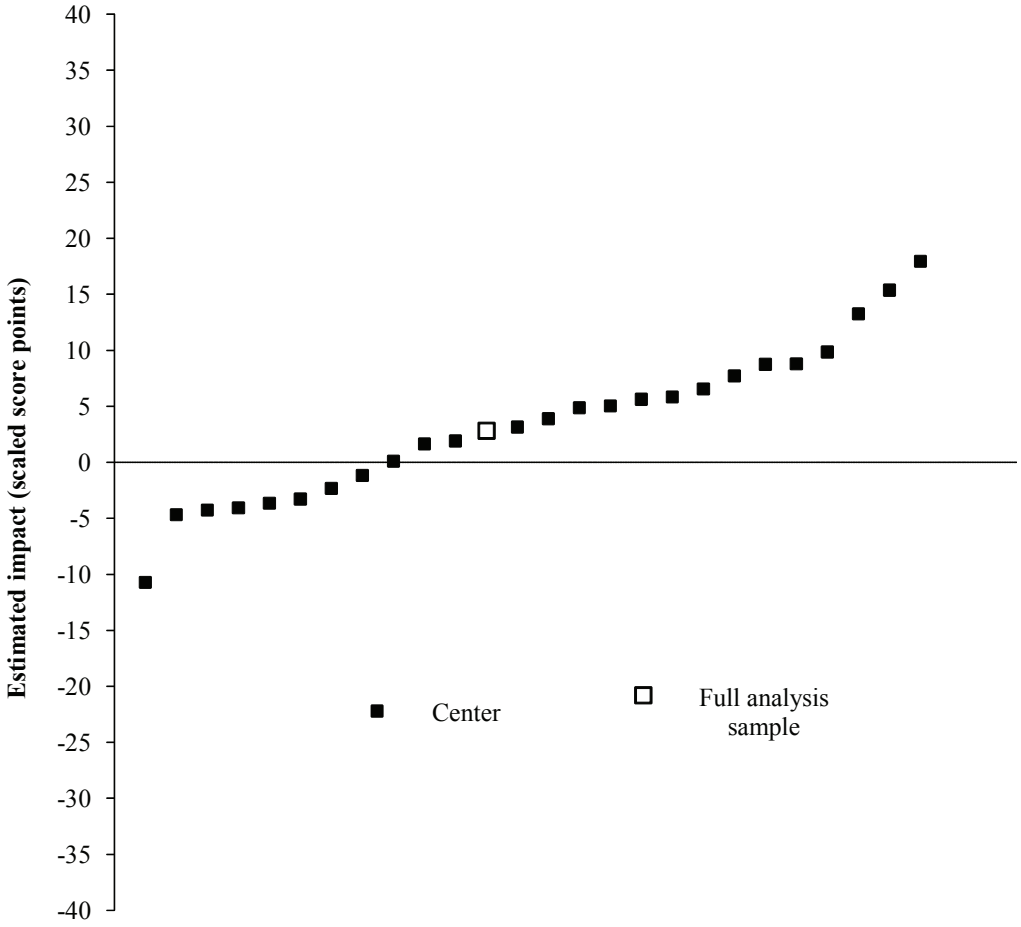
[68]Twenty-four of the 25 centers are included in this next section because one of the school characteristics could not be determined for one center.

[69]This is not a multilevel model of students nested within teachers within centers because, for the control group, information about which students were grouped with which teachers was not available.

[70]See Appendix G for details of the model.

Figure 4.2

Impact of the Enhanced Math Program on Student Achievement and Its Distribution Across Centers



SOURCES: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The figure shows the estimated program impact for the student-level analysis sample on students' SAT 10 total math scores (the white box; p-value = 0.01) and how that impact is distributed across the 25 centers in the analysis sample (each dark box). The center-by-center impacts (presented ordinally) are estimated by interacting the treatment indicator with center indicators in an ordinary least squares regression model that also controls for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. Because the study was not designed to detect the impact at the center level (on average, there are only 78 analysis sample students within each center), no statistical tests are conducted to check the significance of the impact estimate for each center. The full analysis sample comprises 1,081 enhanced program group students and 880 regular program group students.

The program implementation measures included in the model are the number of days over the course of the school year that the enhanced math program was offered and whether one or more teachers teaching the enhanced program left during the school year (which could cause a disruption in instruction). School characteristics included in the model[71] are whether the school met its Adequate Yearly Progress (AYP) goals, the proportion of students receiving free or reduced-price lunch, the in-school student-to-teacher ratio, the length of math instruction that students received during the regular school day,[72] and categories for the instructional approach of the math curriculum used during the school day.[73]

Table 4.3 shows estimates based on regression analysis of the relationships of school characteristics and implementation measures with the Mathletics program impact on the SAT 10 total math test scores. This table presents the estimates for the measures hypothesized to be associated with program impacts (that is, school-level mediators) and not estimates for all variables in the model (these are included to increase the precision of the impact estimates). Overall, the full set of school characteristics and implementation measures presented in Table 4.3 are correlated with the program impacts on total math SAT 10 score (p-value = 0.05).

School and implementation characteristics were not correlated with enhanced program effects for the overall math test scores, with two exceptions. Centers meeting adequate yearly progress were associated with higher program impacts (p-value = 0.01). Centers serving schools that employ a curriculum in Group 2 experienced lower program impacts than centers that employed a curriculum similar to Mathletics (p-value = 0.03). With the available information, it is not possible to explain the reasons for these relationships.

---

[71]School characteristic data come from the 2005-2006 National Center for Education Statistics' Common Core of Data (CCD), which compiles school-level demographic data. Data on whether a school met its AYP goals were obtained from each state's Department of Education Web site.

[72]School administrators were asked how many minutes teachers spend a day teaching math or reading to their students. The responses were not a precise number of minutes, so a continuous measure of minutes is not used. Instead, groups were created around the most common response. For math, 24 percent of schools offer 50 to 60 minutes; 32 percent offer 60 minutes; 28 percent offer 60 to 90 minutes; and the remaining 16 percent offer 90 minutes or more. Thus, the natural split for this subgroup is between schools offering 60 minutes or less of school-day math instruction and schools offering more than 60 minutes.

[73]Based on their instructional approaches, school-day curricula were categorized into three groups. Group 1 contains curricula that are unit based, which are typically longer than chapters and are investigation driven with comparatively fewer practice problems and involving interconnected subproblems (for example, Every Day Math, Move-It-Math, Real Math). Group 2 contains curricula that employ a direct instruction approach organized by lessons with spiraled curriculum (for example, Saxon). The left-out group contains curricula that have a format with math topic sections within chapters. Each section contains guided practice problems, numerous computational problems, a few application problems (word problems), and a mixed/cumulative review section at the end of each section and chapter (for example, Scott Foresman-Addison Wesley, Harcourt, McGraw-Hill, Houghton Mifflin) and is similar to the Mathletics curriculum. These are categorizations defined by the authors of this study in consultation with independent experts in math and math education. Currently in the research literature, there is no agreed upon categorization of math curricula.

The Evaluation of Academic Instruction in After-School Programs

Table 4.3

Associations Between School Characteristics and the Enhanced Math Program's Impact on Student Achievement

Interaction Characteristic	Estimated Coefficient	P-Value for the Estimated Coefficient
School		
Curriculum group 1[a]	1.43	0.83
Curriculum group 2[a]	-7.40 *	0.03
More than 60 minutes of math instruction	-1.60	0.76
Student-to-teacher ratio greater than that in the enhanced program[b]	-1.63	0.63
Did not make adequate yearly progress (AYP)	-9.30 *	0.01
Percentage of student body that is low-income[c]	0.02	0.71
Program implementation		
Enhanced teacher left the program during the school year	3.18	0.38
Total days enhanced program was offered	0.20	0.21
F-test of all interaction characteristics *		0.05
Size of student sample (total = 1,879)		
Size of school sample (total = 24)		

(continued)

SOURCES: Student achievement data are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery. Curricula and minutes of instruction were collected from research staff interviews with point persons and phone calls made to schools and districts. AYP status was collected from each state's Department of Education Web site. All other school-level characteristics were collected from the Common Core of Data Web site, http://nces.ed.gov/ccd/. Program implementation characteristics are from the Evaluation of Academic Instruction in After-School Programs attendance data and data from Bloom Associates. All data reflect the 2005-2006 school year.

NOTES: One center is not included in this analysis because it could not be categorized by type of curriculum. This occurred because the school in which it is housed employs two different curricula.

The estimated coefficients represent how the math program impact varies with each school characteristic. They were estimated using a hierarchical linear model, where in the first level (the student level) the following variables are controlled for: treatment status, indicator of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education; in the second level (the center level), the program impact is related to the school characteristic variables listed above. The F-test tested whether the coefficients on the school characteristic variables are jointly equal to zero. Within each center, the analysis sample includes, on average, 78 students.

A two-tailed t-test was applied to each estimated coefficient. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

[a]Based on their instructional approaches, school-day curricula were categorized into three groups. Group 1 contains curricula that are unit based, which are typically longer than chapters, and are investigation driven with comparatively fewer practice problems and involving interconnected subproblems (for example, Every Day Math, Move-It-Math, Real Math). Group 2 contains curricula that employ a direct instructional approach organized by lessons with spiraled curriculum (for example, Saxon). The left-out group contains

Table 4.3 (continued)

curricula that have a format with math topic sections within chapters. Each section contains guided practice problems, numerous computational problems, a few application problems (word problems), and a mixed/cumulative review section at the end of each section and chapter (for example, Scott Foresman-Addison Wesley, Harcourt, McGraw-Hill, Houghton Mifflin) and is similar to the Mathletics curriculum.

[b]Schools are classified as having a high student-to-teacher ratio if the ratio is greater than 13:1.

[c]Student body characteristics are centered on the grand mean of the school sample.

Finally, the two measures of program implementation (teacher departures and hours of Mathletics instruction offered) have no statistically significant relationship to program impacts. Teacher departures are included as a proxy for implementation continuity and strength.

As mentioned above, however, this analysis is nonexperimental because students were not randomly assigned to schools with different characteristics. Thus, the inference that a particular factor caused the impact to be larger or smaller cannot be made. Factors could exist that are correlated with both the program impact and certain school characteristics yet are not controlled for in the analysis. For example, the analysis shows that centers affiliated with schools that use a direct instruction curricular approach are associated with a lower gain from the program, but these schools could also have other characteristics — such as school instructional resources or staffing — that might be related to the effectiveness of the program but were not measured by the study team. Therefore, there might be alternative explanations for the correlations reported in Table 4.3, and the results need to be interpreted with caution.

Conclusion

Overall, the first-year implementation findings of the enhanced math program suggest that the enhanced program was implemented as designed by the developer, using most-to-all of the program materials as intended, and that there was a service contrast between the enhanced and the regular program groups — an important first condition in an evaluation of the enhanced program.

The study finds that the enhanced after-school math instruction improved students' math performance as measured by the SAT 10 test scores, by 2.8 scaled score points, or 0.06 standard deviation in effect size. Similar impacts can be found for the two subscale tests in math as well. The intervention has no statistically significant impact on student academic behaviors as measured by answers from a regular-school-day teacher survey. Correlational analysis that examined the links between program impact on total math scores and certain school characteristics found that the size of the impact did vary with school characteristics; however, the correlational results need to be interpreted with caution, as they do not indicate causality.

The study has completed its second year (school year 2006-2007) of data collection. Because the second-year sample includes students who were part of the study in the first year as well as students who were new to the study in the second year, the new wave of data will shed light both on the cumulative impact of the enhanced after-school program on returning students and on the impact of a more mature program on new students. Those results will be presented in the final report of the project.

Chapter 5

The Implementation of Enhanced After-School Reading Instruction and the Contrast with Regular After-School Services

This chapter begins with a description of the study sample for the evaluation of enhanced after-school reading instruction. It briefly discusses the characteristics of the schools that house the after-school centers and the students in the reading study sample. It then describes the design and implementation findings of the enhanced reading instruction and compares these with the services received by students randomly assigned to the regular after-school program.

The Reading Analysis Sample

Sites in the Reading Study Sample

Table 5.1 shows that, out of the 25 schools that house the after-school centers offering the enhanced reading instruction, 19 are located in large or midsize cities. Students in the schools are predominantly black (60 percent) or Hispanic (24 percent), and 81 percent of all students in these schools come from low-income families.[74] Eleven of the schools (44 percent) did not meet the Adequate Yearly Progress (AYP) goals set by their state under the federal No Child Left Behind Act in school year 2005-2006.[75]

During the regular school day, students in seven of these schools receive more than 90 minutes of reading instruction each day (see Table 5.2), with students in 18 schools receiving 90 minutes or less. As shown in Table 5.2, the school-day reading instructional approach varies, and schools may use different reading curricula across grades 2 through 5.

[74]This information comes from the 2005-2006 National Center for Education Statistics' Common Core of Data (CCD), which compiles school-level demographic data, including school locale, ethnicity, and free or reduced-price lunch status. The proportion of low-income families is defined as the proportion of students in a school who are eligible for free or reduced-price lunch. School locale designations fall into one of eight categories: large city, midsize city, urban fringe of a large city, urban fringe of a midsize city, large town, small town, rural (outside core-based statistical area), and rural (inside core-based statistical area).

[75]Data on whether a school met its AYP goals were obtained from each state's Department of Education Web site.

The Evaluation of Academic Instruction in After-School Programs

Table 5.1

Characteristics of Schools Housing After-School Centers Implementing Adventure Island

School Characteristic	
Number of schools	
School setting[a]	
Large or midsize city	19
Urban fringe of a large or midsize city	4
Large or small town	2
Rural area	0
Schools not making adequate yearly progress (AYP)	11
Composition of student body	
Race/ethnicity of students (%)	
Black	60.11
White	13.48
Hispanic	23.85
Asian	1.97
American Indian	0.44
Low-income students[b] (%)	81.35
Average student-to-teacher ratio	15:1
Sample size (total = 25)	

SOURCES: AYP status was collected from each state's Department of Education Web site. All other school-level characteristics were collected from Common Core of Data Web site, http://nces.ed.gov/ccd/. All data reflect the 2005-2006 school year.

NOTES: Composition of the student body is calculated by averaging the proportion of students within each school (collected from the CCD) across all schools.

[a]National Center for Education Statistics category designations, retrieved August 8, 2007.

[b]A student is defined as low-income if the student is eligible for free/reduced-price lunch.

Characteristics of Students in the Reading Study Sample

The process of sample intake and random assignment produced a full-study sample of 2,063 students for the reading centers (with 57 percent in the enhanced program group and 43 percent in the regular program group). Data collection included response rates for all data sources at or above the target rate of 85 percent. Two-tailed t-tests show that there are no statistically significant differences in response rates between the enhanced and the regular after-school program groups across centers for all outcome measures. (See Appendix C for response

The Evaluation of Academic Instruction in After-School Programs

Table 5.2

Characteristics of the Regular School Day in Schools Housing After-School Centers Implementing Adventure Island

Regular-School-Day Characteristic	Number of Schools
Minutes of reading instruction offered	
Number of schools offering 90 minutes or less	18
Number of schools offering more than 90 minutes	7
Reading materials/curricula[a]	
Basal Readers (Scott Foresman)	
Houghton Mifflin Reading: A Legacy of Literacy	
Open Court Reading (SRA/McGraw-Hill)	
Balanced Literacy	
Guided Reading Model	
International Baccalaureate	
McGraw-Hill	
Scholastic	
Scott Foresman	
Success For All	
Sample size (total = 25)	

SOURCES: Data were collected from research staff interviews with point persons and phone calls made to schools and districts in spring 2007.

NOTES: Data reflect grades 2 through 5 for 24 of the 25 schools housing the after-school centers; data for one school could not be obtained. School and district staff were asked for the names and publishers of the reading curricula and the amount of time spent on reading instruction in each of grades 2 through 5 during the regular school day in the 2005-2006 school year. Responses regarding curricula varied in specificity and include curricula names, such as Houghton Mifflin Reading: A Legacy of Literacy; publishers of curricula, such as McGraw-Hill; and instructional approaches, such as Balanced Literacy.

[a]The number of schools using the listed curricula is not presented because some schools use different curricula for different grades.

rate analysis.) Thus, the sample used in the analysis is limited to students with follow-up data, which is 89 percent of the entire study sample.[76] The final analysis sample used throughout this

[76]The sample used in the impact analysis is defined as students who had both a follow-up achievement test score and a teacher survey. Seventy-six students are excluded because they have a SAT 10 score but no teacher survey; 125 students are excluded because they have a teacher survey but no SAT 10 score; and 34 students are excluded because they have neither source of follow-up data.

report consists of 1,828 students for the reading centers, divided into 1,048 enhanced program students (57 percent) and 780 regular after-school program students (43 percent).[77]

Given these analysis sample sizes, the study is equipped to detect impacts as small as a 0.06 standard deviation for the full sample. This translates into 2.14 scaled score points on the Stanford Achievement Test Series, Tenth Edition (SAT 10) total reading test. The weighted average growth for students in grades 2 through 5 in a nationally representative sample is 10 scaled score points, based on the full-length SAT 10 test. Therefore, a 2.14 scaled score point impact is equivalent to 21 percent of the expected improvement of students in grades 2 through 5 nationally.[78] In addition, the minimum detectable difference in effects for a subgroup comprising half the students in the sample is 0.09 standard deviation, and the minimum detectable effect size (MDES) for a subgroup of a quarter the size of the full analysis sample is 0.12. Details on MDES calculations, given this sample size, are discussed fully in Appendix B.

Using the demographic data received from the applications, as well as the baseline test scores, Table 5.3 presents the baseline characteristics for those students assigned to the enhanced program receiving Success for All's Adventure Island and for those students assigned to the regular after-school program group. It also shows the characteristics of students in subgroups defined by grade level and by baseline reading achievement test score. The information in this table can be used to describe the reading analysis sample of students and to compare the enhanced and regular program research groups used in the impact analysis.

The reading analysis sample is made up of approximately equal numbers of students in the second through fifth grades (sample sizes: 912 for grades 2 and 3; 916 for grades 4 and 5). Like the student body in the schools linked to the after-school centers in the study, most of the sample members are black (61 percent) or Hispanic (26 percent). About half the sample members (48 percent) are male; one in four are overage for grade; and 88 percent were eligible for free or reduced-price lunch. Thirty-eight percent of the students in the reading sample lived in a household with a single adult, and 23 percent of students had a mother who did not finish high school, with 32 percent of students' mothers having a high school diploma or General Educational Development (GED) certificate. Forty percent of the sample scored at a level defined by

[77]Statistical tests were conducted to determine whether the analysis sample is different from the full sample. See Appendix C for details.

[78]Note that since the study targets low-performing students, the actual growth in the sample is different from the national average level.

The Evaluation of Academic Instruction in After-School Programs

Table 5.3

Baseline Characteristics of Students in the Reading Analysis Sample

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Full analysis sample						
Enrollment						
2nd grade	455	266	189			
3rd grade	457	258	199			
4th grade	461	256	205			
5th grade	455	268	187			
Total	1,828	1,048	780			
Race/ethnicity (%)						
Hispanic		24.14	25.80	-1.66	-0.04	0.26
Black, non-Hispanic		61.97	61.15	0.83	0.02	0.57
White, non-Hispanic		8.81	8.75	0.06	0.00	0.96
Asian		1.25	1.51	-0.27	-0.02	0.61
Other		3.83	2.79	1.04	0.07	0.21
Gender (%)						
Male		47.71	49.72	-2.01	-0.04	0.40
Average age (years)		8.73	8.67	0.06 *	0.04	0.04
Overage for grade[a] (%)		26.43	21.52	4.91 *	0.12	0.01
Free/reduced-price lunch (%)						
Eligible (among information providers)		87.58	86.29	1.29	0.04	0.37
No information provided		4.77	3.97	0.80	0.04	0.42
Average household size		1.94	1.88	0.06	0.06	0.27
Single-adult household (%)		39.21	36.34	2.87	0.06	0.21
Mother's education level (%)						
Did not finish high school		25.00	20.10	4.90 *	0.12	0.02
High school diploma or GED certificate		33.40	30.19	3.21	0.07	0.15
Some postsecondary study		37.40	44.06	-6.66 *	-0.13	0.00
No information provided		4.20	5.65	-1.46	-0.06	0.16
SAT 10 reading total scaled scores		564.80	568.37	-3.57 *	-0.09	0.01
Vocabulary/word reading[b]		555.17	560.84	-5.67 *	-0.11	0.00
Reading comprehension		566.06	569.91	-3.85 *	-0.08	0.01
Word study skills[c]		574.27	575.16	-0.88	-0.02	0.59
Sample size (total =1,828)		1,048	780			

(continued)

Table 5.3 (continued)

Characteristic	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Grade subgroups					
Grades 2 and 3					
Overage for grade[a] (%)	23.28	18.38	4.90	0.12	0.06
Mother's education level (%)					
Did not finish high school	25.95	20.72	5.23	0.13	0.07
High school diploma or GED certificate	32.25	27.77	4.48	0.10	0.14
Some postsecondary study	37.98	44.61	-6.63 *	-0.13	0.04
No information provided	3.82	6.90	-3.08 *	-0.13	0.04
SAT 10 reading total scaled scores	537.32	542.18	-4.86 *	-0.12	0.02
Vocabulary/word reading[b]	522.23	530.35	-8.12 *	-0.15	0.01
Reading comprehension	539.66	544.30	-4.63 *	-0.10	0.05
Word study skills	552.51	554.29	-1.78	-0.04	0.44
Sample size (total = 912)	524	388			
Grades 4 and 5					
Overage for grade[a] (%)	29.58	24.67	4.91	0.12	0.09
Mother's education level (%)					
Did not finish high school	24.05	19.47	4.57	0.11	0.10
High school diploma or GED certificate	34.54	32.60	1.94	0.04	0.55
Some postsecondary study	36.83	43.52	-6.68 *	-0.14	0.04
No information provided	4.58	4.41	0.17	0.01	0.90
SAT 10 reading total scaled scores	592.12	594.40	-2.28	-0.06	0.17
Vocabulary	588.04	591.26	-3.22	-0.06	0.14
Reading comprehension	592.41	595.47	-3.06	-0.07	0.14
Word study skills[c]	596.04	596.02	0.02	0.00	0.99
Sample size (total = 916)	524	392			
Prior-achievement subgroups					
Students scoring at below basic level					
Overage for grade[a] (%)	32.49	30.19	2.30	0.06	0.51
Mother's education level (%)					
Did not finish high school	27.92	23.70	4.22	0.10	0.23
High school diploma or GED certificate	35.01	31.12	3.89	0.09	0.29
Some postsecondary study	32.49	37.86	-5.37	-0.11	0.15
No information provided	4.58	7.32	-2.74	-0.11	0.13
SAT 10 reading total scaled scores	547.72	549.49	-1.77	-0.04	0.10
Vocabulary/word reading[b]	533.96	537.67	-3.71	-0.07	0.06
Reading comprehension	547.43	550.01	-2.57	-0.06	0.11
Word study skills[c]	559.77	558.40	1.36	0.03	0.52
Sample size (total = 736)	437	299			

(continued)

Table 5.3 (continued)

Characteristic	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Students scoring at basic level					
Overage for grade[a] (%)	22.95	17.95	5.00	0.12	0.08
Mother's education level (%)					
Did not finish high school	23.95	19.89	4.07	0.10	0.17
High school diploma or GED certificate	32.14	29.97	2.16	0.05	0.51
Some postsecondary study	39.52	45.12	-5.60	-0.11	0.10
No information provided	4.39	5.02	-0.63	-0.03	0.68
SAT 10 reading total scaled scores	573.25	574.83	-1.58	-0.04	0.09
Vocabulary/word reading[b]	565.77	569.72	-3.95 *	-0.07	0.04
Reading comprehension	575.11	577.17	-2.06	-0.05	0.16
Word study skills[c]	580.31	578.51	1.80	0.04	0.36
Sample size (total = 877)	501	376			
Students scoring at proficient level					
Overage for grade[a] (%)	19.42	5.00	14.42 *	0.35	0.01
Mother's education level (%)					
Did not finish high school	18.45	9.62	8.82	0.21	0.18
High school diploma or GED certificate	33.98	35.00	-1.02	-0.02	0.90
Some postsecondary study	45.63	47.98	-2.35	-0.05	0.79
No information provided	1.94	7.39	-5.45	-0.23	0.13
SAT 10 reading total scaled scores	592.04	592.68	-0.64	-0.02	0.74
Vocabulary/word reading[b]	589.67	592.92	-3.25	-0.06	0.48
Reading comprehension	596.57	598.26	-1.68	-0.04	0.64
Word study skills[c]	601.23	596.26	4.97	0.11	0.32
Sample size (total = 201)	103	98			

(continued)

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs application packet and baseline results on the Stanford Achievement Test Series, 10th ed (SAT 10) abbreviated battery.

NOTES: The estimated differences are regression-adjusted using ordinary least squares, controlling for indicators of random assignment strata. The values in the column labeled "Enhanced Program" are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in the next column are the regression-adjusted means using the observed distribution of the enhanced program group across random assignment strata as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated difference effect size for each characteristic is calculated as a proportion of the standard deviation of the regular program group.

F-tests were calculated for the full analysis sample and each subgroup sample in a regression model

Table 5.3 (continued)

containing the following variables: indicators of random assignment strata, reading total scaled score, race/ethnicity, gender, free-lunch status, overage for grade, mother's education, mobility, and family size. The full analysis sample (F-value of 1.97) is significant at the 5 percent level; the second- and third-grade sample (F-value of 1.59) and the fourth- and fifth-grade sample (F-value of 1.57) are significant at the 10 percent level. The F-values for the prior-achievement subgroups are not significant.

There are 7 enhanced program group students and 7 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]A student is defined as overage for grade at the time of random assignment if a student turned 8 before the start of the second grade, 9 before the start of the third grade, 10 before the start of the fourth grade, or 11 before the start of the fifth grade. This indicates that the student was likely to have been held back in a previous grade.

[b]Second-grade students take the word reading subtest, while third- to fifth-grade students take the vocabulary subtest.

[c]The administration of the test to fifth-graders in the spring does not include word study skills.

the baseline test publisher as "below basic" proficiency; 48 percent scored at the "basic" proficiency level; and 11 percent scored at "proficient."[79]

For the reading sample, differences between the enhanced program groups and the regular program group on most characteristics are not statistically significant, with the exceptions being the differences in the percentage overage for grade (higher for the enhanced group),[80] mother's education (lower for the enhanced group), and baseline reading test scores (also lower for the enhanced group).[81] The last characteristic is most important because it is a key outcome measure. These baseline differences are especially noticeable in the second- and third-grade sample. Randomization ensures that the enhanced program students and the regular program students start out similar to each other in terms of baseline test scores and other characteristics. However, there may still be small differences between the groups that are attributable to chance, and these can be statistically significant when the samples are very large. Looking across all the baseline variables as a group by doing an F-test, the observed differences in individual baseline characteristics are greater than would be predicted by chance (that is, there is a statistically significant difference between treatment and control groups) for the full analysis sample, for the second- and third-grade subgroup sample, and for the fourth- and fifth-grade subgroup sample. Therefore, in order to address this problem, all statistically significant baseline differences be-

[79]These percentages are calculated by dividing the sample size of the three achievement test subgroups in the table by the analysis sample. These three groups sum to 99 percent because 14 students performed at the advanced level on the baseline SAT 10.

[80]A student is defined as "overage for grade" at the time of random assignment if a student turned age 8 before the start of the second grade, age 9 before the third grade, age 10 before the fourth grade, or age 11 before the fifth grade. Thus, average age is also significantly higher for the enhanced group.

[81]The baseline test was taken before random assignment but scored approximately one month after the randomization. Thus, baseline test scores had no effect on eligibility for the program or on the random assignment process.

tween the enhanced program group and the regular program groups are controlled for by covariates in the analysis, and sensitivity tests — described briefly in Chapter 6 and more fully in Appendix F — were conducted. (See Appendix F for a detailed description of the impact analysis model and of robustness checks that validate the sample and model.)

The Implementation of Enhanced Reading Instruction

Students randomly assigned to the *enhanced after-school program* group were offered special instruction in reading during an initial 45-minute block of time, while students randomly assigned to the *regular after-school program* group received the existing academic support services in the center (usually, help with homework). Both groups received similar services for the remainder of the afternoon schedule. The enhanced reading instruction involved use of Success for All's Adventure Island, supported by implementation strategies related to staffing, support for instructional staff, and efforts to support student attendance. This section describes how these elements were put in place for the enhanced program group, the implementation challenges encountered, and the responses to these challenges.

Success for All's Adventure Island Reading Program

The Success for All Foundation (SFA) was selected to adapt its school-day reading programs to create a new after-school reading program, which is called Adventure Island and is built around the theme of a tropical island. Adventure Island is a structured reading program, with a prescribed sequence of activities in each daily, 45-minute lesson covering a number of exercises and switching from one activity to the next quickly. It includes key elements identified by the National Reading Panel (2000): phonemic awareness, phonics, fluency, vocabulary, comprehension, and strategic reading. The program builds cooperative learning into its daily classroom routines, which also include reading from a library of selected books and frequent assessments built into lessons to monitor student progress. A key component of the reading program is its assessment model, which is used to group students by their initial reading level, to identify skills in need of emphasis in instruction, and to reassess students and regroup them depending on student progress. Students' initial assignments are made based on an assessment in the fall, and students are reassessed in December and assigned, if appropriate, to a higher level in January. Adventure Island was designed to be offered four days a week for 45 minutes per day, or a total of 180 minutes a week. The enhanced instruction was planned to start up soon after the school year began and to last until the end of the after-school program in the spring.[82]

[82]The actual intensity of services is discussed below, in this chapter.

The reading program for students at the first-grade reading level — labeled *Alphie's Lagoon* — focuses on providing students with a base for literacy with a phonics program designed to build skills in phonemic awareness (the ability to hear and manipulate sounds in words), letter-sound correspondence, word-level blending (blending individual letter sounds to form words), and segmenting (breaking words into sounds). The program also has students read progressively more complex stories with guidance from the teacher, with partners, and, finally, individually. The program emphasizes the development of fluency and comprehension through the daily reading of decodable stories. Brief video segments, embedded into the daily lessons, model critical skills for the teacher and students.

For students at the second-grade reading level and above, the after-school reading program includes three levels of advancing skills (named *Captain's Cove, Discovery Bay,* and *Treasure Harbor*), each of which offers lessons based on fiction and nonfiction texts that provide instruction in vocabulary, advanced phonics, fluency, reading comprehension strategies, and story elements. Partner reading and other cooperative learning techniques are used within each lesson and are designed to build skills and motivation.

The Adventure Island reading program, like its school-day SFA counterparts, is a direct instruction approach, with detailed daily lessons for teachers to follow, SFA materials for instruction, and fast-paced activities. Teachers using this reading program are expected to master the sequence and timing of activities, allowing them to provide a daily lesson with the intended mixture of instructional strategies and topic coverage. The teacher works with the entire group of students at once, with activities during the session that involve cooperative learning (reading and discussion of material) in partnerships and teams. In Alphie's Lagoon (the first-grade level), for example, each day includes phonics instruction, with instruction by the teacher using graphical representations of letters and key sounds, picture cards, and video vignettes that teach letter-sound correspondence, word-level blending, and key vocabulary. Daily lessons also involve reading easily decodable stories and discussing the stories to support early reading skills. Teachers are expected to use SFA classroom management techniques, such as hand signals, special cheers for positive reinforcement, point allocations on a Team Score Sheet to reward students for good attendance and performance, and team and individual prizes for good work.

Use of Assessment to Guide Instruction

For the initial assessment and grouping of students, Adventure Island uses a SFA-developed 10- to 15-minute assessment (called the Word Meaning test) that can be group-administered and covers reading vocabulary, decoding, and word meaning. This test contains a list of target words, and students chose another word that means the same as the target word from a list of four words. Students scoring at the third- to fourth-grade level on the Word Meaning test are placed in Discovery Bay. For students reading below the third-grade level on the

Word Meaning test, an SFA-developed word identification test is individually administered and scored to route students to either Alphie's Lagoon or Captain's Cove. While the regular-school-day version of SFA formally reassesses students every eight weeks, the after-school program design is to reassess students once during a program year. In this project, the reassessment took place just prior to the December vacation. Students were regrouped, if needed, when they returned in January. In addition to this formal reassessment, brief fluency and comprehension assessments were built into lesson plans. In Alphie's Lagoon, phonemic awareness and phonics assessments are administered after every 10 lessons. In Captain's Cove, there are weekly written assessments for phonics, fluency, and comprehension (related to tests on stories read).

Implementation Findings

This section reports on how Adventure Island was implemented in the study centers, drawing on surveys[83] and structured interviews of after-school program staff involved in its operation, conducted by the research staff; structured protocol observations of instructional practice of after-school instructors, conducted by the research staff; structured protocol observations of implementation of Adventure Island, conducted by the district coordinators; and attendance records.

The Amount of Instruction Offered

Ninety-four percent of the after-school program staff teaching Adventure Island reported on the staff survey that they offered an average of 176 minutes of instruction per week either in four 45-minute lessons or in three 60-minute lessons. (Six percent of staff did not respond to the survey; see Appendix C for information about response rates.) The intended amount of instruction was 180 minutes. Table 5.4 provides information on the duration of the Adventure Island program. It shows the number of centers with different numbers of days of Adventure Island offered. All the reading centers offered Adventure Island classes for a minimum of 70 days during the school year, with six centers offering 70 to 79 days of instruction, another four offering 80 to 89 days of instruction, and 15 centers with more than 90 days of instruction.

Student Placement and Progression Through the Levels

In its materials for Adventure Island, SFA describes Alphie's Lagoon as "beginning reading," Captain's Cove as second-grade material, Discovery Bay as third-grade material, and Treasure Island as fourth- and fifth-grade material (Success for All 2004). Figure 5.1 shows how students in each grade were initially placed in the Adventure Island levels in the fall, based

[83]Percentages are based on the number of staff who responded to each item in the after-school staff survey.

Table 5.4

Duration of Adventure Island

Duration	Number of Centers
70 to 79 days	6
80 to 89 days	4
90 to 99 days[a]	3
100 to 109 days	9
110 to 119 days[b]	3
Sample size (total = 25)	

SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs attendance records.

NOTES: The duration of Adventure Island may have varied between classes in a center if, for example, an instructor was not present and a substitute was unavailable or due to other after-school logistics unrelated to Adventure Island. Ranges that include a center in which Adventure Island classes that met for a different number of days than the specified duration for their center are noted.

[a]In one of the centers, a class of 12 students (27.91 percent) met for 89 days.

[b]In one of the centers, a class of 11 students (28.21 percent) met for 108 days.

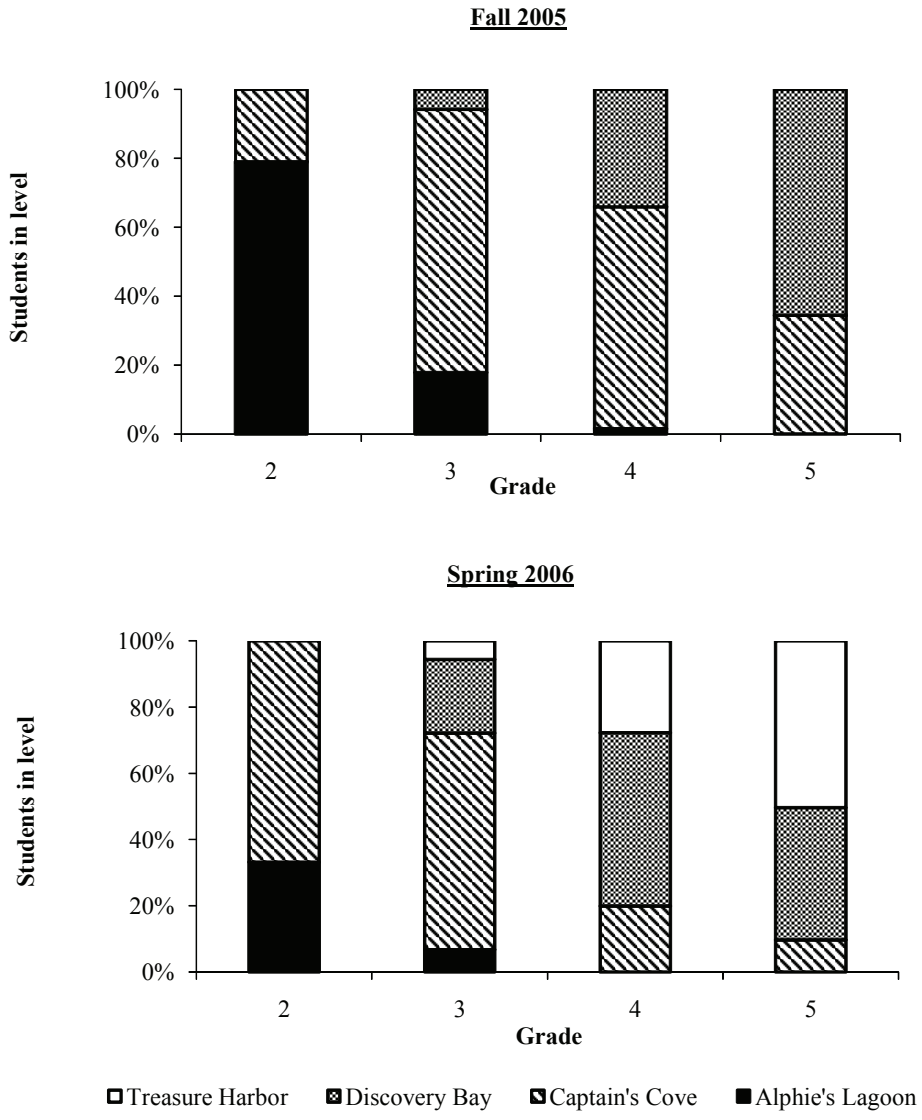
on the initial assessment, and how that changed after the December reassessment. The figure illustrates that the majority of the sample were placed in a level below their actual grade level. In the fall, 79 percent of second-graders (or 210 students) were placed as "beginning readers" in Alphie's Lagoon; 94 percent of third-graders (or 243 students) were placed below the third-grade-level Discovery Bay; and all fourth- and fifth-graders were placed below Treasure Harbor.

In January, after the midyear reassessment and regrouping of students, there was movement of students up the levels of Adventure Island.[84] Starting with the second semester, 66 percent of the second-graders (or 169 students) were placed in Captain's Cove; 22 percent of third-graders (or 56 students) were placed in Discovery Bay or Treasure Harbor; and 27 percent of fourth-graders (67 students) and 50 percent of fifth-graders (125 students) were placed in Treasure Harbor.

[84]Four percent of the fall sample were not reassessed because they were not attending the program when the assessments were administered.

Figure 5.1

The Percentage of Students in Each Adventure Island Level, by Grade Subgroup



SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs application packet and classroom information collected by Bloom Associates.

NOTES: The fall 2005 sample consists of 1,042 students in the reading analysis sample. The spring 2006 sample consists of 1,002 students. Of the 1,048 students in the reading analysis sample, 6 students did not take the SFA placement exam in September 2005, and an additional 40 students did not take the placement exam in December 2005.

Teachers' Reactions to the Content of the Program

When surveyed, 93 percent of after-school teachers (99 of 106) reported on whether the materials of Adventure Island were appropriate for their students. Ninety-eight percent of 99 staff stated that it was "very true" or "sort of true" that "the materials address the topics students need help on,"[85] while 94 percent of 99 staff reported that the materials were "about the right level of difficulty for students who are enrolled," with 4 percent reporting that they were "too easy," and 2 percent saying "too hard."

Measures of Implementation of Adventure Island

The implementation experience is summarized using structured observations conducted by both district coordinators and research staff, described in Appendix D, and structured interviews conducted by research staff.

USE OF INSTRUCTIONAL ELEMENTS

Structured classroom observations of implementation were conducted by district coordinators and were used to provide background information on the implementation of Adventure Island. The protocols used in these observations focused on six elements of the material that were identified by the developer as being key to intended implementation. They include three procedural factors (use of SFA materials, cooperative learning, awarding of points to student teams for performance) and three key topics to be covered (phonics, fluency, and completion of lesson plan).[86] Based on these observations in the two lower-level classes, 19 percent of the Adventure Island classes included four or fewer of the six core elements of the material in the sessions. In the upper-level classes, 13 percent included three or fewer of the five core elements.

Among all 344 observations that took place during the course of the school year, there were 44 classroom observations among 34 teachers in which the class that was observed included 70 percent or fewer of the core elements. During these classroom observations, three issues stand out as consistent problems, two of which deal with methods to improve fluency and one of which deals with covering all the intended lesson elements in the allotted time.[87] In 95 percent of the 44 classroom observations, teachers did not award points for fluency to teams during instruction (a method to encourage improvement in fluency), and 84 percent of the observed teachers did not model or practice fluency during the lesson. Further, in 74 percent of the 44 classroom observations, teachers did not complete all the components of the lessons in the

[85]Specifically, 64 percent reported "very true."

[86]Phonics was emphasized in Alphie's Lagoon and Captain's Cove but not in the upper levels.

[87]A finding of implementation challenges with SFA materials is consistent with prior research conducted and reported by Success for All (Slavin and Madden 1999).

allotted time. Other problems emerged less frequently; 36 percent of the 44 classroom observations involved teachers failing to award team points for cooperative learning (a strategy to encourage cooperation), 30 percent failing to use cooperative learning during the lesson, and 26 percent failing to model or practice reading comprehension strategies.

PACING OF INSTRUCTION

The Adventure Island daily lesson plans contain multiple instructional methods (such as direct instruction and cooperative learning) and specific topics like phonics. The research team observed instruction by a randomly selected half of the Adventure Island teachers and, following this observation, conducted structured interviews with them. During this interview, the teachers were asked, "Can you get through all the material you need to in each session?" All of the 50 teachers interviewed indicated experiencing some challenges related to pacing. Their responses were categorized as follows: 42 percent described pacing as a "consistent problem" and said that, as a rule, they had trouble completing the daily lesson in the allotted time. Another 32 percent said that pacing was "sometimes a challenge," depending on such things as the SFA level that they were teaching or the specific skills that they were covering. Finally, slightly over a quarter of the teachers (26 percent) reported that they were generally able to cover the material in the allotted time and that pacing was "rarely a problem" for them. Figure 5.2 summarizes staff reports of their ability to complete material within each session of Adventure Island. When teachers were asked to identify what, in particular, they found challenging, 30 percent of the 50 randomly selected teachers who were interviewed expressed concern that the pace at which they were expected to cover the material and move on to the next lesson was too fast for students to master the content. They reported feeling that the students needed more time to practice and review before moving on to a new book or skill.

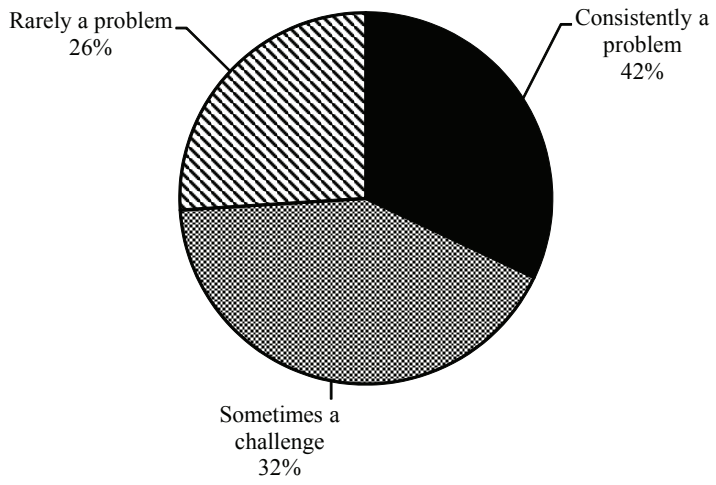
CHARACTERISTICS OF INSTRUCTION

Fifty randomly selected reading teachers were observed in their classrooms by the research team. Table 5.5 shows the results of these observations of instructional practice.[88] Eighty percent of the teachers in the reading centers were rated 3 or higher on the 4-point scale in presenting an organized sequence of instruction and use of materials; 68 percent were rated 3 or higher in making a clear presentation of reading content; and 48 percent were rated 3 or higher on modeled mastery of the content in explaining material. Seventy-four percent of instructors were rated 3 or higher on monitoring student progress during direct instruction, and 65 percent

[88]The scales used to create this measure are discussed in more detail in Appendix D.

Figure 5.2

Staff Reports of Ability to Complete Material Within Each Session of Adventure Island



SOURCE: MDRC calculations are from structured interviews with enhanced program group staff conducted by the research team.

NOTE: Percentages are based on a random sample of 50 enhanced program group staff, all of whom responded to the question "Can you get through all the material you need to in each session?"

were rated 3 or higher on monitoring student progress during independent student work,[89] which was done by breaking the whole group into small clusters so that students focused on the same material using cooperative learning techniques. Sixty-four percent of instructors were rated 3 or 4 on connecting new content to content that students already knew. Eighty-eight percent of Adventure Island instructors were rated 3 or higher on including all students in activities, which is one skill targeted by SFA classroom techniques, such as random drawings of students' names

[89]Some Adventure Island daily lessons did not allow for independent student work. In these instances, observers did not rate the teacher on this practice. For this reason, the sample size of teachers rated on this practice is 26.

Table 5.5

Ratings of Instructional and Classroom Management Practices for Sampled Staff Who Implemented Adventure Island

	Percentage of Adventure Island Staff	
Classroom Practice	Rated 3	Rated 4
Organizes instruction and use of materials in a logical sequence	42.00	38.00
Presents content clearly	30.00	38.00
Uses modeling to explain material	30.00	18.00
Monitors student progress during direct instruction	44.00	30.00
Monitors student progress during independent work[a]	34.62	30.77
Connects new content to content students already know	42.00	22.00
Includes all students in activities	40.00	48.00
Manages classroom behavior effectively	38.00	36.00
Is responsive to students	42.00	28.00
Sample size (total = 50)		

SOURCE: MDRC calculations are from observations of randomly selected enhanced program classes conducted by the research team.

NOTES: Two staff members from each center were randomly chosen to be observed; the sample reported represents 50 out of 100 staff teaching at any given time. Researchers rated enhanced program staff on a 4-point scale. As a general guide, staff received a score of 4 on a classroom practice if the practice was outstanding, 3 if it was good or very good, 2 if it could use improvement, and 1 if it definately needed improvement.

[a]Some Adventure Island daily lessons did not allow for independent student work. In these instances, observers did not rate the teacher on this practice. For this reason, the sample size of teachers rated on this practice is 26.

(called Numbered Heads) to participate in classroom activities or to answer questions. Seventy-four percent were rated 3 or higher on behavior management, and 70 percent were so rated on being responsive to students.

Implementation Strategies Used for Adventure Island

STAFFING

There are two key staffing strategies: (1) hiring certified teachers as instructors, with a preference for experienced teachers, and (2) establishing 10:1 student-to-teacher ratios for instruction. Based on responses to the after-school staff survey, 99 percent of instructors in the

reading sites were certified teachers; 75 percent had more than four years of elementary school teaching experience; and 14 percent of instructors had two or fewer years of experience.[90] In addition, centers identified a substitute teacher who was trained to teach Adventure Island.[91]

Random assignment was conducted in a way to produce enhanced program groups of 10 to 13 students per grade, to allow for some attrition and absences and still maintain an average class size of 10 students. When surveyed near the midpoint of the school year, Adventure Island instructors reported an average of nine students was enrolled in their classes per staff member. When asked, "How many students actually attend this activity on a typical day?" instructors reported that an average of approximately eight students per staff member were present.

During the school year, there were 17 new teachers and substitutes that became part of the staff complement in the reading centers. Of the 109 teachers and substitutes hired at the beginning of the school year, a total of 20 teachers and three substitutes left, but not all the teachers were replaced, because some classes were consolidated. Of the 20 teachers, three taught Alphie's Lagoon; eleven taught Captains Cove; five taught Discovery Bay; and one left prior to the program's starting and was not assigned a level.[92] In total, there were one or more staff departures in 13 of the 25 reading centers. Among these 13 centers that experienced turnover, five centers had two or more teachers leave in the fall.

Out of the 23 staff members who left, four were asked to leave because either their local manager or the project team judged that they were unable to perform their duties in a satisfactory way (three of these four were in the same center). The remaining 19 staff members cited various reasons for leaving: seven stated that their reasons were personal or said that they had an illness; 10 said that they had to leave for professional reasons, such as switching districts or starting a Ph.D. program; and two teachers went to work for a different after-school program.

SUPPORT FOR STAFF

Enhanced program group instructors received training and support in a variety of ways throughout the school year. As reported in Chapter 1, all instructors were hired in time to attend the summer training on Adventure Island prior to the start of the school year. In addition, the training on Adventure Island was repeated in January 2006 for teachers who were onboard at

[90]One of the 100 Adventure Island teachers responding to the after-school staff survey reported no prior experience teaching at the elementary school level.

[91]These substitutes are not included in the survey findings presented in this section unless they became a regular Adventure Island instructor by replacing a teacher who left prior to early spring of 2006, when the after-school staff survey was fielded.

[92]Note that, in the fall, 49 percent of students tested into Captains Cove; thus there were more teachers at that level.

that point but had not been trained in the prior summer. Ninety-seven percent of instructors responding to a staff survey question in early 2006 stated that it was "very true" or "sort of true" that they "received high-quality training to carry out this activity."[93]

Another implementation strategy was to provide all materials needed to teach Adventure Island so staff would not be burdened by purchasing supplies. Eighty-six percent of instructors reported that they had the needed materials to carry out their work, with another 14 percent reporting that this was "sort of true." The implementation plan also called for 30 minutes of paid daily preparation time, and 84 percent of instructors reported that they had 30 minutes or more of paid preparation time each day.

The project also provided ongoing, on-site technical assistance, with SFA representatives visiting each reading site twice during the school year; a project-funded, part-time district coordinator to support implementation; and frequent technical assistance (one or two on-site visits and weekly conversations by phone) with Bloom Associates. Ninety-five percent of instructors surveyed said that it was "very true" or "sort of true" that they received ongoing support for how to teach children in their activity.[94]

ATTENDANCE

Enhanced program group staff followed up with their after-school students who were absent and provided incentives for students to continue attending. The enhanced reading program was offered to students, on average, 96 days over the course of the school year. Students attended, on average, 70 days (or 73 percent of the time). Attendance of students in the enhanced program could have been influenced by the special efforts of staff to monitor absences and follow up to encourage attendance, by incentives for good attendance, as well as by Adventure Island. Because these are offered together as a package for the enhanced group, it is not possible to disentangle the influence of each factor on attendance; the factors could be offsetting or reinforcing.

To put these findings on overall attendance in context, they can be compared with the amount of attendance in the previous random assignment impact study of 21st CCLC elementary school programs commissioned by the National Center for Education Evaluation and Regional Assistance at the U.S. Department of Education's Institute of Education Sciences (IES) and conducted by Mathematica Policy Research (Dynarski et al. 2003, 2004).[95] In this earlier

[93]Seventy percent reported "very true."

[94]Seventy-five percent reported "very true."

[95]Because of differing research questions in the two studies, "attendance" was defined slightly differently. In the current study, attendance was collected on the days when the special instruction met. This means that the "total days attended" count in this study excludes attendances on days that the after-school program operated (continued)

study (mentioned in Chapter 1) that examined whether after-school programs led to improved academic achievement, students in the treatment group attended the after-school program, on average, 63 days during the course of the school year — 7 fewer days than the students in the enhanced reading program. More than half (54 percent) of students in the enhanced reading program attended more than 75 days, and 23 percent attended 51 to 75 days over the course of the school year, while just 40 percent and 15 percent of the students in the earlier study attended that often, respectively. Fifteen percent of students in the enhanced reading program attended for 25 days or fewer, compared with 26 percent of students in the earlier study.[96]

Challenges in Implementing the Adventure Island Reading Program

Next to pacing, the most frequent concern expressed by Adventure Island reading teachers was about preparation time. Depending on the center, Adventure Island reading teachers received between 30 and 60 minutes of paid preparation time each day that they taught Adventure Island after school.

As part of the structured interviews following the classroom observation of half the instructors (randomly selected to be interviewed), the teachers were asked open-ended questions to identify what challenges they encountered implementing Adventure Island and how the program might be improved. The most common challenge specified in these open-ended questions was pacing (8 teachers), followed by the amount of paperwork (6 teachers). However, when asked specifically about their preparation time, 32 percent of teachers (16 of the 50) volunteered that they did not feel the preparation time allotted was sufficient. Five of these teachers cited the amount of paperwork required, time needed to prepare materials, or the time needed to read through the stories and master the "script" that they were expected to use for instruction. However, more than half (9 teachers) made specific comments that indicate that they were not able to use the allocated preparation time in the intended way. They reported that they felt too tired at the end of the day to prepare for the next day or that their preparation time was interrupted by other responsibilities (for example, parent calls and bus duty at the end of the after-school program).

With this background on the implementation of the models of enhanced instruction, the next section shifts to a brief summary of the services offered students who were assigned to the regular after-school program and discusses the differences in services actually received by students in the two groups.

---

but the special instruction was not offered. The Mathematica report collected attendance data for all days that the after-school program operated. This difference in definition means that the difference in attendance in the two studies is somewhat underestimated.

[96]See Dynarski et al. (2004, p. 14).

The Difference in After-School Academic Services Received by the Enhanced Program Group and the Regular Program Group

Any program impacts, which are reported in the next section of this chapter, are produced by the *difference* between the after-school academic services received by the enhanced program group and those received by the regular, "business as usual" program group. This section first describes the academic support services offered to and received by the regular program group and compares these services with those received by the enhanced program group. The section then focuses on differences in attendance in these services between the two groups and concludes with analysis of differences in special academic support received from other sources — during the regular school day and outside school.

The service contrast for which impacts are estimated is described through five interrelated findings. First, the *service offerings* differ: 15 percent of the regular after-school group staff received some form of academic instruction in reading. Overall for the regular program group, homework help and/or tutoring on multiple subjects were the most common academic support offered. Second, *staff members* providing the instruction to the enhanced group students were also more likely to be experienced certified teachers and received more training and support for their instruction than staff for the regular program group. Third, *overall attendance* in the after-school program was greater for students in the enhanced program group. Fourth, students in the enhanced program group received more hours of *academic instruction in reading,* with the average service difference being 48 hours or about 20 percent more total reading instruction over the course of the school year. Finally, *academic support from other sources* (during the regular school day or other out-of-school activities) did not lessen the service contrast produced in the after-school program.

Differences in Service Offerings

The academic support offered to students in the regular program group was different from the support for students in the enhanced program group, in various ways, including the nature of the services offered and the staffing strategy, support provided to the staff, and attendance policies. Because after-school centers that provided formal reading instruction in their regular after-school program were not selected as sites for the evaluation, the regular or "business as usual" programs described in this chapter are not necessarily indicative of the state of after-school programming in the United States in general but, rather, are a reflection of what comparison group members received in this study.

The previously mentioned survey of after-school staff covered both staff providing the enhanced reading instruction and staff providing academically oriented services to students in

the regular after-school program. The findings for the regular after-school program group in this section are based on the latter staff's responses to the survey.[97]

Academic Support Services

Regular after-school program staff were surveyed about the nature of the services offered in the regular after-school program. In the reading sites, the majority of regular after-school program staff (62 percent) reported focusing on mixed subjects, depending on student needs, by providing help with homework or individual or small-group tutoring. Eighteen percent focused on a single subject other than reading (for example, math or science), and 15 percent focused on reading.

The regular after-school program staff members focusing on reading were potentially offering services similar to the special reading instruction provided to students in the enhanced program group. Figure 5.3 presents more detailed information about the services provided by these regular program staff. The figure begins at the top with all regular program staff providing academically oriented services, and 15 percent of regular program staff reported a main focus on reading, with 12 percent of the whole regular program staff (or eight individuals) reporting that they provided academic instruction in reading (as opposed to tutoring, homework help, or a response of some other method of support). Of the eight individuals, five staff reported formally assessing student progress monthly, all of whom also reported using the student assessment to guide the instruction. In addition, three reported that their instruction used a daily lesson plan with supporting materials. Of the eight staff members reporting that they provided academic instruction in reading, one person uses teacher-made material; one relies on suggestions from the school-day teachers; two have children do guided reading; and the remaining four reported that they use some unnamed reading program.

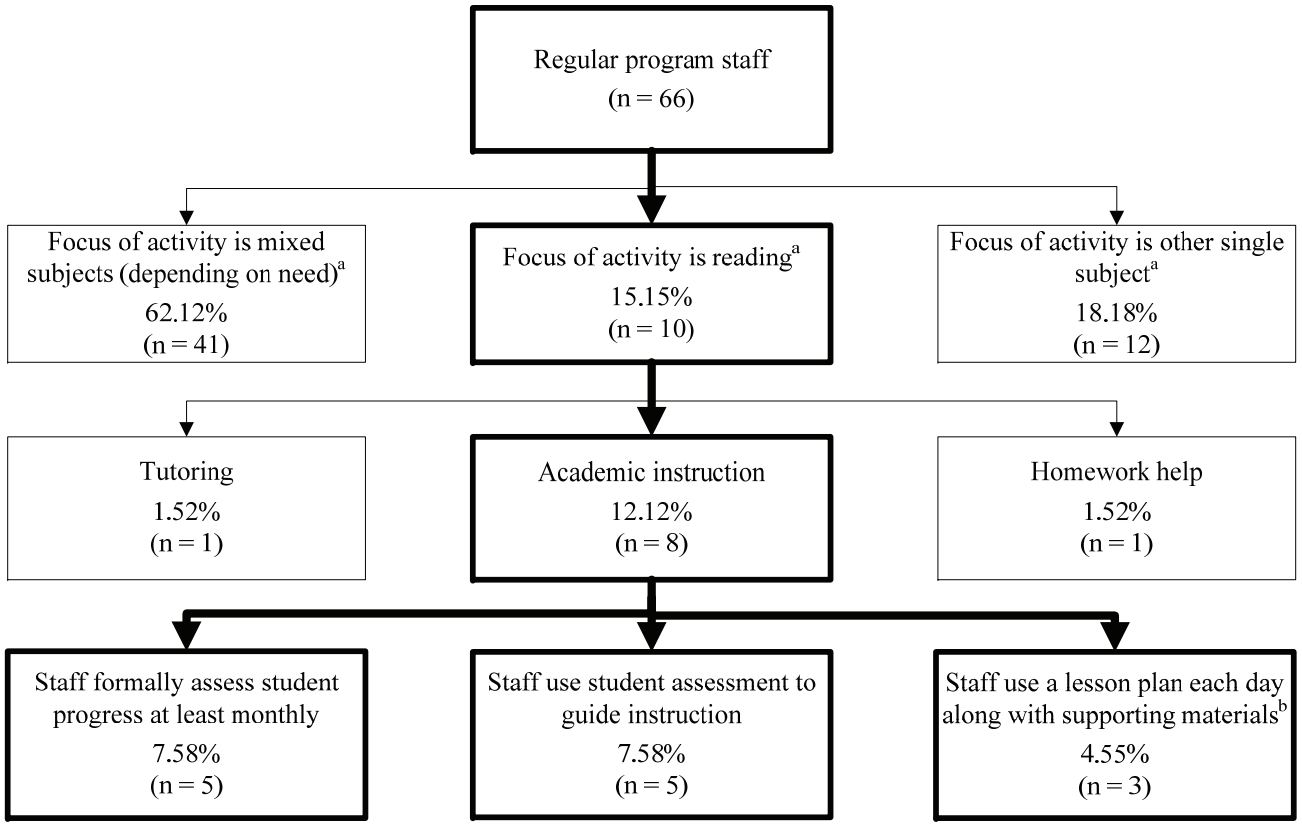
Staff Providing Academic Support Services

In the regular after-school program, certain staff members were involved in providing academic support to students, while other staff members were primarily involved in enrichment or recreational activities. This and the following sections focus on the staff members providing academic support within the after-school program who were surveyed by the research team. As shown in the top panel of Table 5.6, 60 percent of regular program staff were certified teachers; 13 percent had no prior elementary school teaching experience; and 55 percent had more than four years of elementary teaching experience. Among enhanced program staff, 99 percent were certified teachers — a statistically significant difference of 39 percentage points above the percentage for the regular program staff. Seventy-five percent of the enhanced program staff had

[97] Percentages are based on the number of staff who responded to each survey item.

Figure 5.3

Academic Services Offered by Regular "Business as Usual" After-School Program Staff at Centers Implementing Adventure Island



SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs after-school staff survey.

NOTES: Percentages are based on the regular program group sample, comprising 66 staff. Staff for whom values are missing are included in the calculations.

[a]Three of the 66 staff did not respond to this question.

[b]Staff responded "sort of true" or "very true" to the question "I have a lesson plan to follow each day, along with supporting materials."

The Evaluation of Academic Instruction in After-School Programs

Table 5.6

Characteristics of After-School Staff and Support for Staff at Centers Implementing Adventure Island

Service Offering	Enhanced Program	Regular Program	Estimated Difference		P-Value for the Estimated Difference
Staffing strategy					
Certified in elementary education (%)	99.00	59.68	39.32	*	0.00
Years of elementary school teaching experience (%)					
No experience	1.00	12.90	-11.90		
1-2 years	13.00	24.19	-11.19		
3-4 years	11.00	8.06	2.94		
More than 4 years	75.00	54.84	20.16		
			chi-square	*	0.00
Staff-youth ratio (youth enrolled)	1:9	1:14	-4.64	*	0.00
Sample size (total = 165)	100	65			
Support for staff					
High-quality training to carry out activity[b] (%)	97.00	58.33	38.67	*	0.00
Ongoing support from district for how to teach children in activity[b] (%)	95.00	55.00	40.00	*	0.00
Amount of paid preparation time to carry out activity (%)					
None	1.01	36.67	-35.66		
Less than 15 minutes per day	0.00	3.33	-3.33		
15 minutes to less than 30 minutes per day	15.15	25.00	-9.85		
30 or more minutes per day	83.84	35.00	48.84		
			chi-square[a]	*	0.00
Sample size (total = 165)	100	65			

SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs after-school staff survey.

NOTES: All findings are based on staff self-reports. The values reported for the enhanced program group and the regular program group are the unadjusted means for the staff in each group. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. For service offerings where the table presents the distributions across more than two responses, chi-square tests were used to test whether the distributions for the enhanced program group and the regular program group were the same. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The sample size reported represents the number of staff who filled out a survey. The sample size for each service offering varies by as much as 6 for the enhanced program group and 5 for the regular program group due to nonresponses on particular survey items. Staff for whom values are missing are not included in the calculations.

[a]This chi-square test may not be valid due to small sample sizes within the cross-tabulation.

[b]This presents percentages of after-school staff who responded "sort of true" or "very true" when surveyed.

four or more years of elementary school teaching experience; this is a 20 percentage point difference above the percentage for the regular program staff, and a chi-square test reveals that the difference in the distribution of experiences between the two groups of staff is statistically significant.

The enhanced program averaged a student-to-staff ratio of 9:1, while the student-to-staff ratio in the regular after-school program averaged 14:1 — a statistically significant difference of five students per staff member.

The difference in staffing between the enhanced and the regular program groups, which occurred coincident with the implementation of Adventure Island, could contribute to program impacts. However, the effect of having more certified experienced teachers and a lower student-to-teacher ratio after school cannot be disentangled from the effect of the implementation of Adventure Island.

Support for Staff

As the lower panel of Table 5.6 shows, staff providing academic support in the regular after-school programs were less likely than staff for the enhanced programs to report having received high-quality training to carry out their work (a statistically significant difference of 39 percentage points) or to report receiving ongoing support for how to teach children in their activity (a statistically significant difference of 40 percentage points). In addition, they were less likely to report receiving paid daily preparation time. In the reading sites, 37 percent reported getting no paid preparation time, and 35 percent reported getting 30 minutes or more, compared with 84 percent for the enhanced reading program staff — for a difference of 49 percentage points. A chi-square test found that the differences in the paid preparation time are statistically significant.

Differences in Attendance in the After-School Program

As mentioned above, the centers that were selected for the project all expected enrolled students to attend regularly, and none operated as a drop-in program with sporadic attendance. All regular programs took daily attendance (as required for 21st CCLC grantees), but no special staff were assigned to follow up with regular after-school program students who were absent (as the district coordinators did for the enhanced program group).

The first panel in Table 5.7 presents attendance on the days that the enhanced program operated. The first row of data shows the number of days attended, and the second row reports average hours of attendance in reading instruction offered by the after-school program for students in the enhanced and regular program groups. The discussion presents findings for the entire analysis sample and then for subgroups based on grade level.

The Evaluation of Academic Instruction in After-School Programs

Table 5.7

Attendance of Students in the Reading Analysis Sample

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Attendance in after-school program[a]					
Number of days attended	70.34	63.68	6.66 *	0.19	0.00
Total hours of reading instruction received[b]	55.00	6.54	48.46 *	2.74	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	38.67	31.33	7.34 *	0.16	0.00
Number of days per week[d]	1.13	0.78	0.35 *	0.24	0.00
Regular school day[e]					
Students receiving special support (%)	2.41	2.39	0.02	0.03	0.41
Minutes per week of individualized help	86.81	83.66	3.15	0.02	0.52
Sample size (total = 1,828)	1,048	780			

(continued)

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs attendance records, student survey responses, and regular-school-day teacher survey responses.

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size for each measure is calculated as a proportion of the standard deviation of the regular program group.

[a]Attendance in the after-school program is based on the days the enhanced program operated.

[b]Students in the enhanced classes received 45 minutes of instruction (and 60 minutes in one site that met only three days a week) on the days they were present. Total hours is calculated for these students by multiplying each student's total days of attendance by 45 (or 60 in the one site).

Students in the regular program group were not supposed to receive any structured instruction. However, some regular program staff indicated on the survey that they provide structured academic instruction. Total hours is calculated for these students by multiplying the total number of days attended by 45, then by the proportion of regular program staff within the center who reported providing structured instruction. If no regular program staff in a center indicated that they provide structured instruction, then total hours for these students in that center is zero. If no regular program staff in a center answered this question, this calculation could not be performed for these students. Calculated as such, the sample size for the regular program group is 603.

Table 5.7 (continued)

[c]This information comes from student survey responses to questions for each day of the week that ask, "Do you go somewhere else for a reading class or to be tutored in reading?" These calculations are based on a smaller sample than the reported analysis sample by 11 students in the enhanced program group and 8 students in the regular program because these students did not complete a survey.

[d]Students who responded that they do not receive reading support from other out-of-school sources are included in these averages.

[e]This information comes from regular-school-day teacher survey responses. "Special support" refers to special support in reading during the school day (that is, pull-out tutoring, Reading Recovery, assigned to a computer assisted lab, and so on). "Individualized help" refers to individual help from the teacher or an aide with a task or answering a question. Teachers who responded that they did not provide support may or may not have responded that they provided minutes of individualized help. Thus, average minutes includes responses for all students, not just those who received special support.

Attendance in the After-School Program When Adventure Island Operated

STUDENTS IN THE ENHANCED PROGRAM GROUP ATTENDED THE AFTER-SCHOOL PROGRAM MORE THAN STUDENTS IN THE REGULAR PROGRAM GROUP ON DAYS WHEN ADVENTURE ISLAND OPERATED

Students in the enhanced program attended 7 more days over the school year than those in the regular program, a statistically significant difference. Thus, Adventure Island as implemented in the study did not lead to a decline in attendance. Again, attendance of students in the enhanced program could have been influenced by the special staff monitoring and follow-up of absences, by incentives for good attendance, and by the special reading instruction. Because these are offered together as a package for the enhanced program group, it is not possible to disentangle the influence of each factor on attendance; the factors could be offsetting or reinforcing.

For subgroups based on student grade level and baseline achievement, the same pattern of somewhat greater attendance among the enhanced program group is present. Findings for subgroups are presented in Appendix H. The difference in days attended is statistically significant for the two grade-level subgroups and for one of the prior-achievement subgroups (students scoring at the “basic” level at baseline).

Amount of Academic Instruction Received in the After-School Program

STUDENTS IN THE ENHANCED PROGRAM GROUP ATTENDED MORE HOURS OF AFTER-SCHOOL ACADEMIC INSTRUCTION IN READING THAN THOSE IN THE REGULAR PROGRAM GROUP

The lower average hours for students in the regular program group reflects the finding that students in the enhanced program group had better attendance, and 12 percent of regular program staff reported providing academic instruction in reading, whereas all staff in the enhanced program group provided reading instruction.

In the reading sites, students in the enhanced program group averaged 48 more hours of reading instruction than the regular program group, a statistically significant difference. This impact on reading instruction is an estimated 20 percent more reading instruction when taking into account regular-school-day reading instruction. This percentage increase is estimated based upon information on the minutes of school-day reading instruction reported above in this chapter. More specifically, if students receive 90 minutes per day of instruction (as is common for reading) and attend 90 percent of 180 scheduled school days, then they would receive 243 hours of instruction in reading. The 48 hours of extra reading academic instruction is 20 percent more instructional time.

Academic Support in Reading from Other Sources

Surveys of students and regular-school-day teachers provide information on additional sources of reading instruction that students might receive outside after-school programs. The bottom panel of Table 5.7 contains the findings for academic support from other nonschool sources and during the regular school day.

Support Outside School

STUDENTS IN THE ENHANCED PROGRAM GROUP PARTICIPATED IN MORE CLASSES OR ACTIVITIES IN READING OUTSIDE SCHOOL THAN STUDENTS ASSIGNED TO THE REGULAR AFTER-SCHOOL PROGRAM GROUP

Students were surveyed in late fall 2005 and spring 2006 about whether they attended a reading class or activity outside the regular school day that was not part of the after-school program. (The students were not asked to provide details about the class or activity.) They were also asked how many days a week they attended this class or activity. Results from the spring survey are presented in Table 5.7.

A higher percentage of the enhanced program group reported participating in outside reading classes or activities. Thirty-nine percent of the enhanced group reported participating in an outside reading class or activity, compared with 31 percent of the regular program group, and enhanced program students reported participating about one-third of a day more per week (0.35 day) in these activities than regular program students, with both differences being statistically significant.[98]

Support During the Regular School Day

THERE ARE NO STATISTICAL DIFFERENCES IN THE AMOUNTS OF ACADEMIC SUPPORT DURING THE REGULAR SCHOOL DAY BETWEEN STUDENTS IN THE ENHANCED AND THE REGULAR PROGRAM GROUPS

Enhanced program group students received 87 minutes of individualized reading instruction per week (an average of 17 minutes a day), compared with 84 minutes for students in the regular program group, but the difference in minutes is not statistically significant. Finally, there are no statistically significant differences in the percentages of students in the enhanced and the regular program groups who received special instruction in reading.

[98]Findings for the grade-level and prior-achievement subgroups are reported in Appendix H.

Chapter 6

The Impact of Enhanced After-School Reading Instruction

This chapter presents the impact findings for the reading analysis sample, focusing on answering the primary research question: "What are the impacts of the enhanced after-school reading instruction (Adventure Island) on student achievement?" In addition, secondary program effects on certain student academic behaviors — such as homework completion, attentiveness, and disruptiveness in class — are also analyzed. This is followed by the findings from the exploratory analysis on the association between the reading program impacts and the characteristics of the school.

Program Impacts on Student Academic Achievements and Behaviors

Impacts on Student Academic Achievement

In the spring of the first program year, the Stanford Achievement Test, Tenth Edition (SAT 10), abbreviated battery reading test was fielded for all students, and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency and nonsense word fluency tests were administered to students in the second and third grades. The SAT 10 provides a total reading score and subscales on vocabulary, reading comprehension, and (for grades 2 through 4) word study skills.

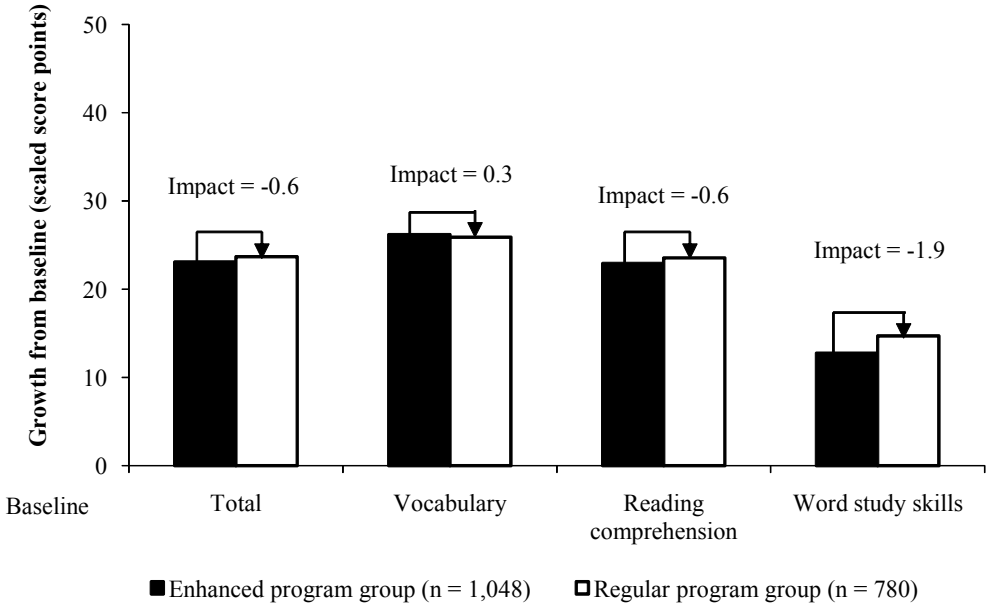
The leftmost pair of bars in Figure 6.1 demonstrates the enhanced program impact for the whole analysis sample. The average total reading score of the enhanced program group increased over the school year by 22.6 scaled score points from baseline, as indicated by the darker bar in the graph. Had there been no intervention, the same group of student would have had a growth of 23.2 scaled score points over the school year (the lighter bar in the graph).[99] The estimated impact for the enhanced program, which is not statistically distinguishable from zero, is –0.6 scaled score point, and the effect size is –0.02 standard deviation.

[99]The reading study sample experienced greater growth in SAT 10 total reading scores from the fall to the spring of the first program year (23 scaled score points) than a nationally representative sample. (The weighted average growth of students in grades 2 through 5 in that sample was approximately 10 scaled score points.) This could be related to the fact that 88 percent of the students in the reading analysis sample were performing "below proficient" in reading at the beginning of the program, indicating that the study sample includes more such students than the national sample.

The Evaluation of Academic Instruction in After-School Programs

Figure 6.1

Student Growth on Test Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Reading Program



SOURCES: MDRC calculations are from baseline and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The estimated impacts on follow-up results are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. Each dark bar illustrates the difference between the baseline and follow-up SAT 10 scaled scores for the enhanced program group, which is the actual growth of the enhanced group. Each light bar illustrates the difference between the baseline SAT 10 scaled score for the enhanced program group and the follow-up scaled score for the regular program group (calculated as the follow-up scaled score for the enhanced group minus the estimated impact). This represents the counterfactual growth of students in the enhanced group.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

Spring administration of the SAT 10 to fifth-graders does not include word study skills. Thus, the sample of students reporting follow-up scores on the word study skills subtest differs from the sample with baseline scores as well as from the sample with follow-up scores on the vocabulary and reading comprehension subtests, which do include fifth-graders.

The rest of the figure shows that impacts on students' performances in the three subcomponents of the SAT 10 — vocabulary, reading comprehension, and word study skills — are not statistically significant. The top panel of Table 6.1 provides detailed information about these estimates.

To determine whether the program was effective for older or younger students, impacts were estimated separately for the second- and third-graders and for the fourth- and fifth-graders. The second panel of Table 6.1 shows that the estimated difference between the enhanced and regular after-school program groups for both the younger and the older subgroups are *not* significantly different from zero after the first year of program implementation. Moreover, a two-tailed t-test shows that program impacts for the two subgroups do not differ statistically from each other. This is also the pattern for impacts on the vocabulary and reading comprehension subtests.[100]

On the other hand, analysis shows that the enhanced reading program produced positive impacts on one of two measures of fluency for the younger students in the study sample.[101] The estimated difference between the enhanced and the regular after-school program groups is 3.7 points (effect size = 0.12) in the nonsense word fluency subtest of DIBELS, which targets the alphabetic principle (including letter-sound correspondence and the ability to blend letters into words in which letters represent their most common sounds). This estimated effect of the program is statistically significant. However, after accounting for multiple comparisons, the estimate is no longer statistically significant.[102]

To determine whether the program was effective for students with different prior achievement levels, students were divided into three subgroups according to their preintervention reading achievement levels: below basic, basic, and proficient. The bottom panel of Table 6.1 presents the separate impact estimates for students from these three subgroups.[103] The program impacts on total reading scores or on any of the three subtests are not significantly differ-

[100]Two-tailed t-tests were also conducted to see whether program impacts differ by grade level within each subgroup, and no statistically significant differences were found.

[101]These two tests on fluency were administered to second- and third-grade students in the first program year. In the second program year, they were administered to students in all four grades.

[102]The DIBELS nonsense word fluency subtest is one of six reading measures estimated for second- and third-grade students. When accounting for multiple test corrections using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), this estimate is no longer statistically significant.

[103]At baseline, 14 students (7 treatments and 7 controls) from the reading analysis sample performed at the advanced level. The program impact on student total reading scores could not be estimated for this group because of the small sample size.

The Evaluation of Academic Instruction in After-School Programs

Table 6.1

Impact of the Enhanced Reading Program on Student Achievement

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Full analysis sample					
SAT 10 reading total scaled scores	587.42	588.04	-0.62	-0.02	0.51
Vocabulary	580.94	580.63	0.31	0.01	0.82
Reading comprehension	588.72	589.36	-0.64	-0.02	0.59
Word study skills (grades 2-4)[a]	586.39	588.33	-1.94	-0.05	0.24
Sample size (total = 1,828)	1,048	780			
Grade subgroups					
Grades 2 and 3					
DIBELS					
Oral fluency score	70.54	68.27	2.26	0.07	0.12
Nonsense word fluency score	64.53	60.82	3.72 *	0.12	0.03
Sample size (total = 931)	537	394			
SAT 10 reading total scaled scores	569.42	570.41	-0.99	-0.03	0.46
Vocabulary	557.05	557.56	-0.51	-0.01	0.80
Reading comprehension	571.54	571.77	-0.23	-0.01	0.90
Word study skills	579.28	582.86	-3.58	-0.09	0.06
Sample size (total = 912)	524	388			
Grades 4 and 5					
SAT 10 reading total scaled scores	605.43	605.57	-0.15	0.00	0.91
Vocabulary	604.84	603.54	1.29	0.03	0.47
Reading comprehension	605.89	606.81	-0.92	-0.02	0.57
Sample size (total = 916)	524	392			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 reading total scaled scores	577.48	575.60	1.88	0.05	0.19
Vocabulary	568.88	566.55	2.33	0.05	0.27
Reading comprehension	579.82	577.51	2.32	0.06	0.21
Word study skills[a]	572.06	572.02	0.05	0.00	0.99
Sample size (total = 736)	437	299			
Students scoring at basic level					
SAT 10 reading total scaled scores	591.61	593.22	-1.62	-0.05	0.25
Vocabulary	585.88	587.20	-1.32	-0.03	0.52
Reading comprehension	592.00	593.40	-1.41	-0.04	0.43
Word study skills[a]	591.06	595.05	-3.99	-0.10	0.10
Sample size (total = 877)	501	376			

(continued)

Table 6.1 (continued)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Students scoring at proficient level					
SAT 10 reading total scaled scores	606.71	611.03	-4.32	-0.12	0.27
Vocabulary	604.77	605.96	-1.19	-0.03	0.84
Reading comprehension	607.70	613.85	-6.15	-0.16	0.25
Word study skills[a]	610.92	609.20	1.72	0.04	0.78
Sample size (total = 201)	103	98			

SOURCES: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery, and results on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessments.

NOTES: Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: for the full analysis sample, scores range from 374 to 787, 439 to 777, 412 to 739, and 410 to 740; for the second- and third-grade subgroup, scores range from 374 to 765, 439 to 743, 412 to 700, and 410 to 727; and for the fourth- and fifth-grade subgroup, scores range from 434 to 787, 478 to 777, and 484 to 739. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximim score; the maximum score is determined by the number of words a student can read or identify correctly in one minute.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 reading total scaled score is calculated as a proportion of the standard deviation of the regular program group, which is 35.71 based on the analysis sample. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.05. For each SAT 10 and DIBELS subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group.

There are 7 enhanced program group students and 7 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]The sample consists of second- through fourth-graders only because the spring administration of the test to fifth-graders does not include word study skills.

ent from zero for any of the subgroups. In addition, a joint F-test shows that the program impacts are the same across the three subgroups.[104]

Overall, the students in the enhanced reading program group did not experience any statistically significant impact in their performance level on reading tests (SAT 10 total and subtests), above and beyond the level that they would have achieved had there been no enhanced reading program during the first program year. This can be said about both the whole analysis sample and subgroups. Students in grades 2 and 3 may have benefited from the program on their fluency skills, as measured by the nonsense word fluency DIBELS subtest, but this lone significant result could be due to chance, given the multiple number of comparisons performed.

As mentioned in Chapter 5, there are significant differences in baseline test scores between the enhanced and the regular after-school program groups in the reading sample, with those in the enhanced program group having lower scores. As a robustness check, block-by-block baseline differences in test scores were checked, and 12 blocks with the biggest baseline test score differences were excluded from the sample, thereby eliminating the statistically significant differences at baseline.[105] All impacts were reestimated using this restricted sample, and the results are similar to the impact estimates for the analysis sample. (See Appendix F for details.) These results suggest that controlling for the baseline characteristics as covariates in the impact model sufficiently eliminated the observed baseline differences between the enhanced program and the regular program groups. Thus the reading impact results presented above are not affected by the significant baseline differences. Additional robustness tests were also conducted, using the full sample instead of the analysis sample, and using two alternative estimation models, one of which includes prior achievement and the random assignment block indicators as covariates and another that includes the random assignment block indicators as covariates. (In other words, the impact estimates are unadjusted except for the randomization strata.) These tests yield results that are consistent with the ones reported here.[106]

Student scores on locally administered reading tests were also collected and analyzed, and the results were compared with results from the study's test, the SAT 10.[107] Appendix F

[104]The p-value for this test is 0.12. A linear interaction model was also used to test whether the program impacts on the total score and subtests vary linearly with students' prior achievement level. This also indicates that the program impact on the total reading test score does not vary linearly with students' prior achievement (p-value = 0.14).

[105]As described in Chapter 2, random assignment was conducted within each center by grade block.

[106]For detailed descriptions of the tests, see Appendix F.

[107]Note, first, that because the locally administered tests were not available for second-graders in 13 of the 25 schools, the sample on which this analysis was conducted is a subset of the analysis sample. Second, because the locally administered tests differ by site, all test scores were standardized within each study site, and all estimated impacts on this measure are in effect size. For detailed discussion of the sample and the test score standardization, see Appendix E.

compares the impact estimates for local and SAT 10 tests; again, there is no statistically significant program impact on student performance in the locally administered reading tests after the first year of program administration.

Impacts on Student Academic Behaviors

The expected effects of the enhanced reading program on student academic behaviors are uncertain: on the one hand, if students felt better able to do their schoolwork, their classroom behavior may have improved; on the other hand, the additional instructions that students received in the after-school program may cause "fatigue" and, therefore, negatively affect their behavior during the regular school day. To assess this issue, three measures of student academic behavior — How often do they not complete homework? How often are they attentive in class? How often are they disruptive in class? — are drawn from the survey of the sites' regular-school-day teachers. All three measures in this domain are on a scale ranging from 1 to 4, with "1" indicating that the specific behavior *never* occurred and "4" indicating that it occurred *often.* Table 6.2 shows estimated impacts on these measures. The enhanced reading program did not produce statistically significant impacts on any of these measures either for the full analysis sample or for the various subgroups.

Variation in Impacts

Figure 6.2 presents the average impact for the full analysis sample and the distribution of impacts, by center.[108] Eleven of the 25 center-level impact estimates (solid boxes in the figure) are above zero, and 14 of the 25 are negative. The positive estimates range from 0.2 to 9.7 scaled score points, and the negative estimates are between –0.7 and –11.3 scaled scores in magnitude. A composite F-test indicates that the cross-center variation is not statistically significant and cannot be distinguished reliably from the overall mean (p-value = 0.17).

The next section examines to what degree the variation in impacts across centers is related to variation in the regular-school-day characteristics in which the program was operated.

[108]Center-level impacts were estimated by replacing the treatment indicator in the impact model with 25 center-level dummies (interacted with the treatment indicator).

The Evaluation of Academic Instruction in After-School Programs

Table 6.2

Impact of the Enhanced Reading Program on Student Academic Behavior

Student Academic Behavior Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Full analysis sample					
Student does not complete homework	2.42	2.41	0.01	0.01	0.85
Student is disruptive	2.32	2.28	0.04	0.04	0.43
Student is attentive	3.30	3.33	-0.03	-0.04	0.35
Sample size (total = 1,828)	1,048	780			
Grade subgroups					
Grades 2 and 3					
Student does not complete homework	2.39	2.39	0.01	0.01	0.90
Student is disruptive	2.34	2.35	-0.01	-0.01	0.90
Student is attentive	3.27	3.35	-0.08	-0.11	0.07
Sample size (total = 912)	524	388			
Grades 4 and 5					
Student does not complete homework	2.44	2.43	0.01	0.01	0.86
Student is disruptive	2.29	2.21	0.08	0.08	0.20
Student is attentive	3.33	3.31	0.02	0.03	0.66
Sample size (total = 916)	524	392			
Prior-achievement subgroups					
Students scoring at below basic level					
Student does not complete homework	2.58	2.61	-0.03	-0.03	0.67
Student is disruptive	2.43	2.37	0.06	0.06	0.43
Student is attentive	3.12	3.16	-0.04	-0.06	0.48
Sample size (total = 736)	437	299			
Students scoring at basic level					
Student does not complete homework	2.37	2.31	0.06	0.06	0.34
Student is disruptive	2.26	2.25	0.00	0.00	0.95
Student is attentive	3.38	3.38	0.00	0.01	0.92
Sample size (total = 877)	501	376			
Students scoring at proficient level					
Student does not complete homework	1.99	1.85	0.14	0.14	0.37
Student is disruptive	2.14	1.90	0.24	0.23	0.17
Student is attentive	3.64	3.70	-0.07	-0.09	0.57
Sample size (total = 201)	103	98			

(continued)

Table 6.2 (continued)

SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs regular-school-day teacher survey.

NOTES: All survey responses are on a scale of 1 to 4, where 1 equals "Never" and 4 equals "Often."

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size for each outcome is calculated as a proportion of the standard deviation of the regular program group.

There are 7 enhanced program group students and 7 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

The sample size for each outcome varies by the number of regular-school-day teachers who did not respond to the question. Across the analysis sample, the variation ranges from 7 to 13 for the enhanced program group and from 5 to 6 for the regular program group.

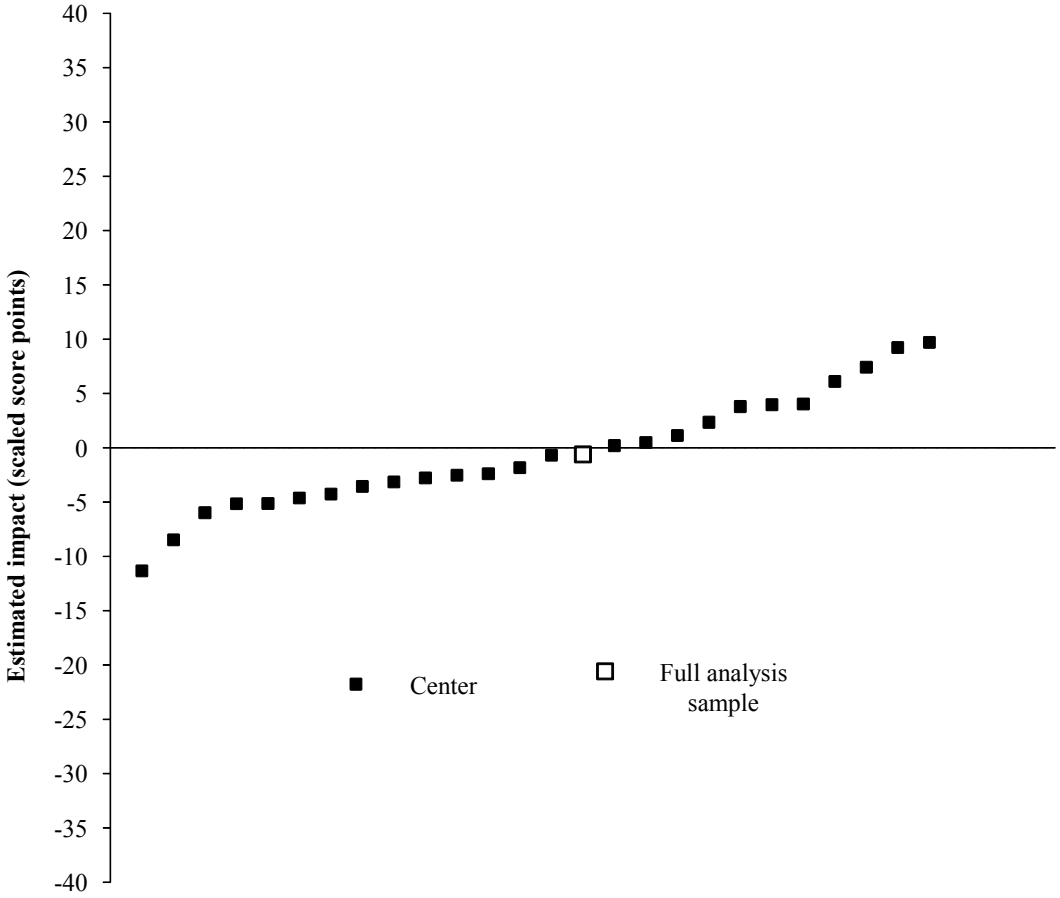
Linking Impact on Total Reading Scores with School Characteristics

Exploratory exercises were conducted to examine whether factors related to program implementation or the school environment in which the enhanced after-school reading program operated are associated with the program impacts for students' academic performance. Even though center-level impacts did not differ at the 5 percent significance level, Figure 6.2 shows that the program effect experienced by each center covers a wide range (from –11.3 to 9.7 scaled score points). Therefore, school characteristics and program implementation measures could still be related to the size and direction of the impact. A multi-level hierarchical model with students nested within centers was utilized to estimate the program impact, and, at the center level of the model, treatment effect was specified as a function of school characteristics as well as of program implementation measures.[109] Notice that this analysis is nonexperimentally based; thus these results should be viewed cautiously and as hypothesis-generating rather than as definitive. Additional analysis of this issue will be conducted when the second year of data is available, but the data available to date allow the team to start examining possible linkages.

[109]See Appendix G for details of the model.

Figure 6.2

Impact of the Enhanced Reading Program on Student Achievement and Its Distribution Across Centers



SOURCES: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The figure shows the estimated program impact for the student-level analysis sample on students' SAT 10 total reading scores (the white box; p-value = 0.51) and how that impact is distributed across the 25 centers in the analysis sample (each dark box). The center-by-center impacts (presented ordinally) are estimated by interacting the treatment indicator with center indicators in an ordinary least squares regression model that also controls for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. Because the study was not designed to detect the impact at the center level (on average, there are only 73 analysis sample students within each center), no statistical tests are conducted to check the significance of the impact estimate for each center. The full analysis sample comprises 1,048 enhanced program group students and 780 regular program group students.

The school characteristics included in the model are the length of reading instruction that students received during the regular school day,[110] whether the school met its Adequate Yearly Progress (AYP) goals, the proportion of students receiving free or reduced-price lunch, and the in-school student-to-teacher ratio. Program implementation measures are the number of days over the course of the school year that the enhanced program was offered and whether one or more instructors teaching the enhanced reading program left during the school year.[111, 112]

Table 6.3 shows estimates of the relationships of school characteristics with the Adventure Island program impact on the SAT 10 total reading test scores. Overall, the full set of school characteristics presented in Table 6.3 is not correlated with the program impact on total reading SAT 10 score (p-value = 0.71), and none of the individual associations (controlling for the other factors listed in Table 6.3) is statistically significant at the 5 percent level.

As mentioned above, since students were not randomly assigned to schools with different characteristics, this analysis is nonexperimental, and the correlational analysis results presented here should not to be interpreted causally.

Conclusion

Overall, data collected during the first-year implementation of the enhanced reading program indicate that teachers reported experiencing difficulty with the pace of instruction. Throughout the year, there was a difference in services offered between the enhanced and the regular program groups. The findings indicate that the first-year implementation of this intervention did not produce statistically significant impacts on students' SAT 10 reading test scores, in terms of both total test scores and subtest scores on word study, vocabulary, and reading comprehension. The estimated impact on one of two measures of fluency is positive and significant.

If one formally adjusts the relevant significance criteria to take account of the multiple reading scores examined (comprehension, fluency, word skill, and so on), the fluency impact is

[110]School administrators were asked how many minutes teachers spend a day teaching math or reading to their students. The responses were not a precise number of minutes, so a continuous measure of minutes is not used. Instead, groups were created around the most common response. For reading, 20 percent offer, on average, less than 90 minutes (in some schools the amount of time varies by grade); about half (52 percent) offer 90 minutes; and the remaining 28 percent offer more than 90 minutes. Thus, the natural split for this subgroup is between schools offering 90 minutes or less and schools offering more than 90 minutes.

[111]School characteristic data come from the 2005-2006 National Center for Education Statistics' Common Core of Data (CCD), which compiles school-level demographic data. Data on whether a school met its AYP goals were obtained from each state's Department of Education Web site.

[112]Not enough was known about the reading curricula used during the regular school day to assess the similarity of the school-day curriculum with the enhanced after-school reading program's materials.

The Evaluation of Academic Instruction in After-School Programs

Table 6.3

Associations Between School Characteristics and the Enhanced Reading Program's Impact on Student Achievement

Interaction Characteristic	Estimated Coefficient	P-Value for the Estimated Coefficient
School		
More than 90 minutes of reading instruction	0.72	0.80
Student to teacher ratio greater than that in the enhanced program[a]	0.73	0.77
Did not make adequate yearly progress (AYP)	-1.00	0.73
Percentage of student body that is low-income[b]	-0.01	0.87
Program implementation		
Enhanced teacher left the program during the school year	-0.08	0.98
Total days enhanced program was offered	-0.16	0.28
F-test of all interaction characteristics		0.71
Size of student sample (total = 1,828)		
Size of school sample (total = 25)		

SOURCES: Student achievement data are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery. Minutes of instruction were collected from research staff interviews with point persons and phone calls made to schools and districts. AYP status was collected from each state's Department of Education Web site. All other school-level characteristics were collected from the Common Core of Data Web site, http://nces.ed.gov/ccd/. Program implementation characteristics are from the Evaluation of Academic Instruction in After-School Programs attendance data and data from Bloom Associates. All data reflect the 2005-2006 school year.

NOTES: The estimated coefficients represent how the reading program impact varies with each school characteristic. They were estimated using a hierarchical linear model, where in the first level (the student level) the following variables are controlled for: treatment status, indicator of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household and mother's education; in the second level (the center level), the program impact is related to the school characteristic variables listed above. The F-test tested whether the coefficients on the school characteristic variables are jointly equal to zero. Within each center, the analysis sample includes, on average, 73 students.

A two-tailed t-test was applied to each estimated coefficient. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

[a]The enhanced program offers a student-to-teacher ratio of at least 13:1.

[b]Student body characteristics are centered on the grand mean of the school sample.

no longer statistically significant.[113] Subgroup analysis results — based both on student characteristics and on school characteristics — are consistent with findings for the full analysis sample. The program also did not produce any significant impacts on students' academic behaviors as measured by answers to a regular-school-day teacher survey. Further exploratory analysis did not provide evidence linking program impact on total reading scores with school characteristics and program implementation measures.

As this report is being written, the study has completed its second year (school year 2006-2007) of data collection. That sample includes students who were part of the study in the first year as well as students who were new to the study in the second year. Thus, the new wave of data will shed light both on the cumulative impact of the enhanced after-school program on returning students and on the impact of a more mature program on new students. Those results will be presented in the final report of the project.

[113]When accounting for multiple test corrections, the Benjamini-Hochberg procedure is used (Benjamini and Hochberg, 1995).

Appendix A

Random Assignment and the Target Sample Sizes

This appendix describes how random assignment was conducted and the size and allocation of the sample assembled.

The Random Assignment Process

At least 15 eligible students were recruited at each grade level in the 25 after-school centers testing each intervention (math or reading), totaling to a research sample of 2,109 students in the math centers and 2,064 students in the reading centers.[1] For programmatic reasons, random assignment was conducted separately within each center, by grade level. (Statisticians call this "blocking" by center and grade.) Even though they were blocked by grade level, the random assignment process for centers took place together, in a batch.

Prior to the point of random assignment, the centers were continuously working to build their sample of students. During this process, centers were urged to identify all potential sample members, rather than a specific number of students. For this and other reasons, until the random assignment rosters were assembled and submitted for random assignment, the exact characteristics of the sample were not known. At this point in the process, the total number of applicants per grade determined the random assignment ratio needed for that center to produce the desired size of the enhanced program group.

The Allocation of the Sample Assembled

In order to assure attendance of approximately 10 students in the enhanced class on any given day, 13 students were assigned to the enhanced program group, as long as at least 21 eligible students in a grade were on the random assignment roster. If there were 15 to 20 eligible applicants in a particular grade, the first 10 random draws were assigned to the enhanced program group so that the class could have the desired minimum number. The abilities of centers to recruit eligible students differed; thus some centers within the study had grades with too few students to produce 10 enhanced program group students with a 1:1 random assignment ratio, while some had grades where there were enough students on the random assignment roster to produce 13 students for the enhanced program group and 13 for the regular program group with a 1:1 ratio. In three cases where there were fewer than 15 students in a grade, students were as-

[1]There are three exceptions to this. One center was able to recruit only 13 third-graders and 13 fifth-graders, and another center was able to recruit only 14 second-graders.

signed in a way that maintained the ratio of two enhanced program group students for every regular program group student.

In instances where the proportion of enhanced program group students to regular program group students differs from 1:1, the power of the sample to detect impacts decreases. To compensate for the smaller sites with 15 students per grade and a 2:1 ratio, larger sites were used to increase the sample size back up to an average of 80 students per center and to move back toward the desired 1:1 ratio. To reflect the random assignment design and control for variance between blocks that random assignment produced, the random assignment block indicators are included as variables in each of the analyses.

Appendix Table A.1 shows the random assignment strategy (the number of enhanced and regular program group students) used for different numbers of students in a grade. If 21 eligible students applied, 13 students were allocated to the enhanced program group, and 8 students were assigned to the regular program group. If 26 students applied, there was a balanced design of 13 enhanced program group students and 13 regular program group students. If 32 students had applied, 13 of them would have been assigned to the enhanced program group, and 19 would go to the regular program group. More than 19 students are not assigned to the regular program group, since it would push the ratio of enhanced program group to regular program group students too far away from the ideal balanced 1:1 design with little increase in statistical precision. However, no sites had more than 32 students in a grade who were available for the study.[2]

The Evaluation of Academic Instruction in After-School Programs

Appendix Table A.1

Planned Random Assignment Ratios Given Varying Numbers of Enrolled Students

Students Enrolled per Grade, per Center	Students Randomly Assigned to Enhanced Program Group	Students Randomly Assigned to Regular Program Group
13	8	5
14	9	5
15-20	10	Remainder
More than 20	13	8-19

[2]In a few cases, exceptions to these rules were made. For example, in one district there was funding for only two teachers to work with regular program group students across all grades. In order to keep the regular program group classes to a manageable size, from a pool of 18 eligible students in a given grade, 12 were allocated to the enhanced program group, and 6 were allocated to the regular program group. Regardless of exceptions, the ratio never went beyond the worst-case scenario of 2:1.

Appendix B

Statistical Precision and Minimum Detectable Effect Size

This appendix reviews the statistical power analysis of the Evaluation of Enhanced Academic Instruction in After-School Programs impact study to determine an acceptable level of precision when estimating the impact of the program. Specifically, it reviews how the sample configuration, use of regression covariates, and other analytic assumptions would affect the precision of the impact estimates. The discussion focuses on achievement test score outcomes because of their prominence in the study.

In the discussion that follows, precision is reported as "minimum detectable effect size" (MDES). Intuitively, a minimum detectable effect is the smallest program impact that could be estimated with confidence given random sampling and estimation error.[1] This metric, which is used widely for measuring the impacts of educational programs, is defined in terms of the standard deviation of student achievement for the underlying population. For example, an MDES of 0.20 indicates that an impact estimator can reliably detect a program-induced increase in student achievement that is equal to or greater than 0.20 standard deviation of the existing student distribution. This is equivalent to approximately four Normal Curve Equivalent (NCE) points on a nationally norm-referenced achievement test and translates roughly into the difference between the 25th and the 31st percentiles.

The discussion that follows presents the smallest impact that the evaluation can reliably detect in effect size. The calculations of MDES for this study account for both within-site and across-site variation in the outcome in question. They also account for random variation across the enhanced program group and the regular program group by including pre-random assignment target test scores (reading or math). Finally, the minimum detectable differences presented here are assumed to be fixed-effect estimates; that is, they do not account for variation across sites in the true impact of the program.[2] This final assumption is justified by the fact that the sites for the study were selected purposefully. Therefore, the results are not generalizable statistically to any larger universe of after-school programs other than the centers included in this particular study.

The first row of each panel in Appendix Table B.1 shows the sample sizes resulting from various configurations of student subgroups for the math program sample and the reading program sample separately. For these rows, the first column shows the actual total number of students in the analysis samples for each subject. Each of the following columns in the table

[1] A minimum detectable effect is defined as the smallest true program impact that would have an 80 percent chance of being detected (have 80 percent power) using a two-tail hypothesis test at the 0.05 level of statistical significance.

[2] The concluding page of this appendix explains how minimum detectable differences are estimated.

shows sample sizes for the subgroups that the study aimed to include. Dividing the full analysis sample into two subgroups according to grade level equally splits the sample and creates two subgroups with 50 percent of the sample size. Defining subgroups based on their prior achievement creates somewhat unequal subgroups, with their sizes ranging from 201 students (for the "proficient" group in the reading sample, which is 11 percent of the full analysis sample) to 1,055 students (for the "basic" group in the math sample, which is 54 percent of the full analysis sample).

The second row of each panel in Appendix Table B.1 shows how the MDES for average achievement scores would vary among sample sizes associated with various configurations of student subgroups.

To see whether there is an overall program impact for math and reading, the analysis will rely on the students in the full analysis sample. For these rows, the first column of numbers indicates that the smallest program impact that could be estimated with confidence (given random sampling and estimation error in the sample) would be 0.06 standard deviation for both math and reading.

In addition to answering questions regarding effects on the full analysis sample of students, the evaluation was designed to allow for the estimation of impacts for subgroups of students defined by pre-random assignment characteristics, including students' grade levels and baseline test scores. For the minimum detectable effect rows, the remaining columns present the estimated MDES for subgroups of students that would comprise 75 percent, 50 percent, 25 percent, or 10 percent of the intended sample. For example, for a subgroup with a quarter of the full analysis sample size (457 to 490 students), the impact estimator can reliably detect a program-induced increase in student achievement that is equal to or greater than 0.12 standard deviation of the existing student distribution.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table B.1

Sample Sizes and Minimum Detectable Effect Sizes for Math and Reading, by Varying Proportions of the Analysis Sample

	Analysis Sample	75% of the Sample	50% of the Sample	25% of the Sample	10% of the Sample
Math					
Sample size	1,961	1,471	981	490	196
Minimum detectable effect size	0.06	0.07	0.08	0.12	0.18
Reading					
Sample size	1,828	1,371	914	457	183
Minimum detectable effect size	0.06	0.07	0.09	0.12	0.20

NOTE: Calculations are based on the formula discussed in Appendix B.

Estimating the MDES

Minimum detectable differences are estimated as follows:

$$MDES = M_{N-J-12} * \sqrt{\frac{\sigma_y^2(1-R^2)}{P(1-P)(N)(\sigma_y^2+\tau_y^2)} + \frac{\omega^2}{J(\sigma_y^2+\tau_y^2)}},$$

where:

M_{N-j-12} = Calculated to be 2.8, assuming a two-tailed test with a statistical power level of 0.80 and a statistical significance level of 0.05 for a sample of J blocks and N students. This multiplier assumes that estimation will include covariates for each block and 12 additional covariates.

σ_y^2 = The (within-block) variance of the outcome in question (assumed to be 1 for the effect size calculations. By definition of effect size metric, this term does not affect the MDES).

R^2 = The explanatory power of the impact regression adjusted for pre-random assignment characteristics, that is, the proportion of the variance in y explained by the experiment and any pre-random assignment characteristics. Based on the collected data, it is assumed to be 0.6.

P = The proportion of students randomly assigned to the treatment group (which equals 0.55 for the math sample and 0.57 for the reading sample).

N = The number of students: equals 1,961 for the math full analysis sample and 1,828 for the reading full analysis sample.

J = The number of grade-center blocks in the study: equals 96 for the math sample and 100 for the reading sample.

τ_y^2 = The cross-block variance in the mean value of the outcome measure y. The variance components of total outcome test scores were estimated for both reading and math and, based on the estimates,

$$\frac{\sigma^2}{\tau^2+\sigma^2} = .51 \text{ for math and } \frac{\sigma^2}{\tau^2+\sigma^2} = 0.55 \text{ for reading.}$$

ω^2 = The cross-site variance in the true impact of the program. The minimum detectable effect sizes presented here are calculated as fixed-effects estimates; that is, they do not account for cross-site variation in the true impact of the program. Thus, ω^2 is assumed to be zero.

Appendix C

Response Rates for Outcome Measures and the Creation of the Analysis Sample

This appendix describes the response rates for the data sources and the creation of the analysis sample used in the math and reading impact analysis. First the math and reading total study samples produced by random assignment are presented. Then the different response rates for the data sources used in the impact analysis are shown. Finally, this appendix compares students who responded and are thus included in the analysis sample with those not in the analysis sample, to make sure that the creation of the analysis sample did not change the specific demographic composition of students created by random assignment.[1]

The Math Sample

The intake and random assignment process produced a full study sample of 2,108 students for the math centers. Appendix Table C.1 shows the baseline characteristics for the full study sample. The response rates within this sample for the data sources are reported in this first panel of Appendix Table C.2.

The first two rows in Appendix Table C.2 show the response rates for the key outcome measures used in the impact analysis — the follow-up SAT 10 total score and the regular-school-day teacher questionnaire. The columns within the table show the percentage of all students who responded to a given measure and the proportion of respondents who are in the enhanced and regular program groups. All response rates are above 90 percent. Ninety-four percent of students (enhanced program group or regular program group) have follow-up SAT 10 math total scores, and the response rates for the teacher questionnaire are between 98 percent and 99 percent. For each data source, there is no significant difference in response rates between the enhanced and regular after-school program groups.[2] The last two rows in the first panel of Appendix Table C.2 report the response rates for the other outcome measures used in analysis: the student survey (to measure the service contrast) and the follow-up state test score (used as a

[1] Attempts were made to collect follow-up data on all students initially randomly assigned into the study, regardless of whether the student was still attending the after-school program. Thus, response rates are not reflective of attrition but, rather, of the ability of data collection staff to gather data from students.

[2] A t-test of the difference between the response rates for each data source was conducted. Differences are not statistically significant at the 0.05 level.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table C.1

Baseline Characteristics of Students in the Math Full Study Sample

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Full study sample						
Enrollment						
2nd grade	513	288	225			
3rd grade	534	291	243			
4th grade	547	297	250			
5th grade	514	292	222			
Total	2,108	1,168	940			
Race/ethnicity (%)						
Hispanic		26.22	23.64	2.58	0.06	0.13
Black, non-Hispanic		46.27	46.19	0.08	0.00	0.96
White, non-Hispanic		21.94	24.99	-3.06	-0.07	0.05
Asian		1.03	1.24	-0.21	-0.02	0.65
Other		4.54	3.94	0.61	0.03	0.49
Gender (%)						
Male		46.83	46.90	-0.07	0.00	0.97
Average age (years)		8.65	8.68	-0.03	-0.02	0.18
Overage for grade[a] (%)		18.41	19.84	-1.44	-0.04	0.39
Free/reduced-price lunch (%)						
Eligible (among information providers)		80.39	79.63	0.76	0.02	0.63
No information provided		3.51	2.57	0.94	0.06	0.21
Average household size		1.92	1.91	0.01	0.01	0.84
Single-adult household (%)		33.45	33.65	-0.20	0.00	0.92
Mother's education level (%)						
Did not finish high school		17.98	18.60	-0.62	-0.02	0.72
High school diploma or GED certificate		34.16	31.31	2.85	0.06	0.16
Some postsecondary study		41.18	44.32	-3.14	-0.06	0.14
No information provided		6.68	5.77	0.91	0.04	0.38
SAT 10 math total scaled scores		568.76	568.66	0.10	0.00	0.94
Problem Solving		573.90	573.19	0.71	0.01	0.61
Procedures		562.55	563.21	-0.66	-0.01	0.70
Sample size (total = 2,108)		1,168	940			

(continued)

Appendix Table C.1 (continued)

Characteristic	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Grade subgroups					
Grades 2 and 3					
Overage for grade[a] (%)	13.99	15.63	-1.64	-0.04	0.45
Mother's education level (%)					
Did not finish high school	19.52	18.62	0.90	0.02	0.71
High school diploma or GED certificate	33.68	30.60	3.08	0.07	0.28
Completed some post-secondary	41.45	45.09	-3.64	-0.07	0.23
No information provided	5.35	5.69	-0.33	-0.01	0.81
SAT 10 math total scaled scores	538.49	537.47	1.02	0.02	0.56
Problem solving	543.81	543.38	0.43	0.01	0.82
Procedures	532.98	530.71	2.28	0.04	0.34
Sample size (total = 1,047)	579	468			
Grades 4 and 5					
Overage for grade[a] (%)	22.75	23.99	-1.24	-0.03	0.63
Mother's education level (%)					
Did not finish high school	16.47	18.58	-2.11	-0.05	0.37
High school diploma or GED certificate	34.63	32.02	2.62	0.06	0.37
Some postsecondary study	40.92	43.56	-2.64	-0.05	0.37
No information provided	7.98	5.85	2.13	0.09	0.17
SAT 10 math total scaled scores	598.51	599.32	-0.81	-0.02	0.67
Problem solving	603.43	602.44	0.99	0.02	0.63
Procedures	591.61	595.18	-3.57	-0.06	0.13
Sample size (total = 1,061)	589	472			
Prior-achievement subgroups					
Students scoring at below basic level					
Overage for grade[a] (%)	27.51	27.24	0.27	0.01	0.95
Mother's education level (%)					
Did not finish high school	22.30	25.98	-3.68	-0.10	0.36
High school diploma or GED certificate	39.03	30.51	8.52 *	0.18	0.05
Some postsecondary study	32.34	37.00	-4.66	-0.09	0.29
No information provided	6.32	6.51	-0.19	-0.01	0.93
SAT 10 math total scaled scores	541.61	540.18	1.43	0.03	0.26
Problem solving	548.12	545.88	2.24	0.05	0.19
Procedures	530.54	529.74	0.80	0.01	0.70
Sample size (total = 516)	269	247			

(continued)

Appendix Table C.1 (continued)

Characteristic	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Students scoring at basic level					
Overage for grade[a] (%)	17.87	19.56	-1.68	-0.04	0.47
Mother's education level (%)					
Did not finish high school	16.95	19.75	-2.80	-0.07	0.24
High school diploma or GED certificate	34.82	33.58	1.24	0.03	0.67
Some postsecondary study	40.99	40.44	0.55	0.01	0.85
No information provided	7.24	6.24	1.01	0.04	0.51
SAT 10 math total scaled scores	564.32	564.54	-0.21	0.00	0.79
Problem solving	569.78	569.47	0.31	0.01	0.78
Procedures	557.43	558.94	-1.51	-0.03	0.32
Sample size (total = 1,125)	649	476			
Students scoring at proficient level					
Overage for grade[a] (%)	10.09	11.25	-1.16	-0.03	0.75
Mother's education level (%)					
Did not finish high school	13.76	6.64	7.12 *	0.18	0.05
High school diploma or GED certificate	28.44	27.34	1.10	0.02	0.82
Some postsecondary study	52.29	61.71	-9.42	-0.19	0.08
No information provided	5.50	4.31	1.19	0.05	0.59
SAT 10 math total scaled scores	601.78	602.30	-0.53	-0.01	0.68
Problem solving	605.39	604.85	0.55	0.01	0.78
Procedures	602.30	604.58	-2.28	-0.04	0.40
Sample size (total = 404)	218	186			

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs a ppli- cation packet and baseline results on the Stanford Achievement Test Series, 10th ed (SAT 10) abbreviated battery.

NOTES: The estimated differences are regression-adjusted using ordinary least squares, controlling for indicators of random assignment strata. The values in the column labeled "Enhanced Program" are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in the next column are the regression-adjusted means using the observed distribution of the enhanced program group across random assignment strata as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated difference effect size for each characteristic is calculated as a proportion of the standard deviation of the regular program group.

F-tests were calculated for the full study sample and each subgroup sample in a regression model containing the following variables: indicators of random assignment strata, math total scaled score, race/ethnicity, gender, free-lunch status, overage for grade, mother's education, mobility, and family size. The F-values are not significant for any of the samples analyzed.

There are 32 enhanced program group students and 31 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]A student is defined as overage for grade at the time of random assignment if a student turned 8 before the start of the second grade, 9 before the start of the third grade, 10 before the start of the fourth grade, or 11 before the start of the fifth grade. This indicates that the student was likely to have been held back in a previous grade.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table C.2

Response Rates to Tests, Surveys, and Observations for Students and After-School Program Staff in the Math Study Sample

Data Source	Full Study Sample	Enhanced Program Group	Regular Program Group
Students[a]			
Key outcome measures			
Follow-up SAT 10[b] (%)	94.17	93.92	94.47
Regular-school-day teacher survey (%)	98.24	97.86	98.72
Additional outcome measures			
Student survey (%)	98.06	98.12	97.98
Follow-up state test score (%)	74.76	74.91	74.57
Full study sample size (total = 2,108)		1,168	940
After-school program staff			
Additional outcome measures			
After-school staff survey[c] (%)		89.57	NA
Interviews and observations[d] (%)		100.00	NA
Sample size[e] (total = 115)			

SOURCES: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery, the Evaluation of Academic Instruction in After-School Programs regular-school-day teacher survey, student survey, and after-school staff survey.

NOTES:

[a]Response rates are calculated from the full study sample for all students in the study and separately for students in each program group.

[b]This calculation is based on responses to the total math scaled score.

[c]Response rates are not calculated for regular program staff because the total sample size of regular program staff is unknown.

[d]The research team observed enhanced group instruction by randomly selecting half (51) of the 102 Mathletics staff teaching at any point in time. Following this observation, they conducted structured interviews with them. The response rate is calculated by taking the number of interviews conducted and dividing it by 51. While 3 instructors of the regular program were observed and interviewed in 2 centers where there was reported to be some structured academic instruction in math, they were not randomly selected, and thus there was no attempt to calculate a response rate for them to this measure.

[e]This is the total number of staff teaching Mathletics over the course of the school year. At a given point in time, 102 staff were teaching classes.

supplementary measure of student's academic performance).[3] Neither of these measures has a statistically significant difference in response rates between the enhanced and the regular after-school program groups. The second panel in Appendix Table C.2 presents the response rates for enhanced program staff measures, such as the after-school staff survey or the interviews and observations.

To keep the sample of students consistent across key outcome measures, an analysis sample was created to contain the students with data from both the follow-up SAT 10 achievement test score and the teacher survey. The flow chart in Appendix Figure C.1 reports the sample sizes of the analysis sample used in the impact analysis. As shown, 19 students are excluded from the math analysis sample because they have a SAT 10 score but no teacher survey; 110 students are excluded because they have a teacher survey but no SAT 10 score; and 18 are excluded because they have neither source of follow-up data. The analysis sample is 93 percent of the full study sample, and the ratio of analysis sample as a proportion of the full study sample is not statistically different between the enhanced program group and the regular program group.[4]

Even though the proportion of students included in the analysis sample is respectably high by social science research standards, it is still less than 100 percent and, therefore, raises two concerns. First, does the analysis sample differ from the full study sample? Second, within the analysis sample, are the enhanced program group and the regular program group still equivalent?

The study team examined the differences in background characteristics between the analysis sample and the rest of the study sample. While the analysis sample reflects the general characteristics of the full study sample (see Appendix Table C.1 for the full study sample's background characteristics and Table 3.3 in Chapter 3 for the analysis group's baseline characteristics), an F-test comparing the students included in the analysis sample and those in the study sample but not the analysis sample indicates that there are systematic differences between them in student characteristics. For example, students are less likely to be included in the analysis sample if their families had moved in the two years prior to the start of this study. Therefore, the students in the analysis sample are not fully representative of the full study sample of 2,108 students. Some caution should be exercised when attempting to generalize the findings beyond those who are included in the impact analysis. Nevertheless, the analysis sample contains 93 percent of students in the full study sample, making the results reflective of the behavior of most of the targeted students.

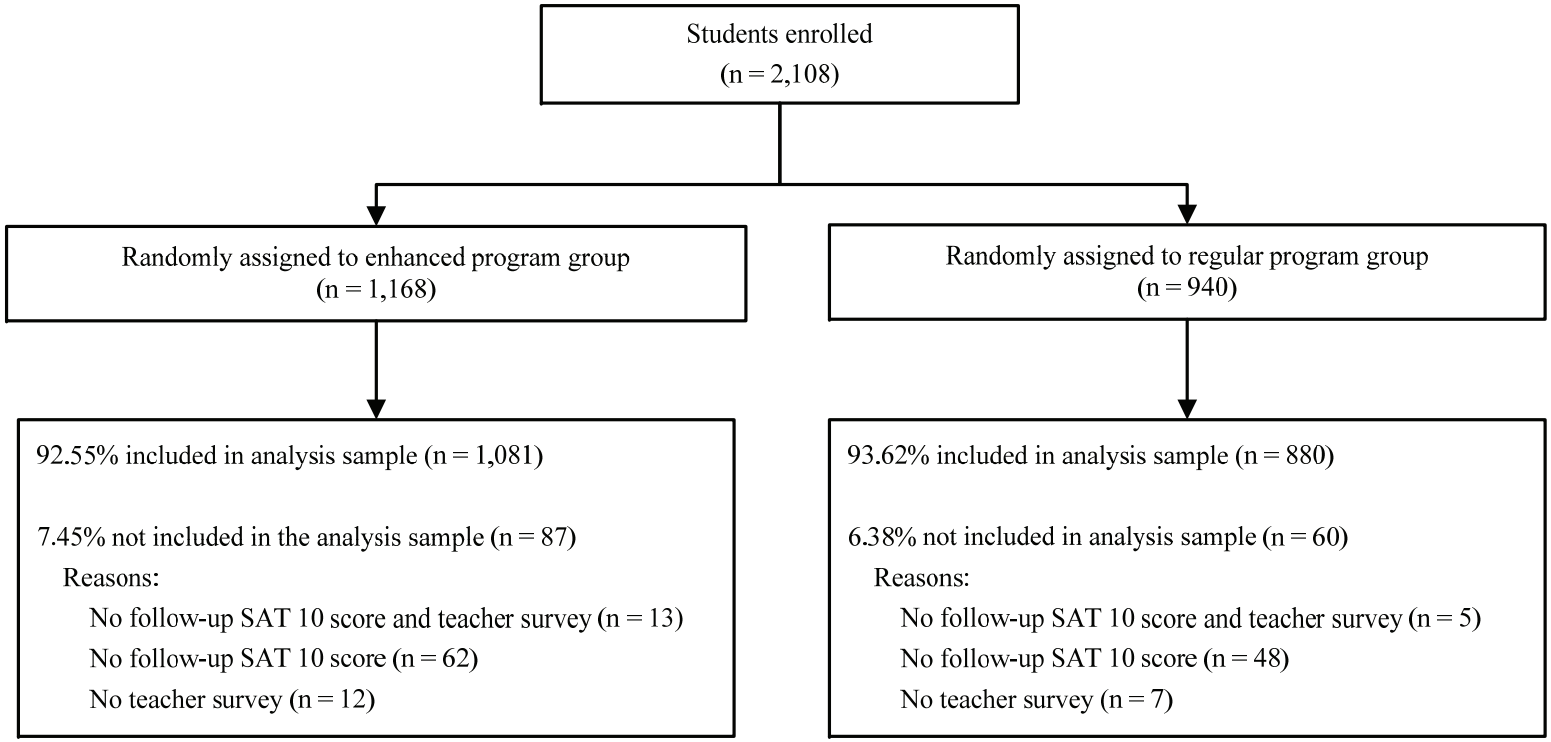
[3]Ten of the 25 schools in the math sample do not test students in grade 2, contributing to a lower response rate for this measure.

[4]Two-tailed t-tests also show that there is no significant variation in the differences in response rates between the enhanced and the regular after-school program groups across math centers, for all outcome measures and the analysis sample.

The Evaluation of Academic Instruction in After-School Programs

Appendix Figure C.1

Flow of Students from Enrollment to Math Analysis Sample



SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs data.

NOTES: This figure explains how the math analysis sample was created from the larger group of students who enrolled in the study. All percentages are based on the number of students randomly assigned to either the enhanced or the regular program group.

In addition, Table 3.3 shows a high degree of similarity between the enhanced program group and the regular program group students in the analysis sample across the baseline characteristics. The characteristic-by-characteristic comparisons and a general F-test all indicate that, overall, there are no systematic differences between these two groups in the analysis sample. The same exercise conducted for each subgroup shows that there also are no systematic differences between the enhanced and the regular program groups at the subgroup level.

The similarity between the student characteristics of the analysis sample and the full study sample, as well as the lack of systematic differences between the enhanced and the regular program groups in the analysis sample, indicate that the analysis sample is appropriate to use in the impact analysis. This conclusion also applies to the samples of students in the subgroup analysis.

The Reading Sample

The intake and random assignment process produced a full study sample of 2,063 students for the reading centers. Appendix Table C.3 shows the baseline characteristics for the full study sample. The response rates within this sample for the data sources used in the impact analysis are reported in the first panel of Appendix Table C.4.

The first four rows in Appendix Table C.4 show the response rates for the key outcome measures used in the impact analysis: the follow-up SAT 10 reading total score, the DIBELS Oral Reading Fluency (ORF) and Nonsense Word Fluency (NWF) scores (fielded to second- and third-graders in the sample), and the regular-school-day teacher questionnaire. The columns within the table show the percentage of all students who responded to a given measure and the proportion of respondents who are in the enhanced and the regular program groups. All response rates are at or above 85 percent. As seen in the table, the response rate for both the enhanced and the regular program group students for the SAT 10 reading total score is between 91 percent and 93 percent. The response rate for both groups for the ORF test is between 88 percent and 90 percent, while the response rate for the other DIBELS portion, the NWF, is between 85 percent and 87 percent. The response rate for all groups for the teacher questionnaire can be rounded to 95 percent. For each data source, there are no significant differences in response rates between the enhanced and the regular after-school program groups.[5] The last two rows in the first panel of Appendix Table C.4 report the response rates for the other outcome measures used in the analysis: the student survey (to measure the service contrast) and the follow-up state

[5]A t-test of the difference between the response rates for each data source was conducted. Differences are not statistically significant at the 0.05 level.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table C.3

Baseline Characteristics of Students in the Reading Full Study Sample

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Full study sample						
Enrollment						
2nd grade	516	296	220			
3rd grade	524	298	226			
4th grade	526	291	235			
5th grade	497	287	210			
Total	2,063	1,172	891			
Race/ethnicity (%)						
Hispanic		23.46	24.24	-0.78	-0.02	0.57
Black, non-Hispanic		63.70	63.38	0.32	0.01	0.81
White, non-Hispanic		8.30	8.22	0.08	0.00	0.93
Asian		1.11	1.42	-0.31	-0.03	0.50
Other		3.42	2.75	0.68	0.04	0.36
Gender (%)						
Male		47.78	49.96	-2.18	-0.04	0.33
Average age (years)		8.72	8.68	0.05	0.03	0.09
Overage for grade[a] (%)		27.22	22.92	4.30 *	0.10	0.02
Free/reduced-price lunch (%)						
Eligible (among information providers)		88.13	86.24	1.89	0.06	0.16
No information provided		5.12	3.94	1.18	0.06	0.22
Average household size		1.92	1.86	0.06	0.06	0.24
Single-adult household (%)		39.53	37.66	1.88	0.04	0.38
Mother's education level (%)						
Did not finish high school		25.09	20.22	4.87 *	0.12	0.01
High school diploma or GED certificate		33.36	30.63	2.73	0.06	0.19
Some postsecondary study		37.29	43.62	-6.33 *	-0.13	0.00
No information provided		4.27	5.54	-1.27	-0.05	0.19
SAT 10 reading total scaled scores		564.36	567.66	-3.31 *	-0.08	0.01
Vocabulary/word reading[b]		554.73	559.87	-5.13 *	-0.10	0.00
Reading comprehension		565.81	569.52	-3.71 *	-0.08	0.01
Word study skills[c]		573.64	574.57	-0.93	-0.02	0.54
Sample size (total =2,063)		1,172	891			

(continued)

Appendix Table C.3 (continued)

Characteristic	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Grade subgroups					
Grades 2 and 3					
Overage for grade[a] (%)	24.58	20.31	4.26	0.10	0.09
Mother's education level (%)					
Did not finish high school	26.77	21.03	5.73 *	0.14	0.03
High school diploma or GED certificate	32.15	28.26	3.90	0.09	0.17
Some postsecondary study	37.37	44.38	-7.00 *	-0.14	0.02
No information provided	3.70	6.33	-2.63 *	-0.11	0.05
SAT 10 reading total scaled scores	536.76	541.98	-5.22 *	-0.13	0.01
Vocabulary/word reading[b]	522.02	530.41	-8.39 *	-0.16	0.00
Reading comprehension	539.02	544.46	-5.44 *	-0.12	0.01
Word study skills	551.87	554.04	-2.17	-0.05	0.31
Sample size (total = 1,040)	594	446			
Grades 4 and 5					
Overage for grade[a] (%)	29.93	25.59	4.34	0.10	0.12
Mother's education level (%)					
Did not finish high school	23.36	19.37	3.99	0.10	0.13
High school diploma or GED certificate	34.60	33.05	1.55	0.03	0.61
Some postsecondary study	37.20	42.84	-5.64	-0.11	0.06
No information provided	4.84	4.74	0.11	0.00	0.94
SAT 10 reading total scaled scores	592.53	593.90	-1.38	-0.03	0.39
Vocabulary	588.30	590.11	-1.81	-0.03	0.39
Reading comprehension	593.30	595.25	-1.95	-0.04	0.32
Word study skills[c]	595.97	595.64	0.32	0.01	0.88
Sample size (total = 1,023)	578	445			
Prior-achievement subgroups					
Students scoring at below basic level					
Overage for grade[a] (%)	33.40	31.08	2.32	0.06	0.48
Mother's education level (%)					
Did not finish high school	27.87	23.84	4.03	0.10	0.22
High school diploma or GED certificate	33.61	32.75	0.86	0.02	0.80
Some postsecondary study	33.81	37.07	-3.26	-0.07	0.34
No information provided	4.71	6.34	-1.63	-0.07	0.33
SAT 10 reading total scaled scores	546.09	547.80	-1.71	-0.04	0.09
Vocabulary/word reading[b]	531.76	535.51	-3.75 *	-0.07	0.04
Reading comprehension	546.20	548.52	-2.33	-0.05	0.11
Word study skills[c]	558.24	556.78	1.47	0.03	0.45
Sample size (total = 835)	488	347			

(continued)

Appendix Table C.3 (continued)

Characteristic	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Students scoring at basic level					
Overage for grade[a] (%)	23.57	19.45	4.12	0.10	0.13
Mother's education level (%)					
Did not finish high school	24.29	19.37	4.91	0.12	0.07
High school diploma or GED certificate	33.21	29.95	3.27	0.07	0.30
Some postsecondary study	38.39	45.56	-7.17 *	-0.14	0.03
No information provided	4.11	5.12	-1.01	-0.04	0.47
SAT 10 reading total scaled scores	573.02	574.77	-1.76 *	-0.04	0.04
Vocabulary/word reading[b]	566.01	569.44	-3.42	-0.06	0.06
Reading comprehension	574.94	577.10	-2.15	-0.05	0.13
Word study skills[c]	579.67	579.26	0.41	0.01	0.82
Sample size (total = 985)	560	425			
Students scoring at proficient level					
Overage for grade[a] (%)	20.00	6.21	13.79 *	0.33	0.01
Mother's education level (%)					
Did not finish high school	18.26	10.48	7.79	0.19	0.19
High school diploma or GED certificate	33.91	33.26	0.65	0.01	0.93
Some postsecondary study	44.35	47.35	-3.00	-0.06	0.71
No information provided	3.48	8.91	-5.43	-0.23	0.14
SAT 10 reading total scaled scores	593.99	594.81	-0.82	-0.02	0.64
Vocabulary/word reading[b]	591.46	593.93	-2.46	-0.05	0.58
Reading comprehension	598.82	600.64	-1.83	-0.04	0.58
Word study skills[c]	602.61	598.81	3.80	0.09	0.39
Sample size (total = 227)	115	112			

(continued)

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs application packet and baseline results on the Stanford Achievement Test Series, 10th ed (SAT 10) abbreviated battery.

NOTES: The estimated differences are regression-adjusted using ordinary least squares, controlling for indicators of random assignment strata. The values in the column labeled "Enhanced Program" are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in the next column are the regression-adjusted means using the observed distribution of the enhanced program group across random assignment strata as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated difference effect size for each characteristic is calculated as a proportion of the standard deviation of the regular program group.

F-tests were calculated for the full study sample and each subgroup sample in a regression model containing the following variables: indicators of random assignment strata, reading total scaled score, race/ethnicity, gender, free-lunch status, overage for grade, mother's education, mobility, and family size.

Appendix Table C.3 (continued)

The full study sample (F-value of 1.74) and the second- and third-grade sample (F-value of 1.73) are significant at the 5 percent level; the fourth- and fifth-grade sample (F-value of 1.58) is significant at the 10 percent level. The F-values for the prior-achievement subgroups are not significant.

There are 9 enhanced program group students and 7 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]A student is defined as overage for grade at the time of random assignment if a student turned 8 before the start of the second grade, 9 before the start of the third grade, 10 before the start of the fourth grade, or 11 before the start of the fifth grade. This indicates that the student was likely to have been held back in a previous grade.

[b]Second-grade students take the word reading subtest, while third- to fifth-grade students take the vocabulary subtest.

[c] The administration of the test to fifth-graders in the spring does not include word study skills.

test score (used as a supplementary measure of student's academic performance).[6] Neither of these measures has a statistically significant difference in response rates between the enhanced and the regular after-school program groups. The second panel in Appendix Table C.4 presents the response rates for enhanced program staff measures, such as the after-school staff survey or the interviews and observations.

To keep the sample of students consistent across key outcome measures, an analysis sample was created to contain the students with data from both the follow-up SAT 10 achievement test score and the teacher survey.[7] The flow chart in Appendix Figure C.2 reports the sample sizes of the analysis sample used in the impact analysis. As shown, 76 students are excluded from the reading analysis sample because they have a SAT 10 score but no teacher survey; 125 students are excluded because they have a teacher survey but no SAT 10 score; and 34 are excluded because they have neither source of follow-up data. The analysis sample is 89 percent of the full study sample, and the ratio of analysis sample as a proportion of the full study sample is not statistically different between the enhanced program group and the regular program group.[8]

[6]Thirteen of the 25 schools in the reading sample do not test students in grade 2, contributing to a lower response rate for this measure.

[7]The sample of students responding to DIBELS is unique, in that it includes only second- and third-graders. Thus, it was not used to create the reading analysis sample, nor is it limited to those students in the analysis sample. There are 96 students included in the DIBELS findings who are not part of the analysis sample: 32 of them have a SAT 10 score but no teacher survey; 53 of them have a teacher survey but no test score; and 11 have neither a SAT 10 score nor a teacher survey.

[8]Two-tailed t-tests also show that there is no significant variation in the differences in response rates between the enhanced and the regular after-school program groups across reading centers, for all outcome measures and the analysis sample.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table C.4

Response Rates to Tests, Surveys, and Observations for Students and After-School Program Staff in the Reading Study Sample

Data Source	Full Study Sample	Enhanced Program Group	Regular Program Group
Students[a]			
Key outcome measures			
Follow-up SAT 10[b] (%)	92.44	93.34	91.25
DIBELS oral reading fluency (%)	89.52	90.40	88.34
DIBELS nonsense word fluency (%)	85.96	86.70	84.98
Regular-school-day teacher survey (%)	94.67	94.71	94.61
Additional outcome measures			
Student survey (%)	96.27	96.67	95.74
Follow-up state test score (%)	74.84	75.77	73.63
Full study sample size (total = 2,063)		1,172	891
After-school program staff[c]			
Additional outcome measures			
After-school staff survey (%)		94.34	NA
Interviews and observations[d] (%)		100.00	NA
Sample size[e] (total = 106)			

SOURCES: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery, results on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessments, and the Evaluation of Academic Instruction in After-School Programs regular-school-day teacher survey, student survey, and after-school staff survey.

NOTES:

[a]Response rates are calculated from the full study sample for all students in the study and separately for students in each program group.

[b]This calculation is based on responses to the total reading scaled score.

[c]Response rates are not calculated for regular program staff because the total sample size of regular program staff is unknown.

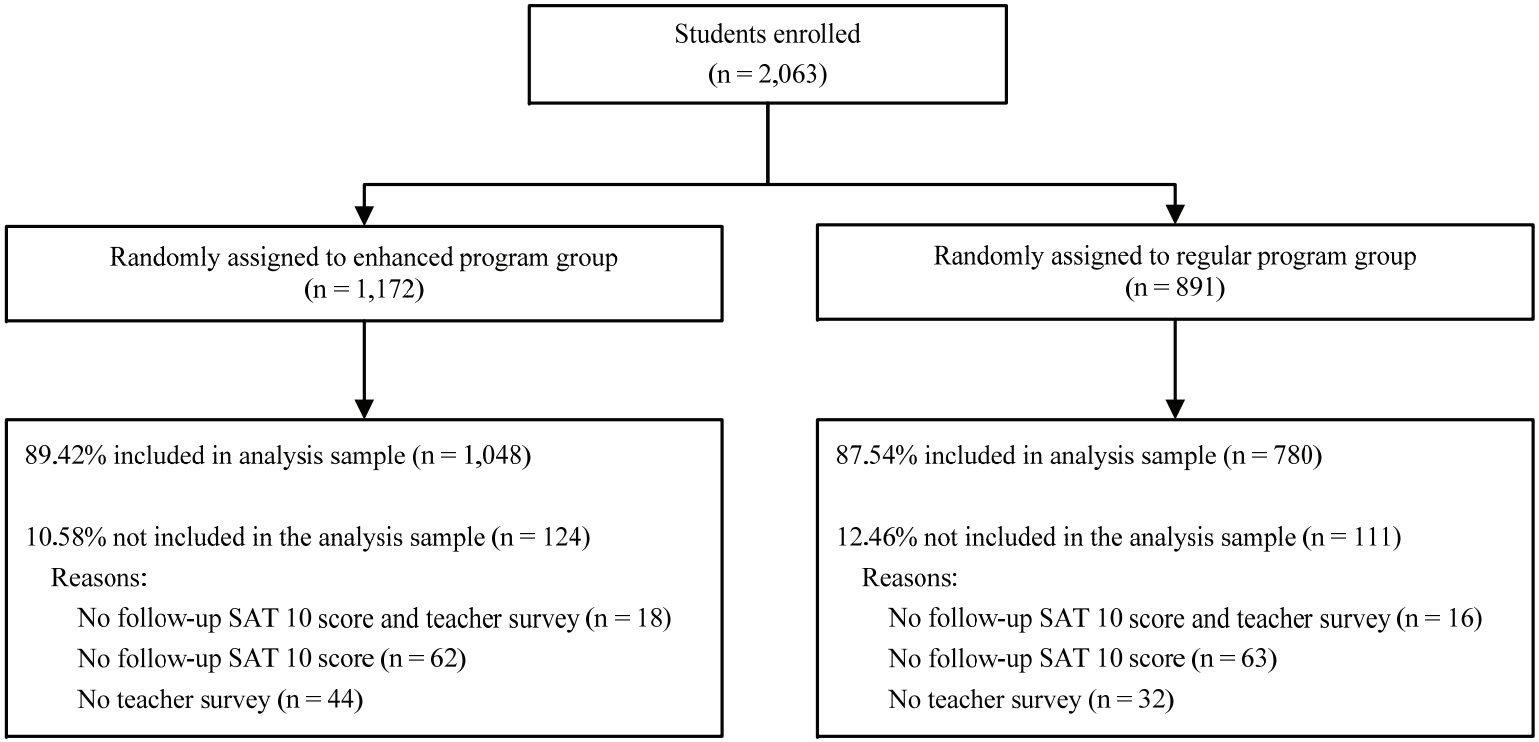
[d]The research team observed instruction by randomly selecting half (50) of the 100 Adventure Island staff teaching at any point in time; following the observation, they conducted structured interviews with them. The response rate is calculated by taking the number of interviews conducted and dividing it by 50. While 5 instructors of the regular program were observed and interviewed in 5 centers where there was reported to be some structured academic instruction in reading, they were not randomly selected, and thus there was no attempt to calculate a response rate for them to this measure.

[e]This is the total number of staff teaching Adventure Island over the course of the school year. At a given point in time, 100 staff were teaching classes.

The Evaluation of Academic Instruction in After-School Programs

Appendix Figure C.2

Flow of Students from Enrollment to Reading Analysis Sample



SOURCE: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs data.

NOTES: This figure explains how the reading analysis sample was created from the larger group of students who enrolled in the study. All percentages are based on the number of students randomly assigned to either the enhanced or the regular program group.

Similar to the math sample, even though the proportion of students included in the reading analysis sample is respectably high by standards for social science research, it is still less than 100 percent and, therefore, raises two concerns. First, does the reading analysis sample differ from the full study sample? Second, within the reading analysis sample, are the enhanced program group and the regular program group still equivalent?

The study team examined the differences in background characteristics between the analysis sample and the rest of the full study sample. While the analysis sample reflects the general characteristics of the study sample (see Appendix Table C.3 for the full study sample's background characteristics and Table 5.3 in Chapter 5 for the analysis group's baseline characteristics), an F-test comparing the students included in the analysis sample and those in the study sample but not the analysis sample indicates that there are systematic differences between them in student characteristics. For example, students are less likely to be included in the analysis sample if they are overage for grade or if information regarding family mobility prior to the start of this study is missing. Therefore, the students in the analysis sample are not fully representative of the full study sample of 2,063 students.

As discussed in Chapter 5 and shown in Table 5.3, for the reading analysis sample, differences between the enhanced program and the regular program groups on most characteristics are not statistically significant, with the exceptions being the differences in the percentage overage for grade (higher for the enhanced group), mother's education (lower for the enhanced program group), and baseline reading test scores (also lower for the enhanced program group).[9] An overall F-test across all available baseline characteristics indicates that there is a statistically significant difference at the 0.05 level between treatment and control groups for the full reading analysis. To control for these observed baseline differences, all baseline characteristics that exhibited statistically significant differences between the enhanced program and the regular program groups are included as covariates in the impact analysis model. Sensitivity tests were also conducted to ensure that the observed baseline differences do not cause selection bias in the impact analysis. (See Appendix F for details of the tests.)

As a result of these sample differences, some caution should be exercised when attempting to generalize the findings beyond students who are included in the impact analysis. Nevertheless, the analysis sample contains 89 percent of students in the full study sample, making the results reflective of the behavior of most of the targeted students.

[9]The baseline test was taken before random assignment but was scored approximately one month after the randomization. Thus, scores were not available to determine eligibility for the study or during the random assignment process.

Appendix D

Structured Protocol Observations

Observations of Implementation of Mathletics and Adventure Island

Structured protocol observations of after-school classes were conducted by local district coordinators who work on-site and were trained by Bloom Associates on the use of their respective structured protocol of implementation. These data were systematically collected to serve two purposes: (1) to provide technical assistance and (2) to describe implementation. District coordinators submitted to Bloom Associates an average of three observations for each teacher over the school year. The write-ups include a checklist of specific intended content coverage and instructional strategies of the enhanced program.

Observation forms (one for the math program and one for the reading program) were developed for this project by Bloom Associates and were reviewed by the research team and the curriculum developers, and they were used by the district coordinators during their formal observations to document whether classes used the curricular materials as intended. The protocols allow the observer to track what portions of the intended lesson are present during the class observed, what is missing entirely, and what has been modified in some way. In addition to the checklist, the write-ups on the forms document how the class was conducted, in light of the structure designed by Harcourt School Publishers or Success for All (SFA). The observation write-ups capture answers to the question "Did they do it?"

Observations of Mathletics

Appendix Box D.1 presents the guidelines for assigning points, based on which Mathletics instructional elements were recorded on the observation form as being present during the enhanced class. Bloom Associates, the curriculum developers, and the research team developed this list to summarize the observations. For the math program, a teacher could receive a maximum score of 6 points per observation by using all the instructional elements (shown in Appendix Box D.1), which include the following: sole use of the curricular materials throughout the instructional period, establishment of routines that allow for smooth transitions between the parts of the instructional session and maximizing time on task, inclusion of a teacher-led warm-up and cool-down for all students, provision of direct and differentiated instruction during the workout, use of other workout components (such as skill packs) appropriately, and inclusions of all the components in the allocated times.

Appendix Box D.1

Math Instructional Elements: Guidelines for Assigning Points

For each of the six areas listed below (uses of curriculum materials, classroom management, warm-ups and cool-downs, direct/differentiated instruction, appropriate use of other program components, structure of lesson and pacing), the district coordinator was instructed to indicate evidence of fidelity by checking bulleted items that were present. Points by area are assigned as indicated. For some of the areas, all bulleted items needed to be checked to be awarded points. In other places, an "or" indicates that only one of the bulleted items needed to be checked. Each classroom observation was recorded as a sum of the points awarded based on this protocol and point distribution scheme. NOTE: There are a total of 6 possible points for the enhanced math curriculum.

Uses curriculum materials. 1 point is awarded if:

- Observer checked box indicating students are engaged in a teacher-led Harcourt Warm-up and Cool-Down exercise;
- Observer checked box indicating the teacher provides direct instruction to small groups using page 1-2 of Skill Pack in both rotations; and
- Observer checked box indicating students work independently on the other components, such as:
    - pages 3-4 of skill packs,
    - Harcourt software connected to instruction plan, or
    - play the 24 Game and/or other Harcourt board games

[Note: A point was not given if the notes section indicated that other materials were used under any of the categories.]

Classroom management. 1 point is awarded if:

- Observer checked box indicating that during the workout portion of the class, teacher directs students to stations using established method of communication and students move quickly; or
- Notes indicate teacher uses recommended management strategies such as Popsicle sticks, rotation charts, timers, etc.

Warm-ups and cool-downs. For each, 1/2 point is awarded if:

- Observer checked box indicating students are engaged in a teacher-led or supported Harcourt numbered warm-up (or cool-down) assignment; and
- Notes indicate that all students participated (e.g., the teacher checked all students' work as she circulated…)

(continued)

Appendix Box D.1 (continued)

Direct/differentiated instruction (to individuals and small groups in rotations). 1 point is awarded if:

- Observer checked box indicating teacher provides direct instruction to small groups using pages 1 and 2 of skill pack in both rotations

Appropriate use of other components. 1 point is awarded if:

- Observer checked box indicating students moved to different activities during rotations, such as:
  - skill pack pages 3 and 4,
  - use of Harcourt software connected to the instructional plan, or
  - Harcourt board games/24 game
- When looking at the numbers of students (and their names in the notes section) assigned to component parts of the workout session, within each rotation, there is distribution across the activities mentioned above

Structure of lesson and pacing. 1 point is awarded if:

- Observer checked box indicating each component section (Warm-ups, Workout Session and Cool-downs) is completed in the allotted timeframe

Each class was observed, on average, three times during the year. For each class, observation scores were averaged together.[1] Appendix Table D.1 indicates to what extent instructional elements were present; 93 percent of classes implementing Mathletics received a score of more than 5 points, on average. In other words, a class that was observed three times may have received 5 of 6 possible points during two of the observations and 6 of 6 possible points during a third observation. The average score for that class is 5.3.

Observations of Adventure Island

Appendix Box D.2 presents the guidelines for assigning points, based on which Adventure Island instructional elements were recorded on the observation form as being present during the enhanced class. The instructional elements recorded for the reading program include slightly different components for the higher and lower reading levels, with a maximum score of

[1]Classroom scores are each teacher's mean score across all observations; when more than one teacher taught a class (for example, a teacher left the program in the middle of the year and was replaced), their mean scores are averaged together. This produces one score per grade at each center and indicates, for example, the average level of implementation that a student in a fourth-grade class at that center experienced.

Appendix Table D.1

Distribution of Structured Protocol Observation of Implementation Scores Across Mathletics Classrooms

Average Score	Percentage of Classrooms Receiving Score
Less than or equal to 1	0.00
Greater than 1 to 2	0.00
Greater than 2 to 3	0.00
Greater than 3 to 4	0.00
Greater than 4 to 5	7.45
Greater than 5 to 6	92.55
Sample size (total = 94)	

SOURCE: Structured protocol observations of implementation conducted by local district coordinators.

NOTES: Enhanced classes were observed, on average, three times during the year by district coordinators and were given a score by Bloom Associates. Classroom scores are each teacher's mean score across all observations; when more than one teacher taught a class, their mean scores are averaged together. All enhanced classes were scored on a scale of 1 to 6.

5 points per observation for Discovery Bay and Treasure Harbor classes and 6 points per observation for Alphie's Lagoon and Captain's Cove classes.[2] The instructional elements (shown in Appendix Box D.2) are a mixture of procedural factors (use of curricular materials, implementation of cooperative learning strategies, awarding of points to reward cooperative learning and the use of fluency techniques, and completion of lesson plan in the allotted time) and indicators for whether key topics were covered (phonics, fluency, and comprehension).

Each class was observed, on average, three times during the year. For each class, observation scores were averaged together.[3] Appendix Table D.2 indicates to what extent instructional elements were present. For the lower reading levels, 31 percent of classes implementing Adventure Island received a score of more than 5 points, on average. In other words, a class that

[2]Alphie's Lagoon classes (which focus on beginning-reader skills) and Captain's Cove classes (which focus on second-grade reading skills) include topics that cover phonics. Discovery Bay classes (which focus on third-grade reading skills) and Treasure Harbor classes (which focus on fourth-grade reading skills) do not include phonics as a key element.

[3]Classroom scores are calculated by taking each teacher's mean score for a specific Adventure Island level, then averaging those scores across all teachers with a score for that level at that center. This produces one score per level at each center and indicates, for example, the average level of implementation that a student in an Alphie's Lagoon class at that center experienced.

Appendix Box D.2

Reading Instructional Elements: Guidelines for Assigning Points

The Success for All (SFA) Adventure Island curriculum consists of four levels: Alphie's Lagoon, Captain's Cove, Discovery Bay, and Treasure Harbor. For each of the eight areas listed below (uses curriculum, models comprehension, completes lesson in allotted time, uses cooperative learning strategies, awards points for cooperative learning, models fluency, awards points for fluency, teaches phonics in Alphie's Lagoon and Captain's Cove), the district coordinator was instructed to indicate evidence of fidelity by checking bulleted items that were present. Points by area are assigned as indicated. For some of the areas, all bulleted items needed to be checked to be awarded points. In other places, an "or" indicates that only one of the bulleted items needed to be checked. Each classroom observation was recorded as a sum of the points awarded based on this protocol and point distribution scheme. NOTE: There are a total of 6 possible points for the Alphie's Lagoon and Captain's Cove curricula. There are a total of 5 possible points for the Discovery Bay and Treasure Harbor curricula.

Uses curriculum. 1 point is awarded if:

- Observation checklist includes name of SFA book title/day filled in on top portion; and
- Check marks assigned to relevant lesson segments and the notes sections refer to SFA curriculum as appropriate

Models comprehension. 1 point is awarded if:

- For Alphie's Lagoon, observer checked box indicating
  - story preview/review,
  - partner word and sentence reading, and
  - guided group or guided partner reading segments, when applicable
- For Captain's Cove, Discovery Bay, and Treasure Harbor, observer checked box indicating
  - the Build Background, Reading Comprehension, and Mini Lesson segments; and
  - the relevant teacher and students practice routines are highlighted or noted, such as:
    - teacher helps students make connections between their prior knowledge and the skill being taught;
    - teacher models strategy/skill;
    - teacher prompts students to review previously read text each day and make predictions, supported by evidence;
    - teacher reads aloud from the student (or secondary) text and presents additional instruction/modeling of the strategy/skill; or
    - teacher closely monitors student reading and prompts strategy use as necessary

(continued)

Appendix Box D.2 (continued)

Completes in allotted time. 1 point is awarded if:

- For all curricula,
  - the observer checks yes on the 2 prompts (1) did class begin on time and (2) timing and pacing
- For Captain's Cove, Discovery Bay, and Treasure Harbor,
  - the lesson segment check boxes (with time segments) are checked, and the notes sections do not indicate a problem with time

Uses cooperative learning strategies. 1/2 point is awarded if:

- The observer highlights or notes key words from the teacher and students practices sections of the observation protocol, such as –
  - uses Think-Pair-Share;
  - numbered heads; or
  - students actively participate in partnerships and teams

Awards points for cooperative learning. 1/2 point is awarded if:

- The observer checked box indicating "the teacher awards points for cooperation" on the Team Score Sheet section of the guide; or
- The notes section of appropriate lesson segments and/or observer comments in the general notes section at the end of the protocol indicate that cooperative learning points were awarded

Models fluency. 1/2 point is awarded if:

- In Alphie's Lagoon, the observer
  - highlights or notes key words from the teacher and student practices column of the protocol, such as —
    - teacher models fluent reading, or
    - students work with partners to read words, sentences and stories;
- In Captain's Cove, Discovery Bay, and Treasure Harbor, the observer
  - checks and/or notes key words from the sections for Partner reading and Fluency portions such as —
    - students practice fluency; or
    - teacher closely monitors practices
- In Captain's Cove, the observer checks marks in the Reading Olympics check box

Awards points for fluency. 1/2 point is awarded if:

- For all levels, the observer checks "teacher awards points for fluency"; or
- There are references in the notes sections that teacher awarded points for fluency

(continued)

Appendix Box D.2 (continued)

Teaches phonics in Alphie's Lagoon and Captain's Cove. 1 point is awarded if:

- For Alphie's Lagoon, observer checked box indicating
  - All applicable lesson segment sub-headings for the following three routines: Fast Track Phonics, Partner Word and Sentence reading, and Guided Group reading; or
  - The corresponding teacher and student practices descriptors are highlighted or referred to in notes sections
- For Captain's Cove, observer checked box indicating
  - Sail Along lesson segment; or
  - The corresponding teacher and student practices descriptors are highlighted or referred to in notes sections

was observed three times may have received 5 of 6 possible points during two of the observations and 6 of 6 possible points during a third observation. The average score for that class is 5.3. For the higher reading levels, 35 percent of classes received a score of more than 4 points, on average. In other words, a class that was observed three times may have received 4 of 5 possible points during two of the observations and 5 of 5 possible points during a third observation. The average score for that class is 4.5.

Observations of Reading and Math Instructional Practices

Observations of instructional practice were conducted by the research team using the same protocol in both math and reading sites. It is a tool developed by Public/Private Ventures (P/PV) to assess a variety of instructional variables of after-school activities. P/PV has been refining the instrument for over 10 years. P/PV has used the instrument in four previous studies of after-school programs, most recently in the CORAL (Communities Organizing Resources to Advance Learning) evaluation, which is an outcomes evaluation of an after-school literacy initiative funded by the Irvine Foundation. For the CORAL study, the instrument yielded reliable scales for such constructs as adult-youth relationships, instructional quality, and classroom management (Arbreton, Goldsmith, and Sheldon 2005).

To create the instrument, P/PV researchers reviewed both the literature on instructional practices linked to positive student learning outcomes and the after-school literature on practices linked to increased participation, to generate a set of underlying variables, or "constructs and

The Evaluation of Academic Instruction in After-School Programs

Appendix Table D.2

Distribution of Structured Protocol Observation of Implementation Scores Across Adventure Island Classrooms

Average Score	Percentage of Classrooms Receiving Score
Alphie's Lagoon and Captain's Cove classrooms	
Less than or equal to 1	2.08
Greater than 1 to 2	0.00
Greater than 2 to 3	0.00
Greater than 3 to 4	16.67
Greater than 4 to 5	50.00
Greater than 5 to 6[a]	31.25
Sample size	48
Discovery Bay and Treasure Harbor classrooms	
Less than or equal to 1	0.00
Greater than 1 to 2	7.50
Greater than 2 to 3	5.00
Greater than 3 to 4	52.50
Greater than 4 to 5	35.00
Greater than 5 to 6[a]	NA
Sample size	40

SOURCE: Structured protocol observations of implementation conducted by local district coordinators.

NOTES: Enhanced classes were observed, on average, three times during the year by district coordinators and were given a score by Bloom Associates. Classroom scores are calculated by taking each teacher's mean score for a specific Adventure Island level, then averaging those scores across all teachers with a score for that level at that center.

[a]Alphie's Lagoon classes, which focus on beginning-reader skills, and Captain's Cove classes, which focus on second-grade reading skills, are scored on a scale of 1 to 6. Discovery Bay classes, which focus on third-grade reading skills, and Treasure Harbor classes, which focus on fourth-grade reading skills, are scored on a scale of 1 to 5.

subconstructs," that seemed relevant to an after-school setting.[4] P/PV also included classroom management and adult responsiveness because those have been correlated with positive student learning outcomes (Grossman, Campbell, and Raley 2007; Miller 2006). Dimensions related to the context of the activity — such as the adequacy of the classroom space, materials, and the time allotted for completion — were also included in the observation instrument because they can affect students' ability to benefit from the activity. Finally, the observation instrument included descriptive characteristics of the activity, such as the schedule and number of adults and students present.

Constructs

The observational instrument gathers information of four overarching constructs: Instructional Delivery, Classroom Management, Cooperative Learning, and Space/Material/Time. This section describes the set of items that the team assessed to measure each construct. (The "Q" followed by a number indicates the question number for that item on the observation scales form.) The responses for all the items were done using a 4-point scale, where 1 is a low or negative rating and 4 is a high or positive rating. The following are the definitions and exact instructions that were given to observers indicating what the numbers mean:

4 = Outstanding. A score of 4 should be given when the dimension being rated is exemplary. The behaviors observed are both positive and in terms of their quality and intensity are outstanding examples of the construct; and nothing about the activity (in terms of this construct) can be improved upon. This score should be used relatively infrequently. As with all scores, ratings of 4 must be thoroughly backed up with detailed examples and descriptions of the activity along this construct.

3 = Good or very good. The activity was strong, with numerous examples of positive behaviors and no negative examples. However, while positive, the examples were not particularly outstanding. It might be helpful to think of this score as "one step down" from a score of 4 — good, but you can imagine better.

[4]Constructs are underlying variables that cannot be directly measured, such as "instruction." A construct can theoretically be made up of several subconstructs, such as organization and instructional clarity. To get a gauge — albeit an indirect gauge — of the underlying construct, a measure is created that is a collection of single-question items believed to be related to the underlying construct. (These measures are often referred to as "scales." Later, this appendix describes the scales used in this study.) This appendix uses the word "construct" to imply the underlying variables, "scale" or "measure" to indicate the indirect gauge of the construct, and "item" to specify the single question that is partially correlated with the underlying construct (DeVellis 2003).

2 = Could use improvement. There is some positive (but weak) evidence of the construct, but in contrast to a score of 3, there are also more negative examples. Significant improvement would be necessary for the activity to be considered good. A "2" may also be given in instances in which no positive behaviors are noted, if there were no negative examples either.

1 = Definitely needs improvement. There is little, if any, evidence of the construct, or predominantly negative examples. This score is also appropriate in cases where the activity is not a "bad" activity, but is simply not designed to address the construct. For example, an activity in which the adult meets with youth one-on-one without any peer interaction would receive a "1" on Peer Cooperation.

Instructional Delivery

This construct describes the manner in which the lesson is presented and its ability to create meaningful connections for youth. The construct includes the following six items.

ORGANIZATION (Q2)

This item evaluates the instructor's organization in presenting the lesson. Organization is key to successfully conveying information and instructions to youth, gaining youth's respect, and taking advantage of the limited time available in the after-school hours. Organized instructors have all materials at-hand and are prepared to start the activity on time, make efficient use of instructional time, and remain on task throughout the lesson. In assessing this item, observers must consider whether the staff appeared prepared to present the whole lesson. Did the instructors keep students focused on the activity's goals? Did they present topics with a logical sequence? On the other hand, did instructors often have to "back track" because they forgot to mention key points (making the activity seem poorly planned and disorganized)? Were they not organized enough to move smoothly from one activity to the next during the lesson?

MODELING BEHAVIOR (Q3)

This item evaluates the instructor's skill in showing students how to use the techniques being taught. In assessing this item, observers were instructed to think about whether modeling occurred during the course of the activity and whether the instructor missed obvious opportunities for modeling. When asked a question, did the teacher provide the answer or help the children think through the steps that would help them get the answer themselves?

CLARITY OF PRESENTATION (Q4)

This item assesses whether the instructor presented the goals and instructions for the activity clearly, enabling youth to move through each step of the activity without confusion. In assessing this item, observers were asked to consider the following: Did instructors explain the goals of the activity to youth in a way they could understand? Did instructors give clear and accurate directions?

Clarity of presentation is also reflected in youth's responses to the activity. Did youth know how to proceed? Did they seem confused? Were instructions provided to youth in manageable "chunks" or thrown at them in a confusing, fast-paced manner that seemed to lose them along the way? At the same time, because there may be instances in which the instructor presents materials in an extremely clear manner, yet youth are still disengaged, observers were told to base their assessment on the instructor's presentation, not the youth's response.

CONNECTION-MAKING (Q6)

This item assesses the instructor's ability to connect specific activities with other lessons and material covered and students' experiences. Successful connection-making allows youth to see the relationships between what they learn one day and the next, between their personal experiences and the material or between what they learn in school and in the after-school activity. When these connections are clear, it is easier to see why an activity is meaningful. Creating these connections also helps remind students of what they have already learned, thus making it more likely that they retain this learning.

To assess this item, observers were asked to consider the extent to which the instructor provided a context for the activity. Did s/he make connections between the current lesson and past lessons, such as explaining how the current activity relates to previous activities? Did s/he contrast or compare new information with previously learned material? Did s/he relate the current lesson to future lessons? Did s/he clearly explain how any games or activities relate to the material covered? The teacher also might ask youth what they know about the topic, referencing something in the neighborhood, or connecting the material to media or pop culture that interests the youth. Observers were instructed to assess the extent to which the instructor clearly placed each activity within the context of other material, lessons, and concepts. An activity that seemed isolated or disconnected to other material or the students' lives would score low on this item.

BALANCES INDIVIDUAL INSTRUCTION AND GROUP ACTIVITY (Q7)

There are two items in Q7. Q7a focuses on the structure of the activity and whether it was primarily an individual or group activity, as measured by the proportion of time devoted to each. Q7b focuses primarily on how well the instructor transitioned and moved between group

and individual activities. (If an activity is entirely group or individual, observers were instructed to rate Q7b as N/A since there is no transitioning between group and individual work.) Group activities were those that include the entire class. Small group and individual work were considered individual activities.

Classroom Management

This construct looks at how the instructor interacted with the students and whether the instructor managed students' behavior during the activity in ways that are appropriate for the age of youth involved and the type of activity. Successful and appropriate behavior management is essential to quality activities because it provides a positive environment for student learning (National Research Council Institute of Medicine 2004).

ADULT MANAGEMENT (Q9)

This item assesses the quality and effectiveness of the techniques staff use to manage youth behavior during the activity. How staff deal with youth who misbehave, become distracted, or disrupt the activity are key aspects of the measure. Staff's management techniques should enable the activity to precede smoothly, and at the same time, should be firm but warm. This can be displayed in a number of ways, but in all cases the adults are able to redirect the youth and win their cooperation without yelling or resorting to critical, punitive or negative discipline tactics. If behavioral issues do occur, the teacher handles them calmly and resolves them quickly and successfully. The adult handles any discipline challenges that arise without getting noticeably angry, frustrated or becoming embroiled in "power struggles" with youth. The staff may be strict with youth, but are able to correct their behavior while maintaining a positive regard and respect for the youth.

TEACHER'S INCLUSIVENESS OF YOUTH (Q10)

This item assesses the extent to which staff try to include all youth in the activity. Staff may show inclusiveness by directing questions to youth who appear isolated. Does the teacher talk to every youth at least once? Do any youth appear to be isolated, without any attention from staff?

ADULT RESPONSIVENESS (Q11)

This item assesses the quality of adult responsiveness toward students in the activity. Adult responsiveness is important for youth because it can make youth feel successful and help them benefit from the activity.

One form of adult responsiveness is the extent to which adults offer guidance to help youth understand and succeed at the task at hand, whether by providing extra information or

encouragement for youth who need it, or making themselves accessible by walking around the room or sitting at a table with youth. Adult Responsiveness includes efforts that are specifically focused on helping *all* youth to reach the goals of the activity, not just a few. Even if youth don't accept offers of help, these efforts should be noted.

MONITORING (Q5)

Teachers use monitoring techniques to assess students' progress and provide feedback. These techniques may involve asking questions to check students' understanding of the concepts being taught, circulating throughout a classroom to check on individual or group progress or providing opportunities for young people to self-assess their learning (such as checking their own work).

Within activities, a teacher must monitor both individual and group progress--and in different ways. Therefore, there are two items in Q5. Q5a focuses on how the instructor monitors and provides feedback to individuals during the direct instruction segment(s). Q5b focuses on how the instructor monitors individual progress when they are working independently and not under direct instruction (i.e., computer work, test taking, independent reading).

Cooperative Learning

Activities that are strong in peer cooperation should enable youth to interact positively with and learn from their peers. Research has shown that activities that encourage cooperative learning enhance students' desire to attend the activity more frequently (Grossman et al. 2007). This section assesses the character of the activity's peer learning environment. Q8 has two parts.

COOPERATIVE LEARNING (Q8a)

The first item focuses on the extent to which the activity requires working in pairs or collaborative problem solving (e.g., team games) and the proportion of time youth actually spend in cooperative learning activities.

MONITORING OF COOPERATIVE LEARNING (Q8b)

The second item focuses on how effectively the instructor monitors cooperative learning and actively encourages youth to work together.

Quality of Space/Material/Time

APPROPRIATENESS OF SPACE (Q12)

Dimensions of space or materials are considered under this construct — e.g., crowding, lighting, noise, quality of materials and adequacy of time. To be appropriate, an activity should not have a major problem with any dimension.

Maximizing Scoring Consistency

Observational data were collected by 16 researchers from MDRC and P/PV.

- A one-day training in Philadelphia was held for the 16 people who conducted the observations, during which time each construct and item was discussed, focusing on its behavioral indicators, how it differed from other items, and what the different rankings of scores meant.
- A scoring manual that included definitions of each item and the types of behaviors that would be positive and negative indicators was produced. The manual was distributed during the training, and observers were instructed to review it prior to conducting an observation.
- During the first two site visits, a researcher who was familiar with the instrument was paired for the same observation with a researcher who was less familiar with the instrument. Each pair rated the activity separately and then met to compare scores and resolve any discrepancies. Discussions of discrepancies served to clarify the scoring system and the definition of each item and thus increase consistency among observers.

Given the number of researchers involved in data collection, conducting traditional inter-rater reliability among all researchers was not feasible. Instead, to maximize the consistency in scoring among the group of researchers, P/PV subjected the observers' ratings of each item to review by a single P/PV researcher, who took the following steps:

- Each observer was asked to submit, along with the observation form, a detailed running record of the entire activity as well as narrative summaries of each construct.
- After the forms, narrative summaries, and running records were sent to P/PV, the P/PV researcher reviewed all the descriptive notes and ratings and compared the numerical rating scores for each item against the narrative summary of the construct and the details of the running record to check for inter-

nal rating consistency. If the numerical rating was not consistent with the narrative summary and running record, a suggested rating was written on the form by the reviewing researcher, and the form was sent back to the observing researcher with directions to review the manual again to ensure that the initial rating considered the scoring guidelines. Following this, the observing researcher provided further justification for the initial rating or accepted the change. Using a single reviewer to check each observer's ratings against the recorded details of the activity maximized consistency across observers.

Appendix E

Outcome Measures

This appendix describes the measures selected for each of the two outcome domains assessed in the study: academic achievement and academic behavior. (See Appendix Table E.1 for a summary of basic descriptive information about each outcome measure.)

Academic Achievement

At the heart of this study is a question about the impact of the enhanced after-school program on the academic achievement of students. Past evaluations, including the prior evaluation of after-school programs by Mathematica Policy Research (Dynarski et al. 2003, 2004), have relied on a nationally normed achievement test of the type used by districts or states to monitor academic performance.

Recognizing that policymakers are interested in such standardized tests, the research team, working with its Technical Work Group and the Department of Education, focused its efforts on identifying an appropriate test of math and reading for the study to administer at baseline and the end of the school year.

Study-Administered Math and Reading Test Instrument Selection

There were several criteria for selecting the achievement tests. The test used in the evaluation needed to cover grades 2 through 5 with a common framework for reporting scores and needed to have various versions, or "forms," allowing administration in both the fall (baseline) and the spring (follow-up). An effort was made to consider what tests are already being used in the study school districts and to not duplicate the testing already happening. Additionally, it was important that the test be:

1. **Accepted by the research community as a reasonable test.** In reading, there is a fairly developed view of what the key skills are for early reading (based on the National Reading Panel), and it is important that a reading test for the early grades actually measure these key skills. In math, there is not such a consensus (based on the National Mathematics Advisory Panel 2008).

2. **Seen as a policy-relevant measure of achievement.** The test should be seen as measuring the kinds of things that schools are being expected to teach and as testing them in a way that is similar to state and local accountability systems.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table E.1

Descriptive Information on Each Outcome Measure

	General Information	Norm Sample and Psychometric Properties
Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery	Commercially available. Math test contains two subtests: problem solving and procedures. Reading test contains three subtests: word study skills, reading comprehension, and vocabulary.	Normed to a national sample of 250,000 students in spring 2002 and of 110,000 students in fall 2002. The average student in the norm sample has a normal curve equivalent score of 50, and the standard deviation of normal curve equivalent scores is 21.06. Internal consistency (KR-20) reliability coefficients range from 0.77 to 0.95 for abbreviated multiple-choice battery test and subtests.
Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	Commercially available. Contains a set of standardized, individually administered measures of early literacy development, used to monitor the development of pre-reading and early reading skills.	Benchmark and progressive goals initially were derived based on data from all schools participating in the DIBELS Data System during the 2000-2001 and 2001-2002 academic years. Test-retest reliability for elementary students ranges from 0.92 to 0.97.
State-administered tests	Norm-referenced tests are commercially available. Criterion-referenced tests are developed specifically for a state and are not commercially available. (See Appendix Tables E.2 and E.3 for a listing of the tests.)	No norming and psychometric properties are available for the criterion-referenced tests. For the norm-referenced tests: • TerraNova reading assessment: Normed to a national sample of 171,000 students. Internal consistency coefficients range from 0.76 to 0.97 for the complete battery test. • Scantron math assessment: Inter-testlet internal consistency coefficients range from 0.512 to 0.876. Correlations between individual units and overall score range from 0.747 to 0.876. Highly predictive correlation with the Iowa Test of Basic Skills as well as the Dakota State Test of Educational Progress. This computer-adaptive test stops testing the student once it reaches a reliability coefficient of 0.91.
Regular-school-day teacher survey	Questions constructed by MDRC or adapted from questions used in other after-school evaluations.[a] Survey items cover issues on homework completion and academic behavior in class.	This measurement is not nationally normed.

SOURCES: Harcourt Assessment (2004); Dynamic Indicators of Basic Early Literacy Skills (2007a); Salvia and Ysseldyke (2001); Scantron Corporation (2005); Dynarski et al. (2004).

NOTE: [a]Three single-item questions used as school-day academic behavior outcomes were drawn from the "Elementary School Teacher Survey" used for the National Evaluation of the 21st Century Community Learning Centers Program study.

3. **Feasible to administer in the after-school setting.** The realities of after-school programs and the staffing available to field the tests create some constraints in administration. Thus, the goal was to pick a test that is relatively straightforward for staff without special expertise to administer to groups of students and that takes no longer than an hour or possibly 90 minutes to administer.

4. **Scored in a way that can be combined across grades in the analysis.** In order to conduct the analysis on the full sample, the test must yield scores or measures that can be combined across grades.

5. **Sensitive to improvements at the bottom range of the achievement distribution.** The most important target group for enhanced instruction in after-school programs is students who are not doing well in school, so the goal was to pick a test that is good at picking up the changes at the low end of the distribution.

From these criteria, a list was created of possible tests, and this list was presented to the Technical Working Group, along with a memo explaining the rationale for why each test was on the list. From this list the Stanford Achievement Test, Tenth Edition (SAT 10), abbreviated battery was chosen.[1]

The SAT 10 abbreviated battery is a group-administered multiple-choice test of one hour or less. This test is widely used, nationally recognized, similar to tests that are part of state and/or local accountability systems (so it has policy relevance), and is relatively easy to administer. Based on the Technical Data Report by Harcourt:

> Stanford 10 full-length and Stanford 10 Abbreviated are both expressed on the same underlying ability scale. Although the relationship of raw score to ability may differ from one test form to another, the relationship of ability (scaled score) to percentile rank is the same. There is in essence a single norm set which applies equally to any Stanford 10 form linked to the underlying Stanford 10 scale. Thus, any information that pertains to norms for the Stanford 10 full-length test applies equally to Stanford 10 Abbreviated. Because the abbreviated form is a core subset of items on the full-length form, all of the validity information for the full-length form applies equally

[1]The SAT 10 is published by Harcourt Assessment, a sister organization of Harcourt School Publishers, which is the creator of the new math curriculum. However, the SAT 10 operates separately, and the Harcourt math curriculum is not especially aligned with the "Stanford" test.

> to the abbreviated form. The only real difference is that since the abbreviated form has fewer items, it does not measure with quite the same precision as the full-length test due to the slightly lower reliability. (Harcourt Assessment 2004, p. 46)

The SAT 10 abbreviated battery is normed to a national sample of 250,000 students in spring 2002 and of 110,000 students in fall 2003. The average student in the norm sample has a Normal Curve Equivalent (NCE) score of 50, and the standard deviation of NCE scores is 21.06. The internal consistency (KR-20) coefficients range from 0.77 to 0.95 for the abbreviated multiple-choice battery test and subtests. There is well-documented evidence of its content, criterion-related, and construct validity (Harcourt Assessment 2004). The test was administered at both baseline and follow-up, covering the topic (reading or math) addressed in the curriculum to be tested in the site.

The reliability coefficients of the abbreviated measure for the total reading score for grades 2 through 5 range from 0.90 to 0.93 for the spring test and from 0.93 to 0.95 for the fall test. For total math score, the reliability measures for grades 2 through 5 range from 0.89 to 0.92 for the spring test and from 0.88 to 0.92 for the fall test. For more details, see Appendix C of the Stanford Achievement Test Series, Tenth Edition, Technical Data Report (Harcourt Assessment 2004).

The math test contains two subtests — problem-solving and procedures — that measure content and process. Problem-solving measures the skills and knowledge necessary to solve problems in mathematics through geometry and measurement; patterns, relationships, and algebra; and data, relationships. and probability. Procedures measure the ability to apply the rules and methods of arithmetic to problems that require arithmetic solutions through computation with whole numbers, decimals, and fractions (Harcourt Assessment 2007).

The reading test contains three subtests — word study skills, reading comprehension, and vocabulary — that reflect and support a balanced, developmental curriculum and sound instructional practices. Word study skills measures structural and phonetic analysis, such as identifying and decoding compound words and contractions and recognizing sounds of consonants and vowels. Reading vocabulary measures students' understanding of the printed word, synonyms, and multiple-meaning words. Reading comprehension measures students' initial understanding, interpretation, and critical analysis of reading passages (Harcourt Assessment 2007).

Study-Administered Fluency Test Instrument Selection

In addition to the SAT 10 test, the research team was advised to include a measure of fluency at follow-up for the younger students in the reading sample. Younger students are more

likely to first show improvement in fluency before improving in overall comprehension, as measured by the SAT 10 standardized test (National Reading Panel 2000). Individually administered tests that are both short and fairly easy to administer were considered. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) was selected and administered at follow-up to second- and third-graders in the reading centers, in addition to the SAT 10.

The DIBELS are "a set of standardized, individually administered measures of early literacy development. They are designed to be short (one minute) fluency measures used to monitor the development of pre-reading and early reading skills" (Dynamic Indicators of Basic Early Literacy Skills 2007a). DIBELS benchmark and progressive goals initially were derived based on data from all schools participating in the DIBELS Data System during the 2000-2001 and 2001-2002 academic years. And test-retest reliability for elementary students ranges from 0.92 to 0.97 (Dynamic Indicators of Basic Early Literacy Skills 2007a). Numerous additional studies have replicated the predictive utility of these goals in other, diverse samples. In this study, students were tested on measures of fluency — oral reading fluency (ORF) and nonsense word fluency (NWF).

The ORF assesses a child's skill in reading connected text. "Student performance is measured by having students read a passage aloud for one minute. Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds are scored as accurate. The number of correct words per minute from the passage is the oral reading fluency rate" (Dynamic Indicators of Basic Early Literacy Skills 2007b). Students in the study were asked to read three passages, and their median score was used in the analysis.

The NWF assesses a child's knowledge of "letter-sound correspondence and of the ability to blend letters into words in which letters represent their most common sounds (Dynamic Indicators of Basic Early Literacy Skills 2007c). The student is presented an 8.5-x-11-inch sheet of paper with randomly ordered vowel-consonant and consonant-vowel-consonant nonsense words (for example, sig, rav, ov) and is asked to produce verbally the individual letter-sound of each letter or to verbally produce, or read, the whole nonsense word. "For example, if the stimulus word is 'vaj,' the student could say /v/ /a/ /j/ or say the word /vaj/ to obtain a total of three letter-sounds correct. The student is allowed 1 minute to produce as many letter-sounds as he/she can, and the final score is the number of letter-sounds produced correctly in one minute. Because the measure is fluency based, students receive a higher score if they are phonologically recoding the word and receive a lower score if they are providing letter sounds in isolation" (Dynamic Indicators of Basic Early Literacy Skills 2007c).

School Record Data

The study also collected information about student performance on the locally administered tests from school record data and used these test scores as a supplementary measure of students' academic performance. The locally administered tests are also more likely to be a full battery and might measure math or reading more reliably than the abbreviated version of SAT 10 used by the study. On the other hand, these locally administered tests also may be testing a slightly different set of skills than tested by the abbreviated SAT 10. Thus, they provide a different measure of reading or math skill.

Each school district has its own specific test, so the closest measure to a total reading and total math score was used. (See Appendix Tables E.2 and E.3 for a list of math tests and reading tests available to the study sites.) In order to pool across the sites and estimate overall impact for the sample, each student's test score was standardized in the following way:

$$Z_{ij} = \frac{(Y_{ij} - \overline{Y}_j)}{s.d._j(Y_{ij})}$$

where:

Z_{ij} = the standardized score for student *i* from site *j*.

Y_{ij} = the raw score for student *i* from site *j* in the locally administered test.

$\overline{Y}_j$ = the average raw score for site *j* in the locally administered test.

$s.d._j(Y_{ij})$ = the standard deviation of the raw test scores for site *j*.

This transformed measure was then used as an outcome for student achievement.

Academic Behavior

Measures of students' academic behaviors come from the regular-school-day teacher survey conducted in the spring of the first program year. For each student in the study sample, the regular-school-day teacher was asked to fill out a short survey about any special academic support that the student receives during the school day and how the student behaved in the regular-school-day class. Specifically, teachers rated their students on the following:

Q6. How often does this student NOT complete homework?

Q7. How often is this student disruptive?

Q9. How often is this student attentive in class?

For each of these questions, the teacher was asked to choose from (1) Never, (2) Not very often, (3) Sometimes, and (4) Often. The answers, therefore, were coded on the scale of 1 to 4, with 1 indicating "Never" and 4 "Often."

The Evaluation of Academic Instruction in After-School Programs

Appendix Table E.2

Math District Tests, by State

Test	Criterion- or Norm-Referenced	Test Content
Standardized test administered to study students		
Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery	Norm-referenced	Number Sense and Operations; Patterns, Relationships, and Algebra; Geometry and Measurement; Data, Statistics, and Probability; Communication and Representation; Estimation; Mathematical Connections; Reasoning and Problem Solving; Mathematical Procedures
State-administered tests		
California Standards Tests (CST)	Criterion-referenced	Grade 3: Number Sense - Place Value, Addition and Subtraction; Number Sense - Multiplication, Division, and Fractions; Algebra and Functions; Measurement and Geometry; Statistics, Data Analysis, and Probability Grade 4: Number Sense - Decimals, Fractions, and Negative Numbers; Number Sense - Operations and Factoring; Algebra and Functions; Measurement and Geometry; Statistics, Data Analysis, and Probability Grade 5: Number Sense - Estimation, Percents, and Factoring; Number Sense - Operations with Fractions and Decimals; Algebra and Functions; Measurement and Geometry; Statistics, Data Analysis, and Probability
Connecticut Mastery Test (CMT)	Criterion-referenced	Numerical and Proportional Reasoning; Geometry and Measurement; Working with Data: Probability and Statistics; Algebraic Reasoning: Patterns and Functions; Integrated Understandings
Georgia Criterion Referenced Competency Tests (CRCT)	Criterion-referenced	Number Sense and Numeration; Geometry and Measurement; Patterns and Relationships; Statistics and Probability; Computation and Estimation; Problem Solving
Florida's Comprehensive Assessment Test (FCAT)	Criterion-referenced	Number Sense, Concepts, and Operations; Measurement; Geometry and Spatial Sense; Algebraic Thinking; Data Analysis and Probability

(continued)

Appendix Table E.2 (continued)

Test	Criterion- or Norm-Referenced	Test Content
Pennsylvania System of School Assessment (PSSA)	Criterion-referenced	Numbers and Operations; Measurement; Geometry; Algebraic Concepts; Data Analysis and Probability
Scantron Math (administered by the State of Kansas)	Norm-referenced	Algebra; Geometry; Measurement; Data Analysis & Probability; Number & Operations
Stanford Achievement Test Series, 10th ed. (SAT 10) full battery (administered by the State of Alabama)	Norm-referenced	Number Sense and Operations; Patterns, Relationships, and Algebra; Geometry and Measurement; Data, Statistics, and Probability; Communication and Representation; Estimation; Mathematical Connections: Reasoning and Problem Solving; Mathematical Procedures
Texas Assessment of Knowledge and Skills (TAKS)	Criterion-referenced	Numbers, Operations, and Quantitative Reasoning; Patterns, Relationships, and Algebraic Reasoning; Geometry and Spacial Reasoning Measurement; Probability and Statistics; Mathematical Processes and Tools
Wisconsin Knowledge and Concepts Examinations - Criterion Referenced Test (WKCE-CRT)	Criterion-referenced	Mathematical Process; Number Operations and Relationships; Geometry; Measurement; Statistics and Probability; Algebraic Relationships

SOURCES: Information on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery, was retrieved from the Harcourt Assessment Web site. State test names, formats, and contents were provided by in-house district data, test assessment Web sites, and state Department of Education Web sites.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table E.3

Reading District Tests, by State

Test	Criterion- or Norm-Referenced	Test Content
Standardized test administered to study students		
Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery	Norm-referenced	Reading Comprehension - initial understanding, interpretation, and critical analysis of reading passages; Reading Vocabulary - understanding of the printed word, synonyms, and multiple meaning words; Word Study Skills - structural and phonetic analysis, such as identifying and decoding compound words and contractions and recognizing sounds of consonants and vowels
State-administered tests		
California Standards Tests (CST)	Criterion-referenced	Word Analysis; Reading Comprehension; Literary Response and Analysis; Writing Strategies; Written Conventions
Georgia Criterion Referenced Competency Tests (CRCT)	Criterion-referenced	Vocabulary; Comprehension; Reading for Literacy Comprehension; Reading for Information; Reading Skills and Vocabulary Acquisiton; Functional and Media Literacy
Florida's Comprehensive Assessment Test (FCAT)	Criterion-referenced	Words and Phrases in Context; Main Idea, Plot, and Purpose; Comparisons and Cause/Effect; Reference and Research
Louisiana Educational Assessment Program (LEAP)	Criterion-referenced	Read, Comprehend, and Respond to a Range of Materials; Write Competently; Use Conventions of Language; Apply Speaking and Listening Skills [not assessed]; Locate, Select, and Synthesize Information; Read, Analyze, and Respond to Literature; Apply Reasoning and Problem-Solving Skills
New Mexico Standards Based Assessment (NMSBA)	Criterion-referenced	Reading and Listening for Comprehension; Writing and Speaking for Expression; Literature and Media

(continued)

Appendix Table E.3 (continued)

Test	Criterion- or Norm-Referenced	Test Content
New York State English Language Arts	Criterion-referenced	Understand Story Events; Draw Conclusions; Make Predictions; Identify the Main Idea; Use Text to Understand Unfamiliar Vocabulary Words; Identify Supporting Details; Identify Point of View; Evaluate Ideas Based on Prior Knowledge; Follow Ideas and Events in the Text; Distinguish Fact from Opinion; Understand Features That Distinguish Genres; Use Figurative Language to Interpret Text
Pennsylvania System of School Assessment (PSSA)	Criterion-referenced	Comprehension and Reading Skills; Interpretation and Analysis of Fiction and Non-Fiction Text
TerraNova (administered by the State of Michigan)	Norm-referenced	Basic Understanding; Analyze Text; Evaluate and Extend Meaning; Identify Reading Strategies
Texas Assessment of Knowledge and Skills (TAKS)	Criterion-referenced	Basic Understanding; Literary Elements; Analysis Using Reading Strategies; Analysis Using Critical-Thinking Skills
Wisconsin Knowledge and Concepts Examinations - Criterion Referenced Test (WKCE-CRT)	Criterion-referenced	Determine the Meaning of Words and Phrases in Context; Understand Text; Analyze Text; Evaluate and Extend Text

SOURCES: Information on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery, was retrieved from the Harcourt Assessment Web site. State test names, formats, and contents were provided by in-house district data, test assessment Web sites, and state Department of Education Web sites.

Appendix F

Estimating Effects and Assessing Robustness

This appendix provides a detailed discussion of the statistical model used to estimate the program impacts and other related statistical issues. It also discusses various tests that were used to assess the robustness of the impact estimates reported in the text and provides the results for these tests.

Analysis of Program Impacts

The program impact analysis involves examining outcome measures constructed from the follow-up student achievement tests, a survey of regular-school-day teachers, and student records from participating districts, with key outcomes listed in Chapter 2. Note that all the listed outcomes are measured at the level of individual students. These outcomes are used to calculate the estimates of impacts of each of the two academic programs separately (the math or the reading program) by comparing outcomes for the enhanced program group and the regular program group within the after-school centers and grade levels.

The Model

Impacts of reading and math programs were estimated separately. For each outcome, the basic model used in the analysis is the following:

$$Y_{ijk} = \gamma_0 Y_{-1,ijk} + \beta_0 T_{ijk} + \sum_k \sum_j \gamma_{1kj} B_{ijk} + \sum_S \gamma_{2s} X_{sijk} + \varepsilon_{jk} \qquad (1)$$

where:

T_{ijk} = one if student i from grade j in center k is assigned to the enhanced program and zero otherwise.

Y_{-1ijk} = the pretest score for student i from grade j in center k before random assignment.[1]

[1]Pretest scores are scaled scores from the SAT 10 (SAT 9 for a couple of centers) reading and math tests administered in the fall of 2005, before the start of the after-school program. Total scores for each subject were used in the analysis of respective samples.

B_{ijk} = block dummy, one if student *i* is in a particular random assignment block, defined by grade j, center *k,* and zero otherwise, $k = 1$ to 25, $j = 1$ to 4.

X_{sijk} = the *s* other student-level covariates for student *i* from grade *j* in center *k*.

ε_{ijk} = a student-level random error respectively, assumed to be independently and identically distributed.

The coefficient, β_0, represents the overall impact of being randomized to enhanced instruction instead of the regular after-school program *for an average student in the sample.* The traditional t-statistic for this coefficient tests whether the estimated average impact for the sample of students in the study centers is statistically significantly different from zero. This analysis does not attempt to generalize statistically beyond the observed sample of sites; thus, the traditional t-test is appropriate.

There are several features worth noting in this model:

- β_0 is a "fixed-effect" estimate that addresses the question: What is the program effect of enhanced instruction for the average student in the sample? This approach is taken because the goal of this study is to conduct an efficacy study of the effects of a new approach, and sites are not selected to be a random sample of a larger population of sites.
- Ordinary Least Squares (OLS) regression is used to estimate Equation (1).
- Indicators for each of the blocks used in the random assignment process (B_{ijk}, defined by the center and the given grade level of the student on the baseline questionnaire) are included in the model to reflect the design feature (that is, differential rates of treatment assignment, by block) and control for the variation in mean outcome level (which can be due to different characteristics of centers, school settings, and so on) across blocks.
- The model controls for individual-level pretest measure. This information can increase the precision of impact estimates, especially for fixed-effect models, because pretests substantially reduce random posttest error, which is the sole source of uncertainty in a fixed-effect model.
- Other baseline covariates are added to the model to improve precision. These covariates include student's gender, race/ethnicity, free/reduced-price lunch status, age, whether a student is from a single-adult household, whether a student is overage for grade, and the mother's education level.

The design also allows the research to detect effects among subgroups of students that are defined by characteristics depicting a student's pre-random assignment condition. To be parsimonious, subgroups on two theoretically relevant and policy-relevant characteristics were examined: subgroups based on students' grade levels and baseline academic performances.

Other Analytical Issues

Missing Covariates

For the baseline achievement test, there are 22 missing cases (11 for math and 11 for reading). For other covariates, there are very few (5 percent or less) missing cases.[2] To keep the sample as complete as possible, the missing values were imputed with the mean value of the center-by-grade-by-treatment-status block to which the student belongs.[3] If more than 5 percent of the observations are missing data for a given variable, then a dummy variable indicating whether a student is missing this covariate or not was also included.

Missing Outcome Measures

Missing data for outcomes pose a problem that is more serious and more difficult to solve because it requires omitting sample members from the impact analysis, which can produce selection bias if this attrition is substantial and nonrandom. As discussed in Appendix C, response rates in this study were in general above 85 percent, and the student characteristics of the full study sample and the analysis sample are similar. Therefore, of the full sample, 147 math (7 percent) and 235 reading (11 percent) students with missing outcome measures were excluded from the impact analysis sample.

[2]Among the students in the reading sites, 4 are missing a race/ethnicity indicator; 83 are missing a free lunch status indicator; 19 are missing information about single-adult household; and 92 are missing information about mother's education. Among the students in the math sites, 2 are missing a race/ethnicity indicator; 58 are missing a free lunch status indicator; 36 are missing information about single-adult household; and 121 are missing information about mother's education. (No students are missing indicators of gender or age.)

[3]Rather than imputing the missing reading or math SAT-10 total scaled score, the mean score for the missing subtest raw score was imputed, and then the subtest raw scores were added, and that student was assigned a scaled score for the given raw score. Thus, if there is an actual score for one or more of the subtests, the imputed total score will incorporate the actual subtest scores.

Additional Tests and Checks

For the Math Sample

In addition to the math program impact results presented in Chapter 4 of the report, the program's impacts on student performance in locally administered math tests were also estimated, to compare with those on SAT 10 tests. The locally administered tests are mostly full-battery tests and might measure math skills more reliably than the abbreviated tests used by the study.

An important caveat for this comparison relates to data availability. The locally administered test data were not always available for second-graders in those study sites that start testing students in the third grade. As a result, all second-graders were excluded from this analysis, and the total sample size for the locally administered test analysis is 1,310 for math.

Appendix Table F.1 presents the estimated program impacts on student performance in locally administered tests for math. Because these test scores were standardized within each study site, all estimated impacts are in effect size units.[4] The table also shows the program impact on the study-administered SAT 10 tests for the sample of students whose local test scores were available for comparison purpose. Because second-graders were excluded from the analysis, the table does not show impact estimates for total scores for the subgroup of second- and third-graders.

For the math sample, all five estimates have the same sign in both measures. The estimated effect sizes for the local tests are in the same direction but with differing magnitudes than those estimated for study-administered SAT 10 total scores, and they are not statistically significant. On the other hand, the program impacts on the SAT 10 math total scores are statistically significant for the subgroup of fourth- and fifth-graders and for the subgroup of students who performed at "basic" level before the program started. This pattern for subgroup findings is the same as the one shown in Table 4.1 for the math analysis sample. Furthermore, the following checks were conducted to see whether the impact estimates on SAT 10 test scores are robust:

- All impacts were reestimated for the sample of all SAT 10 respondents to make sure that no imbalance was created when the full study sample was limited to the analysis sample.

[4]Appendix E describes the standardization of the test score variable.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.1

Impact of the Enhanced Math Program on Student Achievement for Grades 3 to 5

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
State test analysis sample					
State test scaled scores	0.02	-0.01	0.03	0.03	0.49
SAT 10 math total scaled scores	620.62	618.35	2.26	0.05	0.08
Sample size (total = 1,310)	729	581			
Grade subgroup					
Grades 4 and 5					
State test scaled scores	0.03	-0.06	0.09	0.09	0.08
SAT 10 math total scaled scores	626.33	622.27	4.06 *	0.09	0.01
Sample size (total = 921)	515	406			
Prior-achievement subgroups					
Students scoring at below basic level					
State test scaled scores	-0.72	-0.77	0.06	0.06	0.54
SAT 10 math total scaled scores	594.09	591.54	2.56	0.06	0.34
Sample size (total = 347)	184	163			
Students scoring at basic level					
State test scaled scores	0.07	0.02	0.05	0.05	0.40
SAT 10 math total scaled scores	618.83	614.98	3.86 *	0.09	0.04
Sample size (total = 679)	397	282			
Students scoring at proficient level					
State test scaled scores	0.72	0.83	-0.11	-0.11	0.39
SAT 10 math total scaled scores	653.78	656.57	-2.79	-0.06	0.46
Sample size (total = 239)	126	113			

(continued)

SOURCES: MDRC calculations are from results on state tests administered in the 2005-2006 school year and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: State test data were not available for most second-graders because many of the study sites begin testing students in the third grade, and, as a result, all second-graders are excluded from this analysis. In addition, the state test analysis sample is restricted to those from the full analysis sample for whom a state test score was obtained. The resulting state test analysis sample represents 88 percent of the third- through fifth-graders in the full analysis sample and is used to calculate the SAT 10 and state test findings presented.

Each student's state test score was converted into a standardized score because school districts in different states administer different tests. See Appendix E for details.

Appendix Table F.1 (continued)

Based on the SAT 10 national norming sample, math total scaled scores have the following possible ranges: for the state test analysis sample, scores range from 428 to 796; for the fourth- and fifth-grade subgroup, scores range from 450 to 796.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the state test score is calculated as a proportion of the state test score standard deviation of the regular program group from the state test analysis sample. The estimated impact effect size of the SAT 10 math total scaled score is calculated as a proportion of the standard deviation of the regular program group from the full analysis sample, which is 44.64. The standard deviation of a SAT 10 national norming sample with the same grade composition as the full analysis sample is 39.00.

There are 22 enhanced program group students and 23 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

This change in the sample added 19 observations for the math sample. Appendix Table F.2 presents student achievement impact results for math, using the SAT 10 respondents from the full study sample. The general patterns of the findings do not change at all.

- All impacts were reestimated with a model that has no covariates other than the "block" (random assignment unit) indicators, the treatment status indicator, and prior achievement.

In other words, the following model was used to estimate the program impacts:

$$Y_{ijk} = \sum_{m}\sum_{n} \gamma_{0mn} B_{ijk} + \beta_0 T_{ijk} + \gamma_1 Y_{-1,ijk} + \varepsilon_{ijk} \qquad (2)$$

The variables are defined as before. Because this study is based on a randomized experiment, both sets of estimates — those with or those without controlling for other baseline characteristics — provide an unbiased estimate of the treatment effect. The precision of the estimated impact, however, is likely improved by controlling for other baseline characteristics.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.2

Impact of the Enhanced Math Program on Student Achievement for the SAT 10 Respondent Sample

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
SAT 10 respondent sample					
SAT 10 math total scaled scores	604.73	601.99	2.73 *	0.06	0.01
Problem solving	605.85	603.40	2.45 *	0.05	0.04
Procedures	604.91	600.81	4.11 *	0.08	0.01
Sample size (total = 1,980)	1,093	887			
Grade subgroups					
Grades 2 and 3					
SAT 10 math total scaled scores	582.83	581.07	1.76	0.04	0.29
Problem solving	584.43	583.60	0.83	0.02	0.62
Procedures	583.21	579.11	4.10	0.08	0.08
Sample size (total = 984)	542	442			
Grades 4 and 5					
SAT 10 math total scaled scores	626.26	622.55	3.71 *	0.08	0.01
Problem solving	626.88	622.76	4.12 *	0.09	0.01
Procedures	626.26	622.25	4.01 *	0.07	0.05
Sample size (total = 996)	551	445			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 math total scaled scores	583.67	580.85	2.82	0.06	0.22
Problem solving	585.91	582.93	2.98	0.07	0.23
Procedures	579.20	576.70	2.50	0.05	0.43
Sample size (total = 474)	243	231			
Students scoring at basic level					
SAT 10 math total scaled scores	600.28	597.00	3.28 *	0.07	0.03
Problem solving	601.53	598.03	3.50 *	0.08	0.03
Procedures	600.48	595.58	4.89 *	0.09	0.02
Sample size (total = 1,062)	616	446			
Students scoring at proficient level					
SAT 10 math total scaled scores	634.24	631.27	2.98	0.07	0.31
Problem solving	633.61	629.98	3.63	0.08	0.22
Procedures	639.74	637.54	2.20	0.04	0.61
Sample size (total = 384)	205	179			

(continued)

Appendix Table F.2 (continued)

SOURCE: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The SAT 10 respondent sample is composed of all students from the full study sample who have a follow-up SAT 10 math total score.

Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: for the SAT 10 respondent sample, scores range from 389 to 796, 414 to 776, and 413 to 768; for the second- and third-grade subgroup, scores range from 389 to 741, 414 to 719, and 413 to 715; and for the fourth- and fifth-grade subgroup, scores range from 450 to 796, 468 to 776, and 485 to 768.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 math total scaled score is calculated as a proportion of the standard deviation of the regular program group from the analysis sample, which is 44.64. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.00. For each subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group from the analysis sample.

There are 29 enhanced program group students and 31 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

As can be seen from Appendix Table F.3, dropping these covariates from the model affected the precision of the impact estimates but did not affect the magnitudes or the patterns of the impact findings, as one would expect from a randomized experiment.

- All impacts were reestimated with a model that has no covariates other than the "block" (random assignment unit) indicators and the treatment status indicator.

In other words, the following model was used to estimate the program impacts:

$$Y_{ijk} = \sum_{m}\sum_{n} \gamma_{0mn} B_{ijk} + \beta_0 T_{ijk} + \varepsilon_{ijk} \qquad (3)$$

The variables are defined as before and, as can be seen from Appendix Table F.4, dropping covariates from the model and controlling only for the randomization strata did not affect the magnitudes of the impact findings, but statistical significance levels differ in some cases due to less statistical precision.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.3

Impact of the Enhanced Math Program on Student Achievement for the Analysis Sample Without Demographic Characteristics as Model Covariates

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Analysis sample					
SAT 10 math total scaled scores	605.10	602.16	2.94 *	0.07	0.01
Problem solving	606.15	603.70	2.45 *	0.05	0.04
Procedures	605.30	600.74	4.56 *	0.08	0.00
Sample size (total = 1,961)	1,081	880			
Grade subgroups					
Grades 2 and 3					
SAT 10 math total scaled scores	583.23	581.13	2.10	0.05	0.21
Problem solving	584.82	583.77	1.04	0.02	0.54
Procedures	583.55	578.93	4.62	0.09	0.05
Sample size (total = 971)	533	438			
Grades 4 and 5					
SAT 10 math total scaled scores	626.37	622.63	3.74 *	0.08	0.01
Problem solving	626.91	623.08	3.83 *	0.09	0.02
Procedures	626.46	622.02	4.45 *	0.08	0.03
Sample size (total = 990)	548	442			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 math total scaled scores	584.29	581.92	2.37	0.05	0.30
Problem solving	586.30	583.81	2.49	0.06	0.31
Procedures	580.17	578.19	1.99	0.04	0.53
Sample size (total = 467)	239	228			
Students scoring at basic level					
SAT 10 math total scaled scores	600.52	597.29	3.24 *	0.07	0.03
Problem solving	601.74	598.29	3.45 *	0.08	0.04
Procedures	600.63	595.80	4.83 *	0.09	0.02
Sample size (total = 1,055)	612	443			
Students scoring at proficient level					
SAT 10 math total scaled scores	634.67	632.03	2.64	0.06	0.36
Problem solving	634.02	631.71	2.32	0.05	0.43
Procedures	640.08	637.08	3.00	0.06	0.48
Sample size (total = 380)	202	178			

(continued)

Appendix Table F.3 (continued)

SOURCE: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: for the analysis sample, scores range from 389 to 796, 414 to 776, and 413 to 768; for the second- and third-grade subgroup, scores range from 389 to 741, 414 to 719, and 413 to 715; and for the fourth- and fifth-grade subgroup, scores range from 450 to 796, 468 to 776, and 485 to 768.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment and baseline math total scaled score. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 math total scaled score is calculated as a proportion of the standard deviation of the regular program group, which is 44.64 based on the analysis sample and the model controlling for demographic characteristics. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.00. For each subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group.

There are 28 enhanced program group students and 31 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

In summary, the program impacts on the locally administered math test have the same sign as the study-administered SAT 10 impacts but are not statistically significant. The two robustness checks demonstrated that the math impact results reported in Chapter 4 are not affected by the various sample restriction and the alternative model specifications.

For the Reading Sample

Similar to the analysis for the math program, the program impacts on student performance in locally administered reading tests were estimated to compare with those on SAT 10 reading tests.

The locally administered test data were not available for all second-graders in those study sites that start testing students in the third grade. As a result, all second-graders were excluded from this analysis, and the total sample size for the locally administered test analysis is 1,238 for reading.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.4

Impact of the Enhanced Math Program on Student Achievement for the Analysis Sample With a Random Assignment Indicator as the Only Model Covariate

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Analysis sample					
SAT 10 math total scaled scores	605.10	601.99	3.11 *	0.07	0.03
Problem solving	606.15	603.53	2.62	0.06	0.08
Procedures	605.30	600.55	4.76 *	0.09	0.01
Sample size (total = 1,961)	1,081	880			
Grade subgroups					
Grades 2 and 3					
SAT 10 math total scaled scores	583.23	580.61	2.62	0.06	0.23
Problem solving	584.82	583.29	1.53	0.03	0.47
Procedures	583.55	578.31	5.24	0.10	0.07
Sample size (total = 971)	533	438			
Grades 4 and 5					
SAT 10 math total scaled scores	626.37	622.78	3.59	0.08	0.07
Problem solving	626.91	623.22	3.69	0.08	0.08
Procedures	626.46	622.17	4.29	0.08	0.09
Sample size (total = 990)	548	442			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 math total scaled scores	584.29	580.72	3.57	0.08	0.14
Problem solving	586.30	582.48	3.82	0.09	0.15
Procedures	580.17	577.11	3.06	0.06	0.35
Sample size (total = 467)	239	228			
Students scoring at basic level					
SAT 10 math total scaled scores	600.52	597.41	3.11	0.07	0.06
Problem solving	601.74	598.40	3.33	0.07	0.06
Procedures	600.63	595.95	4.68 *	0.09	0.04
Sample size (total = 1,055)	612	443			
Students scoring at proficient level					
SAT 10 math total scaled scores	634.67	632.16	2.50	0.06	0.41
Problem solving	634.02	631.84	2.19	0.05	0.47
Procedures	640.08	637.23	2.85	0.05	0.52
Sample size (total = 380)	202	178			

(continued)

Appendix Table F.4 (continued)

SOURCE: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: for the analysis sample, scores range from 389 to 796, 414 to 776, and 413 to 768; for the second- and third-grade subgroup, scores range from 389 to 741, 414 to 719, and 413 to 715; and for the fourth- and fifth-grade subgroup, scores range from 450 to 796, 468 to 776, and 485 to 768.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment strata. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed distribution of the enhanced program group across random assignment strata as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 math total scaled score is calculated as a proportion of the standard deviation of the regular program group, which is 44.64 based on the analysis sample and the model controlling for demographic characteristics. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.00. For each subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group.

There are 28 enhanced program group students and 31 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

Appendix Table F.5 presents the estimated program impacts on student performance in locally administered tests for reading. Because these test scores were standardized within each study site, all estimated impacts are in effect size.[5] The table also shows the program impact on the study-administered SAT 10 tests for the sample of students whose local test scores were available for comparison purpose. Because second-graders were excluded from the analysis, the tables does not show impact estimates for total scores for the subgroup of second- and third-graders.

For the reading sample, the estimated impact effect size using the local test is –0.01, and that using the study-administered SAT 10 total test score is –0.01 too. None of these estimates are statistically different from zero. Overall, the locally administered tests do not yield qualitatively different findings about the program impact.

In addition, the following checks were conducted to see whether the estimated reading program impacts reported in Chapter 6 are robust:

[5]Appendix E describes the standardization of the test score variable.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.5

Impact of the Enhanced Reading Program on Student Achievement for Grades 3 to 5

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
State test analysis sample					
State test scaled scores	-0.04	-0.03	-0.01	-0.01	0.92
SAT 10 reading total scaled scores	598.74	599.07	-0.33	-0.01	0.77
Sample size (total = 1,238)	720	518			
Grade subgroup					
Grades 4 and 5					
State test scaled scores	-0.05	-0.03	-0.01	-0.01	0.81
SAT 10 reading total scaled scores	605.38	605.71	-0.33	-0.01	0.81
Sample size (total = 830)	486	344			
Prior-achievement subgroups					
Students scoring at below basic level					
State test scaled scores	-0.40	-0.40	0.00	0.01	0.95
SAT 10 reading total scaled scores	585.68	583.33	2.35	0.07	0.16
Sample size (total = 564)	342	222			
Students scoring at basic level					
State test scaled scores	0.20	0.24	-0.04	-0.04	0.56
SAT 10 reading total scaled scores	606.61	608.85	-2.24	-0.06	0.19
Sample size (total = 580)	335	245			
Students scoring at proficient level					
State test scaled scores	0.97	0.70	0.27	0.28	0.26
SAT 10 reading total scaled scores	640.22	632.81	7.41	0.21	0.42
Sample size (total = 90)	41	49			

(continued)

SOURCES: MDRC calculations are from results on state tests administered in the 2005-2006 school year and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: State test data were not available for most second-graders because many of the study sites begin testing students in the third grade, and, as a result, all second-graders are excluded from this analysis. In addition, the state test analysis sample is restricted to those from the full analysis sample for whom a state test score was obtained. The resulting state test analysis sample represents 90 percent of the third- through fifth-graders in the full analysis sample and is used to calculate the SAT 10 and state test findings presented.

Appendix Table F.5 (continued)

Each student's test score was converted into a standardized score because school districts in different states administer different tests. See Appendix E for details.

Based on the SAT 10 national norming sample, reading total scaled scores have the following possible ranges: for the state test analysis sample, scores range from 416 to 787; for the fourth- and fifth-grade subgroup, scores range from 434 to 787.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the state test score is calculated as a proportion of the state test score standard deviation of the regular program group from the state test analysis sample. The estimated impact effect size of the SAT 10 reading total scaled score is calculated as a proportion of the standard deviation of the regular program group from the full analysis sample, which is 35.71. The standard deviation of a SAT 10 national norming sample with the same grade composition as the full analysis sample is 39.05.

There are 2 enhanced program group students and 2 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

- All impacts were reestimated for the sample of all SAT 10 respondents to make sure that no imbalance was created when the full study sample was limited to the analysis sample.

This change in the sample added 76 observations for the reading sample. Appendix Table F.6 presents student achievement impact results for reading using the SAT 10 respondents from the full study sample. The general patterns of the findings do not change at all.

- All impacts were reestimated with a model that has no covariates other than the "block" (random assignment unit) indicators, the treatment status indicator, and prior achievement.

The model used here is the same as Equation (2). As can be seen from Appendix Table F.7, dropping these covariates from the model affected the significance level of the impact estimates but did not affect the magnitudes or the patterns of the impact findings, as one would expect from a randomized experiment.

- All impacts were reestimated with a model that has no covariates other than the "block" (random assignment unit) indicators and the treatment status indicator.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.6

Impact of the Enhanced Reading Program on Student Achievement for the SAT 10 Respondent Sample

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
SAT 10 respondent sample					
SAT 10 reading total scaled scores	587.33	587.75	-0.42	-0.01	0.64
Vocabulary	580.74	580.39	0.35	0.01	0.79
Reading comprehension	588.51	589.03	-0.52	-0.01	0.66
Word study skills (grades 2-4)[a]	586.52	588.25	-1.73	-0.05	0.29
Sample size (total = 1,904)	1,092	812			
Grade subgroups					
Grades 2 and 3					
SAT 10 reading total scaled scores	569.10	569.73	-0.63	-0.02	0.63
Vocabulary	556.70	556.69	0.00	0.00	1.00
Reading comprehension	571.02	571.25	-0.23	-0.01	0.90
Word study skills	579.12	582.20	-3.08	-0.08	0.10
Sample size (total = 944)	544	400			
Grades 4 and 5					
SAT 10 reading total scaled scores	605.43	605.48	-0.05	0.00	0.97
Vocabulary	604.69	603.77	0.92	0.02	0.60
Reading comprehension	605.90	606.36	-0.46	-0.01	0.77
Sample size (total = 960)	548	412			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 reading total scaled scores	576.60	574.94	1.67	0.05	0.23
Vocabulary	567.71	565.92	1.78	0.04	0.38
Reading comprehension	579.00	576.54	2.46	0.06	0.17
Word study skills[a]	571.58	572.40	-0.82	-0.02	0.75
Sample size (total = 770)	456	314			
Students scoring at basic level					
SAT 10 reading total scaled scores	591.74	593.37	-1.63	-0.05	0.24
Vocabulary	586.18	587.34	-1.16	-0.03	0.57
Reading comprehension	591.81	593.62	-1.80	-0.05	0.31
Word study skills[a]	591.38	595.05	-3.67	-0.10	0.12
Sample size (total = 912)	521	391			

(continued))

Appendix Table F.6 (continued)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Students scoring at proficient level					
SAT 10 reading total scaled scores	608.36	610.83	-2.47	-0.07	0.54
Vocabulary	605.58	605.76	-0.17	0.00	0.97
Reading comprehension	609.28	612.77	-3.50	-0.09	0.51
Word study skills[a]	612.02	610.62	1.40	0.04	0.82
Sample size (total = 207)	107	100			

SOURCE: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The SAT 10 respondent sample is composed of all students from the full study sample who have a follow-up SAT 10 reading total score.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: for the SAT 10 respondent sample, scores range from 374 to 787, 439 to 777, 412 to 739, and 410 to 740; for the second- and third-grade subgroup, scores range from 374 to 765, 439 to 743, 412 to 700, and 410 to 727; and for the fourth- and fifth-grade subgroup, scores range from 434 to 787, 478 to 777, and 484 to 739.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 reading total scaled score is calculated as a proportion of the standard deviation of the regular program group from the analysis sample, which is 35.71. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.05. For each subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group from the analysis sample.

There are 8 enhanced program group students and 7 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]The sample consists of second- through fourth-graders only because the spring administration of the test to fifth-graders does not include word study skills.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.7

Impact of the Enhanced Reading Program on Student Achievement for the Analysis Sample Without Demographic Characteristics as Model Covariates

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Analysis sample					
SAT 10 reading total scaled scores	587.42	588.23	-0.81	-0.02	0.39
Vocabulary	580.94	580.91	0.03	0.00	0.98
Reading comprehension	588.72	589.54	-0.82	-0.02	0.50
Word study skills (grades 2-4)[a]	586.39	588.47	-2.08	-0.05	0.21
Sample size (total = 1,828)	1,048	780			
Grade subgroups					
Grades 2 and 3					
SAT 10 reading total scaled scores	569.42	570.61	-1.19	-0.03	0.38
Vocabulary	557.05	558.08	-1.03	-0.02	0.61
Reading comprehension	571.54	571.88	-0.33	-0.01	0.85
Word study skills	579.28	582.85	-3.57	-0.09	0.06
Sample size (total = 912)	524	388			
Grades 4 and 5					
SAT 10 reading total scaled scores	605.43	605.85	-0.42	-0.01	0.75
Vocabulary	604.84	603.69	1.15	0.02	0.52
Reading comprehension	605.89	607.11	-1.23	-0.03	0.44
Sample size (total = 916)	524	392			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 reading total scaled scores	577.48	575.57	1.91	0.05	0.19
Vocabulary	568.88	566.39	2.48	0.05	0.24
Reading comprehension	579.82	577.50	2.33	0.06	0.21
Word study skills[a]	572.06	571.54	0.52	0.01	0.84
Sample size (total = 736)	437	299			
Students scoring at basic level					
SAT 10 reading total scaled scores	591.61	593.71	-2.10	-0.06	0.13
Vocabulary	585.88	587.93	-2.05	-0.04	0.31
Reading comprehension	592.00	593.94	-1.94	-0.05	0.28
Word study skills[a]	591.06	595.26	-4.20	-0.11	0.08
Sample size (total = 877)	501	376			

(continued)

Appendix Table F.7 (continued)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Students scoring at proficient level					
SAT 10 reading total scaled scores	606.71	612.62	-5.91	-0.17	0.13
Vocabulary	604.77	606.11	-1.34	-0.03	0.81
Reading comprehension	607.70	615.35	-7.65	-0.20	0.15
Word study skills[a]	610.92	612.44	-1.52	-0.04	0.81
Sample size (total = 201)	103	98			

SOURCE: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: for the analysis sample, scores range from 374 to 787, 439 to 777, 412 to 739, and 410 to 740; for the second- and third-grade subgroup, scores range from 374 to 765, 439 to 743, 412 to 700, and 410 to 727; and for the fourth- and fifth-grade subgroup, scores range from 434 to 787, 478 to 777, and 484 to 739.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment and baseline reading total scaled score. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the pvalue is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 reading total scaled score is calculated as a proportion of the standard deviation of the regular program group, which is 35.71 based on the analysis sample and the model controlling for demographic characteristics. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.05. For each subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group.

There are 7 enhanced program group students and 7 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]The sample consists of second- through fourth-graders only because the spring administration of the test to fifth-graders does not include word study skills.

The model used here is the same as Equation (3) and, as can be seen from Appendix Table F.8, dropping covariates from the model and controlling only for the randomization strata affected the precision of the impact estimates as well as the magnitudes and the patterns of the impact findings. This is because there were significant differences between the enhanced reading program group and the regular reading program group at baseline, which are no longer being controlled for in this model.

- The baseline reading scores of the regular program group and the enhanced program group were statistically different from each other in a number of reading sample blocks. After restricting the sample to those blocks where the baseline scores were similar, all impacts were then reestimated. (This restricted sample is 87 percent of the analysis sample.)

Even with randomization there may be differences in baseline characteristics between the enhanced and regular program groups that are attributable to chance. Recall from Chapter 5 that there were statistically significant differences between the enhanced reading group and the regular reading program group at baseline. As a robustness check, block-by-block baseline differences in test scores were checked, and 12 blocks with the biggest baseline test score differences were excluded from the sample.[6] The remaining sample achieved balance between the enhanced program group and the regular program group at baseline. All impacts were reestimated using this restricted sample. This series of tests yields very similar impact estimates for the reading program sample (see Appendix Table F.9). These results show that controlling for the baseline characteristics as covariates in the impact model sufficiently eliminated the observed baseline differences between the enhanced program group and the regular program group.

In general, the reading impact results reported in Chapter 6 of this report are not affected by the various sample restriction and alternative model specifications.[7]

[6]A block was dropped if the baseline total reading test score difference between the enhanced program and regular program groups within that block was bigger than the overall difference between these groups by more than 1.75 standard deviations.

[7]In addition, 15 percent of parents reported on applications that the primary language spoken at home is Spanish. Since the classes were taught in English, one concern was that students who primarily do not speak English were not able to benefit from the program. Impacts were reestimated for those students who did not indicate that Spanish is the primary language spoken at home, and the results did not change.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.8

mpact of the Enhanced Reading Program on Student Achievement for the Analysi Sample With a Random Assignment Indicator as the Only Model Covariate

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Analysis sample					
SAT 10 reading total scaled scores	587.42	590.68	-3.26 *	-0.09	0.01
Vocabulary	580.94	583.98	-3.04	-0.07	0.08
Reading comprehension	588.72	591.84	-3.12 *	-0.08	0.03
Word study skills (grades 2-4)[a]	586.39	591.30	-4.91 *	-0.13	0.01
Sample size (total = 1,828)	1,048	780			
Grade subgroups					
Grades 2 and 3					
SAT 10 reading total scaled scores	569.42	574.02	-4.60 *	-0.13	0.02
Vocabulary	557.05	562.47	-5.43 *	-0.12	0.04
Reading comprehension	571.54	575.30	-3.76	-0.10	0.10
Word study skills	579.28	585.67	-6.39 *	-0.17	0.00
Sample size (total = 912)	524	388			
Grades 4 and 5					
SAT 10 reading total scaled scores	605.43	607.34	-1.91	-0.05	0.26
Vocabulary	604.84	605.48	-0.65	-0.01	0.77
Reading comprehension	605.89	608.38	-2.49	-0.06	0.18
Sample size (total = 916)	524	392			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 reading total scaled scores	577.48	576.35	1.13	0.03	0.46
Vocabulary	568.88	567.49	1.39	0.03	0.53
Reading comprehension	579.82	578.15	1.67	0.04	0.38
Word study skills[a]	572.06	571.94	0.12	0.00	0.96
Sample size (total = 736)	437	299			
Students scoring at basic level					
SAT 10 reading total scaled scores	591.61	594.75	-3.14 *	-0.09	0.04
Vocabulary	585.88	589.24	-3.35	-0.07	0.12
Reading comprehension	592.00	594.87	-2.87	-0.07	0.12
Word study skills[a]	591.06	596.46	-5.40 *	-0.14	0.03
Sample size (total = 877)	501	376			

(continued)

Appendix Table F.8 (continued)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Students scoring at proficient level					
SAT 10 reading total scaled scores	606.71	613.07	-6.36	-0.18	0.12
Vocabulary	604.77	606.85	-2.09	-0.05	0.73
Reading comprehension	607.70	615.75	-8.05	-0.21	0.14
Word study skills[a]	610.92	612.35	-1.44	-0.04	0.82
Sample size (total = 201)	103	98			

SOURCE: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: for the analysis sample, scores range from 374 to 787, 439 to 777, 412 to 739, and 410 to 740; for the second- and third-grade subgroup, scores range from 374 to 765, 439 to 743, 412 to 700, and 410 to 727; and for the fourth- and fifth-grade subgroup, scores range from 434 to 787, 478 to 777, and 484 to 739.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment strata. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed distribution of the enhanced program group across random assignment strata as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the pvalue is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 reading total scaled score is calculated as a proportion of the standard deviation of the regular program group, which is 35.71 based on the analysis sample and the model controlling for demographic characteristics. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.05. For each subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group.

There are 7 enhanced program group students and 7 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]The sample consists of second- through fourth-graders only because the spring administration of the test to fifth-graders does not include word study skills.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.9

Impact of the Enhanced Reading Program on Student Achievement When Twelve Random Assignment Blocks Are Excluded from the Analysis Sample

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Restricted analysis sample					
SAT 10 reading total scaled scores	588.03	589.27	-1.24	-0.03	0.21
Vocabulary	581.86	582.43	-0.57	-0.01	0.69
Reading comprehension	589.36	590.45	-1.09	-0.03	0.40
Word study skills (grades 2-4)[a]	585.17	587.89	-2.73	-0.07	0.13
Sample size (total = 1,588)	909	679			
Grade subgroups					
Grades 2 and 3					
SAT 10 reading total scaled scores	569.11	571.34	-2.23	-0.06	0.13
Vocabulary	557.30	558.88	-1.58	-0.03	0.47
Reading comprehension	571.13	572.72	-1.58	-0.04	0.42
Word study skills	578.33	583.18	-4.85 *	-0.13	0.02
Sample size (total = 756)	436	320			
Grades 4 and 5					
SAT 10 reading total scaled scores	605.47	605.79	-0.32	-0.01	0.81
Vocabulary	604.49	604.09	0.40	0.01	0.83
Reading comprehension	606.16	606.72	-0.55	-0.01	0.74
Sample size (total = 832)	473	359			
Prior-achievement subgroups					
Students scoring at below basic level					
SAT 10 reading total scaled scores	578.28	576.37	1.91	0.05	0.21
Vocabulary	569.69	567.62	2.07	0.05	0.36
Reading comprehension	580.84	577.99	2.85	0.07	0.14
Word study skills[a]	571.84	572.09	-0.25	-0.01	0.93
Sample size (total = 646)	376	270			
Students scoring at basic level					
SAT 10 reading total scaled scores	592.05	594.27	-2.22	-0.06	0.14
Vocabulary	587.00	588.90	-1.90	-0.04	0.38
Reading comprehension	592.31	594.41	-2.10	-0.05	0.27
Word study skills[a]	589.39	593.81	-4.42	-0.12	0.09
Sample size (total = 763)	435	328			

(continued)

Appendix Table F.9 (continued)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Students scoring at proficient level					
SAT 10 reading total scaled scores	606.23	612.57	-6.34	-0.18	0.10
Vocabulary	603.81	608.27	-4.46	-0.10	0.44
Reading comprehension	607.34	615.85	-8.51	-0.22	0.11
Word study skills[a]	608.96	607.33	1.63	0.04	0.80
Sample size (total = 166)	91	75			

SOURCE: MDRC calculations are from follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The restricted analysis sample excludes 12 random assignment blocks (grades within centers) because, for each one, the baseline total reading test score difference between the enhanced program and regular program groups is bigger than the overall difference between these groups by more than 1.75 standard deviations.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: for the restricted analysis sample, scores range from 374 to 787, 439 to 777, 412 to 739, and 410 to 740; for the second- and third-grade subgroup, scores range from 374 to 765, 439 to 743, 412 to 700, and 410 to 727; and for the fourth- and fifth-grade subgroup, scores range from 434 to 787, 478 to 777, and 484 to 739.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size of the SAT 10 reading total scaled score is calculated as a proportion of the standard deviation of the regular program group, which is 35.71 based on the full analysis sample. The standard deviation of a SAT 10 national norming sample with the same grade composition as the study sample is 39.05. For each subtest, the estimated impact effect size is calculated as a proportion of the standard deviation of the regular program group.

There are 7 enhanced program group students and 6 regular program group students who performed at the advanced level on the baseline SAT 10; they are excluded from the prior-achievement subgroup analysis.

[a]The sample consists of second- through fourth-graders only because the spring administration of the test to fifth-graders does not include word study skills.

Appendix G

Exploratory Analysis

This appendix lays out the strategy used to investigate possible associations between impacts and characteristics of both the schools housing the after-school program and the implementation of the enhanced after-school program. To explore the interface between the enhanced after-school program strategy and these features, an exploratory correlational analysis was conducted. Because students were not randomly assigned to programs with different school characteristics, this analysis is correlational rather than experimental. As such, the results should not be viewed definitively as causal; the associations that are found could be causal or could purely (or partly) reflect selection bias. Thus, these analyses should be viewed as hypothesis generating, not summative.

In this appendix, the correlational methodology is presented, as is a detailed description of the school characteristic measures used in the analysis.

Analytic Approach

Apart from understanding how impacts may vary with various student characteristics, decision makers may also want to know whether this intervention worked better in particular types of schools or in after-school programs implemented in a particular way. Thus, for the sample of math and reading programs, this part of the analysis explores whether school context characteristics or factors of program implementation were associated with impacts.

Data were collected for the following school characteristics, and their correlation with impacts were examined: the instructional approach of the school-day curricula (available for the math sample but not for the reading sample), how much time is spent in the regular school day on instruction in math or reading, whether the school meet its Adequate Yearly Progress (AYP) goals, what proportion of students in the school receive free or reduced-price lunch, and what is the in-school student-to-teacher ratio. For example, students who are struggling during the school day may benefit from an alternative instructional approach after school. Or additional time in math or reading may have a greater benefit for students who have less time on those topics during the school day. To examine these characteristics, centers were categorized by their regular-school-day curricula (which produced three groups — one with curricula similar to that used after school and two others)[1] as well as categorized by the time spent in the regular school

[1]Note that, for the reading sample, this information is not available.

day on instruction in math or reading (more than 60 minutes per day or less for the math sample, more than 90 minutes per day or less for the reading sample).[2]

Additionally, two factors of program implementation were examined: (1) Did one or more of the instructors teaching the enhanced after-school program leave during the school year? (2) How many days was the enhanced after-school program offered?

The analysis — similar to the approach taken in Bloom, Hill, and Riccio (2001) — examines how the variation of both math and reading impacts is associated with school characteristics across centers.

In particular, this analysis used a two-level hierarchical linear model to estimate how the size of the impact is related to school context inputs. The unit of analysis for Level 1 is the individual student. The unit in Level 2 is the study center. Equations (1) and (2) describe this analytical approach. In this random coefficient model, the size of the center-level impact, β_m, is allowed to vary with the school and the after-school setting experienced by the students.

Level 1

$$Y_{im} = \gamma_1 Y_{-1im} + \sum_{100} \alpha_m Block_{im} + \beta_m T_{im} + \sum_j \delta_j X_{imj} + \varepsilon_{im} \qquad (1)$$

where:

Y_{-1im} = the pretest score for student i in block m before random assignment.[3]

$\sum_j X_{imj}$ = student-level characteristic j for student i from center/block m.

$Block_{im}$ = dummy variable equal to 1 if student i was a member of center/block m, otherwise it is zero.

[2]School administrators were asked how many minutes teachers spend a day teaching math or reading to their students. The responses were not a precise number of minutes, so a continuous measure of minutes is not used. Instead, groups were created around the most common response. For math, 24 percent of schools offer 50 to 60 minutes; 32 percent offer 60 minutes; 28 percent offer 60 to 90 minutes; and the remaining 16 percent offer 90 minutes or more. Thus, for math, the natural split for this subgroup is those offering 60 minutes or less of school-day math instruction and those offering more than 60 minutes. For reading, 20 percent offer, on average, less than 90 minutes (in some schools the amount of time varies by grade); about half (52 percent) offer 90 minutes; and the remaining 28 percent offer more than 90 minutes. Thus, for reading, the natural split is those offering 90 minutes or less and those offering more than 90 minutes.

[3]Pretest scores are scaled scores from the SAT10 tests in reading and math administered in the fall of 2005, before the start of the after-school program. Total scores for each subject are used in the analysis of respective samples.

T_{im} = dummy variable equal to 1 if student *i* was assigned to be part of the experimental group in center/block *m,* otherwise it is zero.

ε_{im} = a student-level random error, assumed to be independently and identically distributed.

Level 2

$$\beta_m = \tau_0 + \tau_1 Group1_m + \tau_2 Group2_m + \tau_3 PERIODlong_m + \tau_4 AYP_m + \tau_5 \%FRL_m + \tau_6 S/T_m + \tau_7 TLEFT_m + \tau_8 TOTDYS_m + \mu_m \quad (2)$$

where:

$Group1_m$ = a dummy equal to 1 if, for the centers implementing Mathletics, the school-day curricula are unit based, which are longer than chapters, and are investigation driven with comparatively fewer practice problems and involving interconnected subproblems, and 0 otherwise.

$Group2_m$ = a dummy equal to 1 if, for the centers implementing Mathletics, the school-day curriculum employs a direct instruction approach organized by lessons with spiraled curriculum, and 0 otherwise.[4]

$PERIODlong_m$ = a dummy equal to 1 if the school-day period in the relevant subject is more than 60 minutes for math or 90 minutes for reading, and 0 otherwise.

AYP_m = a dummy equal to 1 if the school met its AYP requirements in 2005-2006, and 0 otherwise.

$\%FRL_m$ = the percentage of students in school *m* who receive free or reduced-priced lunch centered on the grand mean of all schools in the sample.

S/T_m = a dummy equal to 1 if the student-to-teacher ratio in school *m* is greater than the planned student-to-teacher ratio in the after-school program (13:1 for math).

$TLEFT_m$ = a dummy equal to 1 if one of the instructors teaching the enhanced after-school program left the program during the school year, and 0 otherwise.

[4]In three centers, second-graders used different type of curriculum than the one used in other grades. For these centers, the Group1 and Group2 variables are allowed to vary within school by grade. For example, second-graders within the school may identify with Group2 while the other grades identify with Group1.

$TOTDYS_m$ = the number of days that the enhanced after-school program was offered, centered on the grand mean of all centers in the sample.

μ_m = a center/block-level random error, assumed to be independently and identically distributed.

τ_k (where k = 1, 2, …, 10) is the association between the intervention's impact and school characteristic variable *k,* controlling for other characteristics included in Equation (2). For example, τ_1 is the association of the intervention's impact with having a school-day math curriculum that is unit based, controlling for other characteristics included in Equation (2); and τ_3 is the association of the intervention's impact with having longer periods in school on math or reading, controlling for the other characteristics. If τ_3 is statistically significant and positive, it means that having longer periods in school on math or reading is associated with a bigger program impact.

Appendix H

Service Contrast Subgroups

This appendix shows findings for the *difference* between the after-school academic services received by the enhanced program group and those received by the regular, "business as usual" program group, for subgroups based on student grade level and baseline achievement. The tables present differences in attendance in the after-school program, in hours of instruction received, and in special academic support received from other sources — during the regular school day and outside school.

Appendix Tables H.1 and H.2 present differences for the math program grade-level subgroups and the prior-achievement subgroups, respectively. The difference in hours of academic instruction in math for the second- and third-grade subgroup is 49 hours; for the fourth- and fifth-grade subgroup, it is 48 hours.[1] The difference for the "below basic" and "basic" achievement-level subgroups is 49 hours; for the "proficient" subgroup, it is 46 hours. All these differences are statistically significant.

Findings for the reading program subgroups are presented in Appendix Tables H.3 and H.4. The difference in hours of academic instruction in reading is 51 hours for the second- and third-grade subgroup, and it is 46 hours for fourth- and fifth-graders. The difference for students in the "below basic" achievement level is 43 hours; at the "basic" level, it is 51 hours; and at the "proficient" level, it is 53 hours. All these differences are statistically significant.

Overall, for both measures of attendance in the after-school program, in all but one case, the findings for reading and math subgroups based on student grade level and baseline achievement are similar to those found for the analysis sample, with the same pattern of somewhat greater attendance among the enhanced program group.[2]

[1] In addition, tests found that there are no significantly different patterns of service contrast by grade level within the younger and older subgroups.

[2] One subgroup — the reading students scoring at the "proficient level" — has a negative impact estimate of –1.1 (p-value = 0.81) for number of days attended.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.1

Attendance of Students in the Math Analysis Sample, by Grade Subgroup

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Grades 2 and 3					
Attendance in after-school program[a]					
Number of days attended	74.65	62.47	12.18 *	0.37	0.00
Total hours of math instruction received[b]	58.07	9.03	49.04 *	2.78	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	35.65	24.10	11.54 *	0.29	0.00
Number of days per week[d]	1.21	0.69	0.52 *	0.37	0.00
Regular school day[e]					
Students receiving special support (%)	2.21	2.19	0.02	0.05	0.40
Minutes per week of individualized help	42.60	43.56	-0.96	-0.01	0.79
Sample size (total = 971)	533	438			
Grades 4 and 5					
Attendance in after-school program[a]					
Number of days attended	72.29	60.11	12.18 *	0.37	0.00
Total hours of math instruction received[b]	56.30	8.22	48.08 *	2.73	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	21.90	17.75	4.14	0.10	0.07
Number of days per week[d]	0.73	0.50	0.23 *	0.17	0.01
Regular school day[e]					
Students receiving special support (%)	2.27	2.30	-0.04	-0.08	0.15
Minutes per week of individualized help	56.94	55.13	1.81	0.03	0.89
Sample size (total = 990)	548	442			

(continued)

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs attendance records, student survey responses, and regular-school-day teacher survey responses.

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Appendix Table H.1 (continued)

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size for each measure is calculated as a proportion of the standard deviation of the regular program group.

[a]Attendance in the after-school program is based on the days the enhanced program operated.

[b]Students in the enhanced classes received 45 minutes of instruction (and 60 minutes in one site that met only three days a week) on the days they were present. Total hours is calculated for these students by multiplying each student's total days of attendance by 45 (or 60 in the one site).

Students in the regular program group were not supposed to receive any structured instruction. However, some regular program staff indicated on the survey that they provide structured academic instruction. Total hours is calculated for these students by multiplying the total number of days attended by 45, then by the proportion of regular program staff within the center who reported providing structured instruction. If no regular program staff in a center indicated that they provide structured instruction, then total hours for these students in that center is zero. If no regular program staff in a center answered this question, this calculation could not be performed for these students. Calculated as such, the sample sizes for the regular program group are 379 for the second- and third-grade subgroup and 391 for the fourth- and fifth-grade subgroup.

[c]This information comes from student survey responses to questions for each day of the week that ask, "Do you go somewhere else for a math class or to be tutored in math?" These calculations are based on a smaller sample than the reported analysis sample by the number of students who did not complete a survey. For the second- and third-grade subgroup, the sample size is 533 for the enhanced program group and 437 for the regular program group. For the fourth- and fifth-grade subgroup, the sample size is 548 for the enhanced program group and 442 for the regular program group.

[d]Students who responded that they do not receive math support from other out-of-school sources are included in these averages.

[e]This information comes from regular-school-day teacher survey responses. "Special support" refers to special support in math during the school day (that is, pull-out tutoring, remedial math assistance, assigned to a computer-assisted lab, and so on). "Individualized help" refers to individual help from the teacher or an aide with a task or answering a question. Teachers who responded that they did not provide support may or may not have responded that they provided minutes of individualized help. Thus, average minutes includes responses for all students, not just those who received special support.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.2

Attendance of Students in the Math Analysis Sample, by Prior-Achievement Subgroup

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Students scoring at below basic level					
Attendance in after-school program[a]					
Number of days attended	66.80	55.66	11.14 *	0.34	0.00
Total hours of math instruction received[b]	52.63	4.03	48.60 *	2.76	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	41.42	34.47	6.95	0.17	0.10
Number of days per week[d]	1.50	1.08	0.42 *	0.30	0.01
Regular school day[e]					
Students receiving special support (%)	2.37	2.43	-0.06	-0.13	0.19
Minutes per week of individualized help	58.03	68.20	-10.17	-0.15	0.08
Sample size (total = 467)	239	228			
Students scoring at basic level					
Attendance in after-school program[a]					
Number of days attended	74.50	60.04	14.47 *	0.44	0.00
Total hours of math instruction received[b]	57.91	8.55	49.36 *	2.80	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	27.45	20.46	6.99 *	0.17	0.00
Number of days per week[d]	0.84	0.57	0.27 *	0.20	0.00
Regular school day[e]					
Students receiving special support (%)	2.24	2.23	0.01	0.02	0.72
Minutes per week of individualized help	53.52	52.01	1.51	0.02	0.90
Sample size (total = 1,055)	612	443			
Students scoring at proficient level					
Attendance in after-school program[a]					
Number of days attended	77.83	70.79	7.04 *	0.22	0.01
Total hours of math instruction received[b]	60.25	14.64	45.61 *	2.59	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	18.81	12.09	6.72	0.17	0.06
Number of days per week[d]	0.75	0.37	0.38 *	0.27	0.01

(continued)

Appendix Table H.2 (continued)

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Regular school day[e]					
Students receiving special support (%)	2.10	2.08	0.02	0.05	0.55
Minutes per week of individualized help	33.44	30.75	2.69	0.04	0.63
Sample size (total = 380)	202	178			

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs attendance records, student survey responses, and regular-school-day teacher survey responses.

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size for each measure is calculated as a proportion of the standard deviation of the regular program group.

[a]Attendance in the after-school program is based on the days the enhanced program operated.

[b]Students in the enhanced classes received 45 minutes of instruction (and 60 minutes in one site that met only three days a week) on the days they were present. Total hours is calculated for these students by multiplying each student's total days of attendance by 45 (or 60 in the one site).

Students in the regular program group were not supposed to receive any structured instruction. However, some regular program staff indicated on the survey that they provide structured academic instruction. Total hours is calculated for these students by multiplying the total number of days attended by 45, then by the proportion of regular program staff within the center who reported providing structured instruction. If no regular program staff in a center indicated that they provide structured instruction, then total hours for these students in that center is zero. If no regular program staff in a center answered this question, this calculation could not be performed for these students. Calculated as such, the sample sizes for the regular program group are 181 for the group of students scoring at the below basic level, 397 for the group of students scoring at the basic level, and 164 for the group of students scoring at the proficient level.

[c]This information comes from student survey responses to questions for each day of the week that ask, "Do you go somewhere else for a math class or to be tutored in math?" These calculations are based on a smaller sample than the reported analysis sample by the number of students who did not complete a survey. For the group of students scoring at the below basic level, the sample size is 239 for the enhanced program group and 227 for the regular program group. For the group of students scoring at the basic level, the sample size is 612 for the enhanced program group and 443 for the regular program group. For the group of students scoring at the proficient level, the sample size is 202 for the enhanced program group and 178 for the regular program group.

[d]Students who responded that they do not receive math support from other out-of-school sources are included in these averages.

[e]This information comes from regular-school-day teacher survey responses. "Special support" refers to special support in math during the school day (that is, pull-out tutoring, remedial math assistance, assigned to a computer-assisted lab, and so on). "Individualized help" refers to individual help from the teacher or an aide with a task or answering a question. Teachers who responded that they did not provide support may or may not have responded that they provided minutes of individualized help. Thus, average minutes includes responses for all students, not just those who received special support.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.3

Attendance of Students in the Reading Analysis Sample, by Grade Subgroup

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Grades 2 and 3					
Attendance in after-school program[a]					
Number of days attended	73.34	64.61	8.73 *	0.25	0.00
Total hours of reading instruction received[b]	57.14	6.04	51.10 *	2.89	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	45.68	36.39	9.29 *	0.20	0.00
Number of days per week[d]	1.37	0.95	0.42 *	0.29	0.00
Regular school day[e]					
Students receiving special support (%)	2.46	2.41	0.05	0.09	0.10
Minutes per week of individualized help	73.03	69.55	3.47	0.02	0.58
Sample size (total = 912)	524	388			
Grades 4 and 5					
Attendance in after-school program[a]					
Number of days attended	67.33	62.73	4.60 *	0.13	0.01
Total hours of reading instruction received[b]	52.86	7.21	45.65 *	2.58	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	31.59	26.24	5.35	0.12	0.06
Number of days per week[d]	0.88	0.61	0.28 *	0.19	0.00
Regular school day[e]					
Students receiving special support (%)	2.36	2.37	-0.01	-0.02	0.66
Minutes per week of individualized help	100.37	98.73	1.64	0.01	0.83
Sample size (total = 916)	524	392			

(continued)

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs attendance records, student survey responses, and regular-school-day teacher survey responses.

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Appendix Table H.3 (continued)

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size for each measure is calculated as a proportion of the standard deviation of the regular program group.

[a]Attendance in the after-school program is based on the days the enhanced program operated.

[b]Students in the enhanced classes received 45 minutes of instruction (and 60 minutes in one site that met only three days a week) on the days they were present. Total hours is calculated for these students by multiplying each student's total days of attendance by 45 (or 60 in the one site).

Students in the regular program group were not supposed to receive any structured instruction. However, some regular program staff indicated on the survey that they provide structured academic instruction. Total hours is calculated for these students by multiplying the total number of days attended by 45, then by the proportion of regular program staff within the center who reported providing structured instruction. If no regular program staff in a center indicated that they provide structured instruction, then total hours for these students in that center is zero. If no regular program staff in a center answered this question, this calculation could not be performed for these students. Calculated as such, the sample size for the regular program group is 299 for the second- and third-grade subgroup and 304 for the fourth- and fifth-grade subgroup.

[c]This information comes from student survey responses to questions for each day of the week that ask, "Do you go somewhere else for a reading class or to be tutored in reading?" These calculations are based on a smaller sample than the reported analysis sample by the number of students who did not complete a survey. For the second- and third-grade subgroup, the sample size is 521 for the enhanced program group and 386 for the regular program group. For the fourth- and fifth-grade subgroup, the sample size is 516 for the enhanced program group and 386 for the regular program group.

[d]Students who responded that they do not receive reading support from other out-of-school sources are included in these averages.

[e]This information comes from regular-school-day teacher survey responses. "Special support" refers to special support in reading during the school day (that is, pull-out tutoring, Reading Recovery, assigned to a computer-assisted lab, and so on). "Individualized help" refers to individual help from the teacher or an aide with a task or answering a question. Teachers who responded that they did not provide support may or may not have responded that they provided minutes of individualized help. Thus, average minutes includes responses for all students, not just those who received special support.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.4

ttendance of Students in the Reading Analysis Sample, by Prior-Achievement Subgrou

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Students scoring at below basic level					
Attendance in after-school program[a]					
Number of days attended	63.39	60.14	3.25	0.09	0.12
Total hours of reading instruction received[b]	50.03	7.25	42.78 *	2.42	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	40.05	38.85	1.20	0.03	0.74
Number of days per week[d]	1.16	1.02	0.13	0.09	0.27
Regular school day[e]					
Students receiving special support (%)	2.51	2.53	-0.02	-0.04	0.62
Minutes per week of individualized help	111.29	114.49	-3.20	-0.02	0.75
Sample size (total = 736)	437	299			
Students scoring at basic level					
Attendance in after-school program[a]					
Number of days attended	75.30	65.60	9.71 *	0.28	0.00
Total hours of reading instruction received[b]	58.63	7.34	51.29 *	2.90	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	38.80	27.42	11.38 *	0.25	0.00
Number of days per week[d]	1.14	0.68	0.46 *	0.32	0.00
Regular school day[e]					
Students receiving special support (%)	2.36	2.33	0.03	0.06	0.30
Minutes per week of individualized help	72.52	69.19	3.33	0.02	0.59
Sample size (total = 877)	501	376			
Students scoring at proficient level					
Attendance in after-school program[a]					
Number of days attended	75.09	76.19	-1.10	-0.03	0.81
Total hours of reading instruction received[b]	58.15	5.36	52.79 *	2.98	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	32.04	18.91	13.13	0.28	0.12
Number of days per week[d]	0.94	0.46	0.48	0.33	0.07

(continued)

Appendix Table H.4 (continued)

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Regular school day[e]					
Students receiving special support (%)	2.23	2.20	0.03	0.07	0.62
Minutes per week of individualized help	54.24	52.14	2.10	0.01	0.85
Sample size (total = 201)	103	98			

SOURCES: MDRC calculations are from the Evaluation of Academic Instruction in After-School Programs attendance records, student survey responses, and regular-school-day teacher survey responses.

NOTES: The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of random assignment, baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The values in column 1 (labeled "Enhanced Program") are the observed mean for the members randomly assigned to the enhanced program group. The regular program group values in column 2 are the regression-adjusted means using the observed mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each impact estimate. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The estimated impact effect size for each measure is calculated as a proportion of the standard deviation of the regular program group.

[a]Attendance in the after-school program is based on the days the enhanced program operated.

[b]Students in the enhanced classes received 45 minutes of instruction (and 60 minutes in one site that met only three days a week) on the days they were present. Total hours is calculated for these students by multiplying each student's total days of attendance by 45 (or 60 in the one site).

Students in the regular program group were not supposed to receive any structured instruction. However, some regular program staff indicated on the survey that they provide structured academic instruction. Total hours is calculated for these students by multiplying the total number of days attended by 45, then by the proportion of regular program staff within the center who reported providing structured instruction. If no regular program staff in a center indicated that they provide structured instruction, then total hours for these students in that center is zero. If no regular program staff in a center answered this question, this calculation could not be performed for these students. Calculated as such, the sample size for the regular program group is 242 for the group of students scoring at the below basic level, 285 for the group of students scoring at the basic level, and 73 for the group of students scoring at the proficient level.

[c]This information comes from student survey responses to questions for each day of the week that ask, "Do you go somewhere else for a reading class or to be tutored in reading?" These calculations are based on a smaller sample than the reported analysis sample by the number of students who did not complete a survey. For the group of students scoring at the below basic level, the sample size is 427 for the enhanced program group and 297 for the regular program group. For the group of students scoring at the basic level, the sample size is 500 for the enhanced program group and 371 for the regular program group. For the group of students scoring at the proficient level, the sample size is 103 for the enhanced program group and 97 for the regular program group.

[d]Students who responded that they do not receive reading support from other out-of-school sources are included in these averages.

[e]This information comes from regular-school-day teacher survey responses. "Special support" refers to special support in reading during the school day (that is, pull-out tutoring, Reading Recovery, assigned to a computer-assisted lab, and so on). "Individualized help" refers to individual help from the teacher or an aide with a task or answering a question. Teachers who responded that they did not provide support may or may not have responded that they provided minutes of individualized help. Thus, average minutes includes responses for all students, not just those who received special support.

References

Arbreton, A. J., Goldsmith, J., and Sheldon, J. (2005). *Launching Literacy in After-School Programs: Early Lessons from the CORAL Initiative.* Philadelphia: Public/Private Ventures.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A New and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B*(57): 1289-1300.

Bloom, H. S., Hill, C. J., and Riccio, J. (2001). *Modeling the Performance of Welfare-to-Work Programs: The Effects of Program Management and Services, Economic Environment, and Client Characteristics.* MDRC Working Papers on Research Methodology. New York: MDRC. Web site: www.mdrc.org.

Bodilly, S., and Beckett, M. K. (2005). *Making Out-of-School Time Matter: Evidence for an Action Agenda.* Santa Monica, CA: RAND Corporation.

Britsch, B., Martin, N., Stuczynski, A., Tomala, B., and Tucci, P. (2005). *Literacy in Afterschool Programs: Literature Review.* Portland, OR: Northwest Regional Educational Laboratory.

DeVellis, R. F. (2003). *Scale Development: Theory and Applications.* In Applied Social Research Methods Series (vol. 26). Thousand Oaks, CA: Sage Publications.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). (2007a). *DIBELS Data System.* Retrieved September 7, 2007, from http://dibels.uoregon.edu/index.php.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). (2007b). *Oral Reading Fluency, Retell Fluency.*
Retrieved September 7 , 2007, from http://dibels.uoregon.edu/measures/orf.php.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). (2007c). *Nonsense Word Fluency.* Retrieved September 7, 2007, from http://dibels.uoregon.edu/measures/nwf.php.

Dynarski, M., James-Burdumy, S., Moore, M., Rosenberg, L., Deke, J., and Mansfield, W. (2004). *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: New Findings*. U.S. Department of Education, National Center for Education Evaluation and Regional Assistance. Washington, DC: U.S. Government Printing Office.

Dynarski, M., Moore, M., Mullens, J., Gleason, P., James-Burdumy, S., Rosenberg, L., Pistorino, C., Silva, T., Deke, J., Mansfield, W., Heaviside, S., and Levy, D. (2003). *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program, First-Year Findings.* Report submitted to the U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc.

Ferrandino, Vincent L. (2007). *A New Day for Learning.* Denver, CO: Time, Learning, and Afterschool Task Force.

Grossman, J. B., Campbell, M., and Raley, R. (2007). *Quality After-School Time: What Instructors Can Do to Enhance Learning.* Philadelphia: Public/Private Ventures.

Harcourt Assessment. *Stanford Achievement Test Series, Tenth Edition – Abbreviated Battery, Areas of Assessment.* (n.d.) San Antonio, TX: Harcourt Assessment, Inc. Retrieved September 9, 2007, from http://harcourtassessment.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=SAT10A&Mode=summary&Leaf=SAT10A_2.

Harcourt Assessment. *Stanford Achievement Test Series, Tenth Edition – Spring Multilevel Norms Book.* (2003). San Antonio, TX: Harcourt Assessment, Inc.

Harcourt Assessment. *Stanford Achievement Test Series, Tenth Edition – Technical Data Report.* (2004). San Antonio, TX: Harcourt Assessment, Inc.

Kane, T. J. (2004, January 16). *The Impact of After-School Programs: Interpreting the Results of Four Recent Evaluations.* William T. Grant Foundation Working Paper. New York: William T. Grant Foundation.

Miller, B. M. (2003). *Critical Hours: Afterschool Programs and Educational Success*. Brookline, MA: Miller Midzik Research Associates.

Miller, B. (2006). *Pathways to Success: What Counts in After School Programs.* Wellesley, MA: National Institute for Out-of School Time.

Naftzger, N., Kaufman, S., Margolin, J., and Ali, A. (2006, July). *An Overview of the 21st CCLC Program: 2004-05* (pp. 34-40). Naperville, IL: Learning Point Associates.

National Center for Education Statistics, Common Core of Data (CCD). (n.d.) *Public Elementary/Secondary School Universe Survey Data: 2005-06.* Retrieved July 26, 2007, from http://nces.ed.gov/ccd/.

National Center for Education Statistics, Common Core of Data (CCD). (n.d.) *Identification of Rural Locales.* Retrieved August 9, 2007, from http://nces.ed.gov/ccd/Rural_Locales.asp.

National Mathematics Advisory Panel. (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel.* U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

National Reading Panel. (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Literature on Reading and Its Implications for Reading Instruction.* Bethesda, MD: National Institute of Child Health and Human Development.

National Research Council Institute of Medicine. (2004). *Engaging Schools: Fostering High School Students' Motivation to Learn.* Washington, DC: National Academic Press.

Salvia, J., and Ysseldyke, J. E. (2001). *Assessment* (8th ed.). Boston: Houghton Mifflin Company.

Scantron Corporation. (2005). *Above and Beyond: Applying Adaptive Technology to Diagnose Student Performance and Progress.* Web site: http://www.scantron.com/downloads/Performance_Series_White_Paper.pdf.

Shepard, L. A. The Role of Classroom Assessment in Teaching and Learning. (2001). In Virginia Richardson (Ed.), *The Handbook of Research on Teaching* (8th ed., pp. 1066-1101). Washington DC: American Education Research Association.

Slavin, R. E., and Madden, N. A. (1999). *Disseminating Success for All: Lessons for Policy and Practice.* Center for Research on the Education of Students Placed At Risk (CRESPAR) Report No. 30. Web site: www.csos.jhu.edu/crespar/.

Success for All. (2004). *Adventure Island Reading Intervention for Grades 1 to 5: Targeted Reading Intervention for Struggling Students.* Marketing brochure prepared by the Success for All Foundation to describe the program.

U.S. Department of Education, 21st Century Community Learning Centers. (n.d.) *Program Description*.
Retrieved September 7, 2007, from http://www.ed.gov/programs/21stcclc/index.html.