NCEE 2009-4077

U. S. DEPARTMENT OF EDUCATION

The Evaluation of Enhanced Academic Instruction in After-School Programs

Final Report

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

The Evaluation of Enhanced Academic Instruction in After-School Programs

Final Report

September 2009

Authors:

Alison Rebeck Black
Marie-Andrée Somers
Fred Doolittle
Rebecca Unterman
MDRC

Jean Baldwin Grossman
Public/Private Ventures

Project Officer:

Elizabeth Warner
Institute of Education Sciences

NCEE 2009-4077
U.S. Department of Education

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

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
John Q. Easton
Acting Commissioner

September 2009

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.

This report is in the public domain. Authorization to reproduce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should read: Black, A. R., Somers, M.-A., Doolittle, F., Unterman, R., and Grossman, J. B. (2009). *The Evaluation of Enhanced Academic Instruction in After-School Programs: Final Report* (NCEE 2009-4077). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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

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.

Table of Contents

List of Tables, Figures, and Boxes

Table

Table

Table

Table

Table

Figure

Figure

Box

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; 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 Study, 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 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, as well as classroom observations and interviews with school staff working in the regular after-school program. And Laurie Kotloff at P/PV processed and managed the interview data.

At MDRC, Alixandra Barasch coordinated production of this report and worked tirelessly to prepare the final versions of tables, figures, and other supporting documents. Rebecca Kleinman and Adam Wodka assisted with data collection and provided programming and analysis support. Gordon Berlin, James Kemple, Cynthia Miller, Corinne Herlihy, and John Hutchins provided substantive expertise through their thoughtful comments on, and reviews of, this report. And Mario Flecha, Julia Gomez, Diane Singer, Setha Sean, and Genevieve Williams assisted with report production and fact-checking.

Finally, the authors would like to thank Robert Weber and John Hutchins for their thoughtful editing of the report and Stephanie Cowell and Inna Kruglaya for preparing the final text for publication.

The Author

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

The primary purpose of this study is to determine whether providing structured academic instruction in reading or math to students in grades two to five during their after-school hours — instead of the less formal academic supports offered in regular after-school programs — improves their academic performance in the subject. This is the second and final report from the Evaluation of Enhanced Academic Instruction in After-School Programs — a two-year demonstration and random assignment evaluation of structured approaches to teaching math and reading in after-school settings. The study is being conducted by MDRC in collaboration with Public/Private Ventures and Survey Research Management.

The study 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), in response to growing interest in using out-of-school hours as an opportunity to help prepare students academically (Bodilly and Beckett, 2005; Ferrandino, 2007; Miller, 2003). The federal government has been making an investment toward this goal through its 21st Century Community Learning Centers (21st CCLC) funding.[1] A distinguishing feature of after-school programs supported by 21st CCLC funds has been the inclusion of an academic component. Yet, findings from the National Evaluation of the 21st CCLC program indicate that, on average, the 21st CCLC program grants had limited effects on students' academic achievement (Dynarski and others, 2003; Dynarski and others, 2004; James-Burdumy et al., 2005). One possible explanation for this finding is that academic programming in after-school centers is typically not sufficiently intensive, usually consisting primarily of sessions in which students received limited additional academic assistance (such as reading/math tutoring or assistance with homework). In response, IES decided to fund the development, implementation, and evaluation of instructional resources for core academic subjects that could be used in after-school programs.

As part of this study, enhanced after-school programs providing instruction in either reading or math were implemented in after-school centers during two school years. In the first year of the demonstration (2005-2006), the enhanced programs were implemented in 50 after-school centers — with 25 after-school centers offering the enhanced math program and 25 centers offering the enhanced reading program. The study was then extended to include a second year of operations (2006-2007). This report focuses on the 27 after-school centers that

[1]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).

agreed to participate in the study for both years — 15 of which implemented an enhanced after-school math program, and 12 of which offered the enhanced after-school reading program.[2]

The purpose of this report is to address questions that are relevant to both years of implementation, such as whether one-year impacts are different in the second year of program operations and whether students benefit from being *offered* two years of enhanced after-school academic instruction. Therefore, this report presents findings from the 27 centers that have data to address all these study questions.

Key Findings

Enhanced Math Program

- **One year of enhanced instruction produces positive and statistically significant impacts on student achievement.** The impacts in the 15 centers on SAT 10 total math scores are 3.5 scaled score points in the first year (which is statistically significant) and 3.4 scaled score points in the second year of operations (which is not statistically significant). However, the difference in impacts between implementation years is not statistically significant. The impact of 3.5 scaled score points represents approximately one month's worth of extra math learning.

- **Two years of the enhanced program produces no additional achievement benefit beyond the one-year impact.** Several different analyses support this conclusion. An experimental analysis using the two-year sample finds that the estimated impact of offering students the opportunity to enroll in the enhanced program for two consecutive years (2.0 scaled score points, p-value = 0.52) and the estimated impact on these students of their first year of enrollment in the enhanced program (5.2 scaled score points, p-value = 0.07) are not statistically significantly different (p-value = 0.28). A nonexperimental analysis finds that this remains the case after adjustments are made for students in the enhanced program group who did not attend the enhanced program at all in the second year.

- **There was program fidelity across both years of implementation.** Certified teachers were hired, trained, and provided paid preparation time as in-

[2]Findings from all 50 centers are summarized in Appendix A of this report and are presented in the first-year report (Black et al., 2008). The 27 continuing centers are not statistically representative of all 50 centers so the findings from the 27 sites should not be generalized to all 50 centers.

tended; class sizes were approximately 9 students per instructor (intended ratio was 10 students per instructor); and reports from teachers and district coordinators (i.e., locally based technical assistance staff) indicated that teachers were able to cover the expected material in a class session.

- **Students in the enhanced program received math instruction that was more structured and intensive than regular after-school program students.** Students in the enhanced program group were offered formal instruction in math for three hours per week, and students in the regular program received a mix of homework help and other services not focused on math — although 17 percent of regular program group students in the first year, and 27 percent in the second, received some form of math instruction. Overall, during their first year of participation, enhanced program students received between 42 and 48 more hours of after-school math instruction than did students in the regular after-school program, which converts to a 26 to 30 percent increase in formal instruction in math over the course of the school year.

- **No clear lessons emerge for program improvement or targeting the program in particular types of schools.** Analysis exploring the associations between center-level impacts and the characteristics of schools in which centers operated and the implementation of the program produced no strong associations with clear programmatic implications.

Enhanced Reading Program

- **The enhanced program has no impact on total reading test scores after one year of participation.** This is true in both implementation years in these 12 centers.

- **Two years of participation produces significantly fewer gains in reading achievement for students in the enhanced program group.** Experimental analysis finds that offering students two years of the enhanced reading program has a negative and statistically significant impact on their total reading scores. Nonexperimental analysis suggests that this remains the case even after statistical adjustments are made for students in the enhanced program group who did not actually attend the enhanced program in the second year.

- **Though the reading program was staffed and supported as planned, implementation issues — especially related to the pacing of lessons — occurred in both years.** As with math, certified teachers were hired, trained, and provided paid preparation time as intended, and class sizes were approx-

imately 9 students per instructor (intended ratio was 10 students per instructor). However, lesson pacing was a problem in the first year and continued to be in the second year in at least four of the districts.[3]

- **Students in the enhanced program received reading instruction that was more structured and intensive than regular after-school program students.** Students in the enhanced program group were offered formal reading instruction for three hours per week, and most students in the regular program received a mix of homework help and other services not focused on reading — although 17 percent of regular program group students in the first year, and 12 percent in the second, received some form of reading instruction. Overall, during their first year of participation, enhanced program students received between 54 and 56 more hours of after-school instruction in reading than did students in the regular after-school program, which converts to 22 to 23 percent more formal instruction in reading over the course of the school year.

- **No systematic relationship exists between center-level impacts and program implementation or the local school context.**

Research Questions

The overarching purpose of this evaluation is to determine whether providing students with enhanced after-school academic instruction improves their math or reading achievement above and beyond what they would have achieved had they remained in a regular after-school program. In particular, the study examines whether making the enhanced program available to students for one year improves student achievement, and whether that impact differs when the program is in its second year of operation and, thus, more mature, compared to the first implementation year. Therefore, the following impact questions are examined in this report:

- **What is the impact on student achievement of offering students the opportunity to participate in the enhanced after-school program for one school year?**

- **Is this impact different in the second year of implementation than in the first year?**

[3]In the second year, district staff who helped in implementing the model were asked if pacing continued to be a problem for staff. Of all 10 district staff interviewed, four said it was a problem, four said it was not, and two did not answer the question, so it's not clear whether it was or was not a problem in those last two districts.

The study can also examine whether making the enhanced program available to students for two school years — thereby potentially lengthening students' average level of exposure to the program — improves student achievement. Hence, the following question is also addressed in this report:

- **What is the impact of offering students the opportunity to participate in the enhanced after-school programs for two consecutive years?**

To help interpret and understand the magnitude of the impact findings, the study also examines how well the academic services received by the enhanced after-school program group were implemented, whether the implementation differed across implementation years, and whether there is a measurable difference between the services received by students assigned to the enhanced program and the services received by students assigned to the regular after-school program.

The report also examines two questions that cannot be answered based on the experimental design of the study. First, in order to provide information about the treatment for those who actually received it in both years (rather than the effect of offering two years of programming, which includes students who did not actually participate both years), this report examines the relationship between achievement and program participation for those students who participated in both years of the enhanced after-school services. Second, because the enhanced program was offered in a variety of settings, this report also examines the association between impacts on achievement and the variation in the local school context, as well as variation in program implementation. These nonexperimental findings can then be used to help interpret the generalizability of the overall experimental findings, as well as generate possible avenues for program improvement.

Study Design

After-School Centers in the Study

At the start of the study, 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 27 after-school centers that voluntarily agreed to participate in the study for a second year are located in 11 sites within 10 states and include schools and community-based organizations in a variety of municipalities (rural, urban, and suburban) across the country. They provided the same type of enhanced after-school program (math or reading) as they had provided in the first year of the study.

Student Sample and Random Assignment

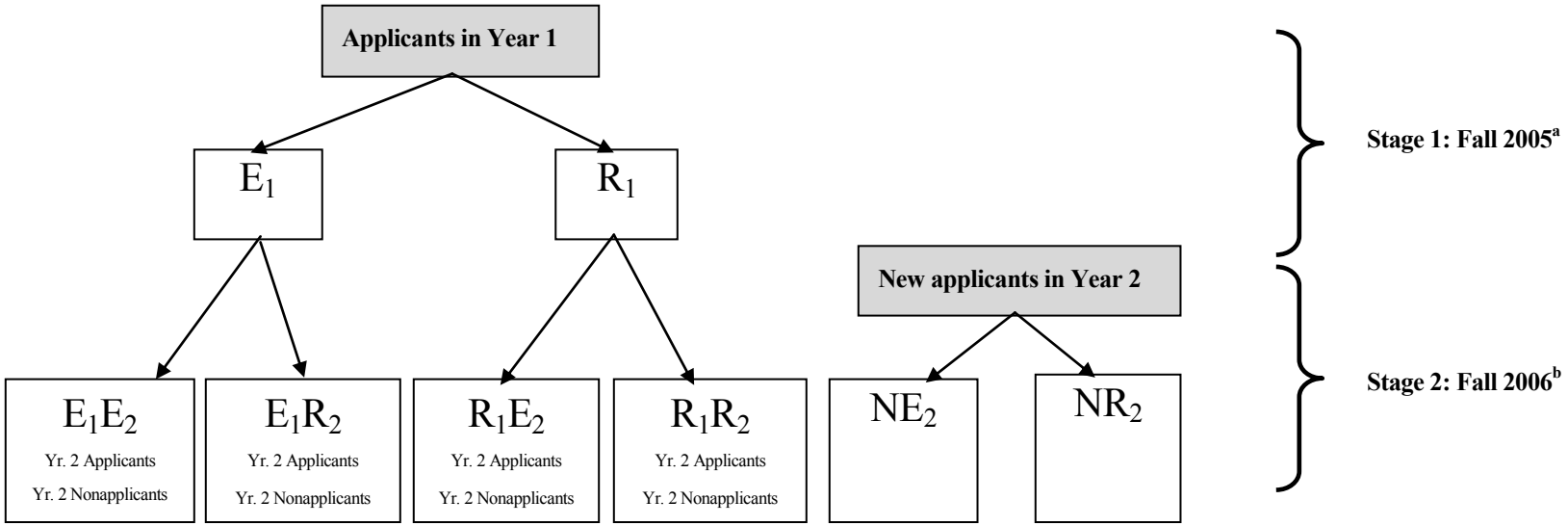
The research design uses a lottery-like process (random assignment) to offer students one of two alternative types of academic support during a 45-minute block of time: the enhanced after-school academic services being tested in this project or the regular after-school services offered in their center. Regular after-school services consisted most commonly of help with homework — although, across both years of implementation, 22 percent of regular program staff in math centers reported providing some form of academic instruction in math and 14 percent of regular program staff in reading sites reported providing some form of academic instruction in reading.

The target population for the study is students in second through fifth grades who are behind grade level in reading or math but not by more than two years. The study sample was recruited from students enrolled in after-school programs and identified by local staff as in need of supplemental academic support to meet local academic standards. Those whose parents then consented to be part of the study and applied for their children to participate in the enhanced program were included in the study sample. Given that instruction in these programs is provided in a small-group format and is not specifically developed to address special needs, students with severe learning disabilities and behavior problems or who could not receive instruction in English were excluded from the sample.

This study is based on a two-stage random assignment design of students, in which students were randomly assigned by grade within each after-school center on two separate occasions — once at the beginning of the first year of the study (first stage in fall 2005, see Stage 1 of Figure ES.1) and then again at the beginning of the second study year (second stage in fall 2006, see Stage 2 of Figure ES.1). (For more details on this two-stage random assignment design, see Box ES.1.) As a result, the sample includes: students who applied to the first year of the study (as described above) and were randomly assigned to either the enhanced program group (E_1) or the regular program group (R_1) and are referred to throughout this report as Cohort 1; students who were not offered the enhanced program in the first year and were applicants in the second year who were either offered the enhanced program (R_1E_2 and NE_2 applicants) or the regular program (R_1R_2 and NR_2 applicants) and are referred to throughout this report as Cohort 2; and students who, through the two-stage random assignment design, were randomly assigned to the enhanced program in both implementation years (E_1E_2 group in Figure ES.1) or assigned to the regular program in both years (R_1R_2 group) and are referred to as the two-year sample. Cohort 1 and Cohort 2 student samples are used to estimate the one-year intent-to-treat impact of the program in the first and second implementation years, respectively. The two-year sample is used to estimate the intent-to-treat impact of offering students the enhanced program for two consecutive years.

Figure ES.1

The Two-Stage Random Assignment Process



E_1 = Enhanced program group, Year 1 E_2 = Enhanced program group, Year 2
R_1 = Regular program group, Year 1 R_2 = Regular program group, Year 2
N = Not in Year 1 study sample (new to the study in Year 2)

NOTES:

[a]In Stage 1 of random assignment, all identified low-performing students who applied to the study were randomly assigned, stratified by grade within each after-school center, to either the enhanced after-school program or the regular after-school program.

[b]Stage 2 of random assignment consisted of two groups, applicants and nonapplicants. Applicants in the second year consisted of newly identified low-performing student applicants in Year 2 and students from Year 1 who applied to the second year of the study. Both of these groups of second year student applicants were randomly assigned, stratified by grade and their first year treatment status (whether they were part of the enhanced or regular after-school program group, or not part of the study in its first year) within each after-school center, to either the enhanced after-school program or the regular after-school program. Nonapplicants are those students from Year 1 who had participated in the first year of the study, but did not apply to the second year of the study. They too were randomly assigned (separately from applicants) by grade and their first year treatment status within each after-school center.

Box ES.1

Two-Stage Random Assignment Design

The study is based on a two-stage random assignment design. At the beginning of the first study year (1st stage in fall 2005, see Stage 1 of Figure ES.1), identified low-performing students who applied to the study were randomly assigned by grade within each after-school center to either the enhanced program group (E_1) or the regular program group (R_1), and are referred to as Cohort 1.

At the end of the first study year, IES decided to extend the study for a second study year to assess both: (1) the one-year impact of the enhanced program and whether that impact changes over time once the site and staff have experience with the program (i.e., a comparison of the one-year impact of the program between the first and second study year), and (2) the impact of extended exposure to the enhanced program (i.e., an estimate of the two-year cumulative effect of being offered the enhanced program both years compared to being offered the regular program both years).

In order to address both these goals for the second study year, a second round of random assignment was conducted consisting of two groups of students, applicants and nonapplicants (2nd stage in fall 2006, see Stage 2 of Figure ES.1). The application process in the second year of the study was conducted the same as in the first year of the study. Applicants in the second year consisted of newly identified low-performing students who were *new applicants in year 2* and students from Cohort 1 who voluntarily applied to the second year of the study. Both of these groups of student applicants in Year 2 were randomly assigned by grade within each after-school center to either the enhanced program group or the regular program group; applicants from Cohort 1 were also randomly assigned by their first year treatment status (whether they were part of the enhanced or regular after-school program group). Randomly assigning for a second time students who participated in the first year, rather than allowing them to maintain their initial randomly assigned grouping, ensured that those who were offered the enhanced program the first year did not receive special treatment once the study was extended.

Nonapplicants are the remaining Cohort 1 students who had participated in the first year of the study, but did not apply to the second year of the study. They too were randomly assigned (separately from applicants) by grade and their first year treatment status within each after-school center. Randomly assigning both the applicants and nonapplicants from Cohort 1 maintains an intent-to-treat sample of Cohort 1 students who are cumulatively offered two years of the program or never offered the program. (Note, fifth-graders from Cohort 1 were excluded from the second stage of the random assignment in fall 2006 because, as sixth-graders, they were no longer eligible for the program and thus did not reapply.)

Impact findings 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 (and all students in the second year), the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) was also administered to measure fluency.

When estimating the impact of one year of exposure to the enhanced instruction separately for each implementation year, the study is equipped to detect an impact of 0.10 standard deviation in math and 0.11 standard deviation in reading in the first year of implementation, and an impact of 0.15 standard deviation in math and 0.14 standard deviation in reading in the second year of implementation.[4] The study is also equipped to detect the impact of offering students two consecutive years of the program that is as small as a 0.21 standard deviation for the math program and 0.23 standard deviation for the reading program.

The following two sections present findings for the enhanced math and reading programs, respectively, based on the 27 after-school centers that participated in both years of the study.

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.

[4]The number of students in the sample is a crucial factor that determines the degree to which the impacts on student achievement and other outcomes can be estimated with enough precision to reject with confidence the hypothesis that the program had no effect. In general, larger sample sizes provide more precise impact estimates. A common way to represent statistical precision is through the "minimum detectable effect size" (MDES). Formally, the MDES is the smallest true program impact (scaled as an effect size) that can be detected with a reasonable degree of power (80 percent) for a given level of statistical significance (5 percent).

- **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, "skill packs" for each topic to provide instruction on the skill in small groups and opportunities for individual practice in its application, 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.

Sites hired certified teachers and operated the enhanced programs with the intended small groups of students, approximately 10 students per instructor. The implementation was supported by the following strategies related to staffing, training and technical assistance, and attendance that were managed and supported by Bloom Associates, Inc.:

- 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 occurred to encourage attendance and address issues preventing it, and incentives to encourage and reward good attendance.

Findings for the Math Program

As mentioned earlier, the math findings presented in this report pertain to the 15 centers that participated in two years of program operations and data collection.

Implementation of the Enhanced After-School Math Program

Overall, the enhanced math program was largely implemented as intended in both years of program operations. Each center was expected to hire certified teachers and to operate with 10 students per instructor. In the first year, for example, 98 percent of instructors were certified

teachers, and the programs operated with the intended small groups of students — on average, in the first year, eight students attended per instructor. The goal was to offer the program for approximately 180 minutes per week, and average offerings were 189 minutes in the first year (a statistically significantly greater amount than intended, p-value = 0.00) and 171 minutes in the second (which does not statistically differ from the amount intended, p-value = 0.45). Instructors were trained by Harcourt staff at the beginning of the year and were provided ongoing assistance.[5] They also received paid preparation time.

Impacts from Offering One Year of the Enhanced Math Program

The impact of enrollment in one year of the enhanced math program on student outcomes is estimated by comparing the outcomes of students who were randomly assigned to enroll in the enhanced after-school math program for one school year with the outcomes of students who were randomly assigned to remain in the regular after-school program during that same school year.[6] This is estimated separately for each implementation year (Cohorts 1 and 2).

On average, students in the enhanced program group in Cohort 1 received 48 more hours of academic instruction in math during the school year than students in the regular program group. This difference — which is statistically significant (p-value = 0.00) — represents an estimated 30 percent increase in total math instruction over and above what is received by these students during the regular school day. In Cohort 2, enhanced program students received 42 more hours — also a statistically significantly greater amount of time (p-value = 0.00) than received by those in the regular program group, and an estimated 26 percent increase in total math instruction. However, the added hours of math instruction was statistically smaller in the second year of implementation (42 hours) than in the first year of implementation (48 hours) (p-value = 0.00).

One year of enrollment in the enhanced after-school program had a positive and statistically significant impact on students' math achievement in Cohort 1 (3.5 scaled score points or 0.09 standard deviation) as measured by SAT 10 total math scores. This statistically significant impact represents a 10 percent improvement over what students in the enhanced group would have achieved had they not had access to the enhanced program, or about one month's extra

[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 (in the first year by Harcourt and Bloom Associates and in the second year by Bloom Associates) and continued support by locally based staff.

[6]Referring back to Figure ES.1, the analysis compared E_1 versus R_1 in the Cohort 1 sample and, in the Cohort 2 sample, R_1E_2 versus R_1R_2 (applicants who had not received the program in the first year) and N_1E_2 versus N_1R_2 (new students in the second year). An overall F-test indicates there is no systematic difference in the baseline characteristics of students in the enhanced and regular program groups in either of the cohort-specific samples.

learning over the course of a nine-month school year. The estimated impact of the enhanced math program on SAT 10 total math scores is not statistically significant for students in the second year of implementation (p-value = 0.07). However, the difference in impacts between implementation years (Cohort 1 and Cohort 2 samples) is not statistically significant. Thus, it cannot be concluded that the enhanced after-school math program was more effective in one implementation year than the other.

One year of enrollment in the enhanced math program also had a positive and statistically significant impact on students' performance on locally administered standardized math tests for Cohort 2 (0.18 standard deviation, p-value = 0.01), and the difference in one-year impacts across cohorts is not statistically significant (p-value = 0.16), so it cannot be concluded that the impact of the enhanced program on locally administered tests differed from one implementation year to the other. However, one year of enrollment did not produce impacts on regular-school-day teacher reports of academic behaviors (homework completion, attentiveness in class, and disruptiveness in class).

Impacts from Offering Two Years of the Enhanced Math Program

The impact of offering students the opportunity to participate in the enhanced program for two consecutive years is estimated using the two-year sample by comparing the outcomes of students who were randomly assigned to either the enhanced after-school program or the regular after-school program for two consecutive school years.[7] However, as mentioned above, to maintain the experimental design, all Cohort 1 students were randomly assigned — both those Cohort 1 students who reapplied in the second year (applicants) and those Cohort 1 students who did not (nonapplicants). Thus, 42 percent of students in the math sample who were offered two years of the enhanced program did not reapply for, and did not receive, the second year of the program services. Hence, the impact findings presented in this section are of a two-year *offer* of services (an intent-to-treat analysis), rather than the impact of *receipt* of two years of the enhanced program — a nonexperimental analysis that is discussed later in this summary.

The estimated impact of offering students the opportunity to participate in the enhanced after-school program for two consecutive years is not statistically significant (2.0 scaled score points on the SAT 10 total score, p-value = 0.52). To place these results into context, the impact of these students' first year in the enhanced program was also estimated and compared to their cumulative two-year impact. Their first-year impact is not statistically significant (5.2 scaled score points, p-value = 0.07). And the estimated impact of assigning students to two years of enhanced services is not statistically different from the impact on these students of their first

[7] An overall F-test indicates there is no systematic difference in the baseline characteristics of students in the enhanced and regular program groups in the two-year sample.

year of access to the program (p-value = 0.28). Hence, for this sample, there is no evidence that offering the enhanced math instruction a second year provides an added benefit.

Figure ES.2 places these impact estimates in the context of the actual and expected two-year achievement growth of students in the enhanced program group. It shows the two-year growth for students in the enhanced program and what their expected growth would have been had they been assigned to the regular program. It also shows the test score growth for a nationally representative sample of students. The test scores of students in the enhanced program group grew 66.3 points over the two years (44.5 points in the first and 21.8 points in the second). Test scores of students in the regular program group grew by 64.3 points (39.4 points in the first year and 24.9 points in the second). These growth rates for the two program groups produce the estimated (not statistically significant) impacts mentioned above, a five-point difference in test scores for this sample after one year and a two-point difference after two years.

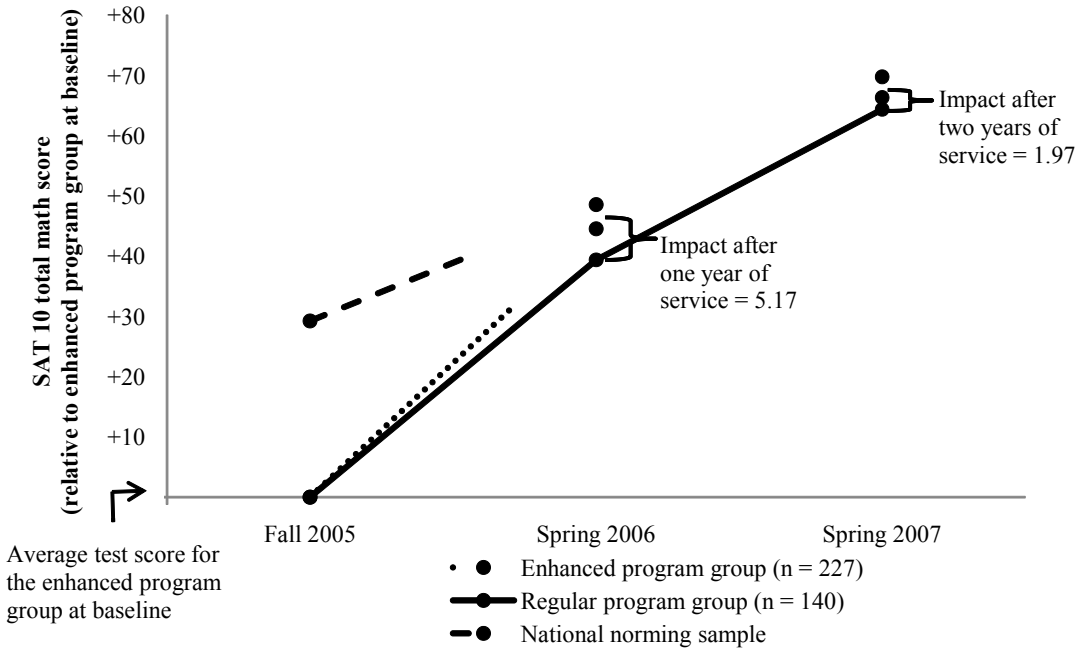
Because not all students in the enhanced program group actually received a second year of enhanced services, a nonexperimental analysis was conducted to examine whether longer exposure to the enhanced program is associated with improved math achievement. This analysis is based on instrumental variables estimation, which makes it possible to statistically adjust for the 42 percent of students in the enhanced program group who never attended the enhanced program in the second year. These findings do not establish causal inferences and thus should be viewed as hypothesis-generating. However, such an analysis may help with interpreting the two-year impacts and provide useful information to program developers.

The findings from this nonexperimental analysis suggest that there is no additional benefit to a second year of enhanced services, even after adjustments are made for students who did not attend a second year. The nonexperimental estimate of receiving two years of enhanced after-school services (3.7 scaled score points for SAT 10 total math scores, p-value = 0.36) does not statistically differ from the 5.2 scaled score points estimated impact of one year of enhanced services (p-value = 0.40). Thus, across both the experimental and nonexperimental analyses, there is no evidence that a second year of the enhanced program — whether offered or received — improves math achievement, over and above the gains produced by the first year of enrollment.

Because the effectiveness of enhanced after-school instruction may be related to factors associated with program implementation or what the students experience during the regular school day, the study also examined whether characteristics of schools and program implementation are correlated with center-level impacts. The analysis is based on center-level impacts in both years of the study (i.e., 30 center-level impacts) and examines whether the impact of one year of enhanced services on SAT 10 total math scores in each after-school center is associated

Figure ES.2

SAT 10 Total Math Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Math Program After One Year and Two Years of Service



SOURCES: MDRC calculations are from baseline and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery. National norming sample calculations are from the SAT 10 (2002 norming sample): Stanford Achievement Test Series: Tenth Edition: Technical Data Report (Harcourt Assessment, 2004, pp. 312-338).

NOTES: The growth line for the enhanced program group is based on the observed mean baseline and follow-up test scores of students assigned to the enhanced after-school program for two consecutive years (baseline is Fall 2005; follow-ups are Spring 2006 and Spring 2007). The growth line for the regular program group represents the test scores that students in the enhanced program group would have obtained had they not been assigned to the enhanced program (calculated as the mean test score for the enhanced program group minus the estimated impact at a given time point). The growth line for the national norming sample is based on the average SAT 10 total math scores for a nationally representative sample of students with the same grade composition in each period as the two-year sample. Specifically, at each point in time (the fall baseline, the first spring, and the second spring), the SAT 10 national norm scores for second-, third-, and fourth-graders are averaged weighting each grade average score according to their proportion in the two-year study sample at baseline. This creates an expected two-year improvement of nationally representative students at the same grade levels as this study's sample. The baseline for the national norming sample is set relative to the average baseline score of the enhanced program group.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data was not collected. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

with (1) the characteristics of the school that housed the after-school center and (2) the characteristics of a center's implementation of the enhanced program.

Though center-level program impacts on total math scores are correlated jointly with the overall set of school context and implementation measures included in the analysis, as well as with some individual measures, no clear lessons emerge for program operations. Program impacts were larger in after-school centers that offered the enhanced program for a greater number of days during the school year, suggesting a positive association between impacts and program dosage. However, this finding is inconsistent with the nonexperimental estimates of two versus one year of enhanced program participation. Program impacts were also larger in centers where one or more teachers left the enhanced program during the school year and in schools that made their Adequate Yearly Progress goals. With the available information, it is not possible to explain the reason for these relationships.

Findings for the Reading Program

Again, the reading findings presented in this report pertain to the 12 centers that participated in two years of program operations and data collection.

Implementation of the Enhanced After-School Reading Program

The enhanced reading program was staffed as intended and offered the intended amount of instruction in both years of program operations. Each center was expected to hire certified teachers and to operate with 10 students per instructor. In the second year, for example, all instructors were certified teachers, and the programs operated with the intended small groups of students — on average, in the second year, nine students attended per instructor. The goal was to offer the program for approximately 180 minutes per week, and average offerings were 177 minutes in the first year and 175 minutes in the second. Instructors were trained by SFA staff at the beginning of the year and were provided ongoing assistance.[8] They also received paid preparation time.

However, in both years of the study, instructors found it challenging to maintain the intended pace of instruction. In the first year of the study, 79 percent of instructors reported that it was consistently or sometimes difficult to include all aspects of the reading program and maintain the intended pace of the daily lesson plan. In the second year of the study, half of the responding district coordinators reported that pacing continued to be a problem.

[8]Ongoing support given to the enhanced program staff consisted of multiple on-site technical assistance visits (by SFA and Bloom Associates) and continued support by locally based staff.

Classroom observations conducted by district coordinators were used to assess the fidelity with which instructors implemented the enhanced reading program. In the classes with students at the first- and second-grade reading levels (in Adventure Island, students are grouped by their initial reading level, not by grade), average fidelity scores did not statistically differ across the first and second years of implementation;[9] in the classes with students reading above the second-grade level, average scores were lower in the second year, by a statistically significant amount (p-value = 0.00).[10] It was also found that, in any given year, implementation of the program lacked consistency, as indicated by variation in the number of program components implemented by teachers.[11] In particular, in the second implementation year, returning teachers in both the lower and upper levels of the program had statistically significantly higher implementation fidelity scores than teachers who were new to the program (p-value = 0.00).

Impacts from Offering One Year of the Enhanced Reading Program

This analysis focuses on the impact of one year of enrollment in the enhanced reading program on student outcomes.[12] The difference between the background characteristics of students in the enhanced and regular program groups, both in Cohort 1 and Cohort 2, was greater than what would be predicted by chance, especially as related to baseline reading achievement test scores and household composition.[13] Measures of student characteristics (including students' baseline test score) were included in the impact model to control for observed differences between the two program groups at baseline. Sensitivity analyses were conducted to gauge whether these covariates adequately control for baseline differences between students in the two program groups. These tests confirm that controlling for students' baseline characteristics — and particularly their pretest scores — produces internally valid estimates of the impact of the enhanced program.

[9]In both years, the average fidelity score was 5.1 out of a total possible score of six components.

[10]The average fidelity score was 4.2 out of a total possible score of five components in the first year; in the second year, it was 3.9.

[11]For example, in the first implementation year, 9 percent of lower-level Adventure Island classes included between three and four of the six measured components; 68 percent included between four and five and 23 percent included between five and six.

[12]As was the case for math, this question is answered by comparing the outcomes of students who were randomly assigned to enroll in the enhanced after-school reading program for one school year and the outcomes of students who were randomly assigned to remain in the regular after-school program during that same school year. Referring back to Figure ES.1, the analysis compared E_1 versus R_1 in the first year sample, R_1E_2 versus R_1R_2 (returning students who had not received the program in the first year) and N_1E_2 versus N_1R_2 (new students) in the second year.

[13]Students in the enhanced group had statistically significantly lower baseline test scores and were more likely to come from a single-adult household.

On average, students in the enhanced program reading group in Cohort 1 received 54 more hours of academic instruction in reading during the school year than students in the regular program group. This difference — which is statistically significant (p-value = 0.00) — represents an estimated 22 percent increase in total reading instruction over and above what is received by these students during the regular school day. In Cohort 2, enhanced program students received 56 more hours — also a statistically significantly greater amount of time (p-value = 0.00) than received by those in the regular program group, and an estimated 23 percent increase in total reading instruction. And the net difference in added hours of instructional reading between implementation years is not statistically significant (p-value = 0.63).

One year of enrollment in the enhanced after-school reading program did not have a statistically significant impact on students' reading achievement (as measured by SAT 10 total reading scores), whether in the first or second year of implementation. It also did not have a significant impact on students' performance on locally administered standardized reading tests, nor did it produce impacts on the DIBELS measures of fluency or on regular-school-day teacher reports of academic behaviors (homework completion, attentiveness in class, and disruptiveness in class).

Impacts from Offering Two Years of the Enhanced Reading Program

The impact of offering students the opportunity to participate in the enhanced reading program for two consecutive years is estimated using the two-year sample in the same way as for the math sample, by comparing the outcomes of students who were randomly assigned to either the enhanced after-school program or the regular after-school program for two consecutive school years.[14] The difference between the background characteristics of students in the enhanced and regular program groups in the two-year sample was greater than what would be predicted by chance, especially related to baseline reading achievement test scores and household composition.[15] Measures of student characteristics (including students' baseline test scores) were included in the impact model to control for observed differences between the two program

[14]Referring back to Figure ES.1, this analysis involves comparing students in E_1E_2 versus R_1R_2. As noted in the discussion of the math findings, the two-year sample includes "nonapplicants" from the first-year study sample who did not reapply to second year of the study. These nonapplicants — who constitute 43 percent of students in the enhanced program group for this analysis — did not actually receive a second year of enhanced after-school services as intended. Hence, the impact findings presented in this section are of a two-year *offer* of services (an intent-to-treat analysis), rather than the impact of two years of *receiving* the enhanced program, which is a nonexperimental analysis discussed later in this summary.

[15]Students in the enhanced program group have lower baseline test scores on average and are more likely to come from a single-adult household.

groups at baseline. Sensitivity analyses were conducted to gauge whether these covariates adequately control for baseline differences between students in the two program groups. These tests confirm that controlling for students' baseline characteristics — and particularly their pre-test score — produces internally valid estimates of the impact of the enhanced program.

The estimated impact of offering students the opportunity to enroll in the enhanced after-school program for two consecutive years is negative and statistically significant (-5.6 scaled score points on SAT 10 total reading scores; p-value = 0.04). To place these results into context, the estimated impact on these students of their first year of program enrollment (-3.6 points) was not statistically significant. And the estimated impact of assigning students to two years of enhanced services does not statistically differ from the impact on these students of their first year of access to the program (p-value = 0.46). Hence, while it can be said that being assigned to two years of enhanced services produces significantly fewer gains on test scores, it cannot be concluded that assigning students to enroll in the enhanced program for two years has a different impact on their reading achievement than assigning them to enroll in one year of the enhanced program.

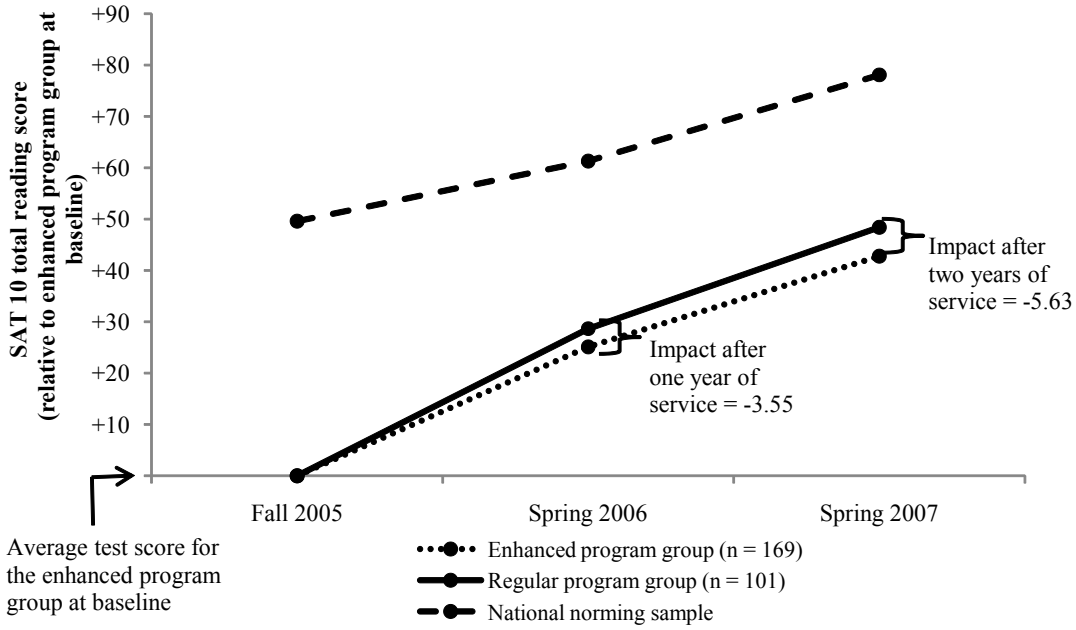
Figure ES.3 places these impact estimates in the context of the actual and expected two-year achievement growth of students in the enhanced program group. It shows the two-year growth for students in the enhanced program and what their expected growth would have been had they been assigned to the regular program. It also shows the test score growth for a nationally representative sample of students. The test scores of students in the enhanced program group grew 25.1 points in the first year and 17.7 points in the second, for a total of 42.8 points. However, the test scores of students in the regular program group also grew, by 28.7 points in the first year and 19.7 points in the second, for a total of 48.4 points. The difference in growth rates between the two program groups produces the two-year impact estimate mentioned above, a -5.6-point difference after two years (in favor of the regular program group).

As in the math analysis, the association between receiving two years of enhanced services and reading achievement was estimated using nonexperimental methods, by statistically adjusting for the 43 percent of students in the enhanced program group who did not attend the program in the second year.[16] Consistent with the experimental estimate for the impact of offering students two years of enhanced services, the association between receiving enhanced academic services for two consecutive years and SAT 10 total reading scores is negative and statistically significant (-7.5 scaled score points, p-value = 0.04). These findings suggest that two years of enhanced after-school services — whether offered or received — produces significantly fewer gains on reading achievement than two years in the regular program group.

[16]The association between receiving two years of enhanced services and reading achievement is estimated using instrumental variables estimation.

Figure ES.3

SAT 10 Total Reading Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Reading Program After One Year and Two Years of Service



SOURCES: MDRC calculations are from baseline and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery. National norming sample calculations are from the SAT 10 (2002 norming sample): Stanford Achievement Test Series: Tenth Edition: Technical Data Report (Harcourt Assessment, 2004, pp. 312-338).

NOTES: The growth line for the enhanced program group is based on the observed mean baseline and follow-up test scores of students assigned to the enhanced after-school program for two consecutive years (baseline is Fall 2005; follow-ups are Spring 2006 and Spring 2007). The growth line for the regular program group represents the test scores that students in the enhanced program group would have obtained had they not been assigned to the enhanced program (calculated as the mean test score for the enhanced program group minus the estimated impact at a given time point). The growth line for the national norming sample is based on the average SAT 10 total reading scores for a nationally representative sample of students with the same grade composition in each period as the two-year sample. Specifically, at each point in time (the fall baseline, the first spring, and the second spring), the SAT 10 national norm scores for second-, third-, and fourth-graders are averaged weighting each grade average score according to their proportion in the two-year study sample at baseline. This creates an expected two-year improvement of nationally representative students at the same grade levels as this study's sample. The baseline for the national norming sample is set relative to the average baseline score of the enhanced program group.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data was not collected. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

The analysis also explored whether the one-year impact estimates for each of the 12 centers are correlated with factors related to program implementation or what the students experience during the regular school day. The analysis is based on center-level impacts in both years of the study (i.e., 24 center-level impacts) and examines whether the impact of one year of enhanced services on SAT 10 total reading scores in each after-school center is associated with (1) the characteristics of the school that housed the after-school center and (2) the characteristics of a center's implementation of the enhanced program. Program impacts on total reading scores are not systematically correlated jointly with either the set of school context and implementation characteristics or with any of those characteristics individually. Thus, the measured local characteristics do not highlight any lessons for settings in which the program will be more effective than average.

Conclusion

This project found that it is possible to implement structured instruction in math and reading for second- through fifth-graders in an after-school setting. The provision of four days of training, ongoing on-site technical assistance, and local program coordinators supported implementation. In both years, math instructors reported few problems implementing Mathletics; teachers implementing the Adventure Island reading program found it challenging to maintain the intended pace of instruction in both years of the study.

It also proved possible to recruit certified teachers who will commit to participate for the full school year. Despite staff turnover across the two years of service offerings, there was growing experience in implementing the programs in the centers. Students also could be recruited each year and retained within each year in the program. The enhanced programs included a combination of extra monitoring of attendance and incentives and encouragement to attend, and students attended the enhanced program as much or more than regular after-school activities, despite initial concerns the program would not be appealing to students or their parents. However, as with most after-school programs (Dynarski and others, 2003; Dynarski and others, 2004), there was substantial dropoff in enrollment across school years (i.e., 42 and 43 percent of students who participated in the enhanced math and reading programs, respectively, in the first year did not attend the enhanced program for a second year).

The enhanced program produced a 26 to 30 percent increase in hours of academic instruction for math and 22 to 23 percent increase for reading, over the school year. For math, this produced one-month worth of extra learning, as measured by math standardized tests. Further, for math, the findings suggest that the benefits of the after-school academic instruction are captured in students' first year of participation, as a second year of math instruction did not

produce any additional benefits for students. However, for reading there were no positive effects on achievement after one year of the program, and findings after two years indicated the enhanced reading program led to slower progress in reading than did the regular after-school programming. In conclusion, these findings are consistent with a growing body of research that finds some evidence of improving achievement through after-school activities (Vandell, Reisner, and Pierce, 2007; Zief, Lauver, and Maynard, 2006).

Chapter 1

Overview of the Study

This is the second and final report from the Evaluation of Enhanced Academic Instruction in After-School Programs — a two-year demonstration and rigorous evaluation of structured approaches to teaching math and reading in after-school settings. The primary purpose of this study is to determine whether providing students in grades two to five with structured academic instruction during their after-school hours — instead of less formal academic supports offered in regular after-school programs, such as help with homework — improves their academic outcomes. The target population for this study is comprised of students who do not meet local academic performance standards.

The study 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), in response to growing interest in using out-of-school hours as an opportunity to help prepare students academically (Bodilly and Beckett, 2005; Ferrandino, 2007; Miller, 2003). The federal government has been making a substantial investment toward this goal through its 21st Century Community Learning Centers (21st CCLC) funding.[1] A distinguishing feature of after-school programs supported by 21st CCLC funds has been the inclusion of an academic component. Yet, findings from the National Evaluation of the 21st CCLC program indicate that, on average, the 21st CCLC program grants had limited effects on participating elementary school students' academic achievement (Dynarski and others, 2003; Dynarski and others, 2004; James-Burdumy, 2005). One possible explanation for this finding is that academic programming in after-school centers is not sufficiently intensive, consisting primarily of sessions in which students receive limited additional academic assistance (such as reading/math instruction or assistance with homework). In response, IES decided to fund the development, implementation, and evaluation of structured models of academic programming in after-school settings.

As part of this study, enhanced after-school programs that provide instruction in either reading or math were implemented in after-school centers during two school years. In the first year of the demonstration (2005-2006), the enhanced programs were implemented in 50 after-school centers — with 25 after-school centers offering the enhanced math program and 25

[1]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).

centers offering the enhanced reading program. The study was then extended to include a second year of operations (2006-2007). This report focuses on the 27 after-school centers that agreed to participate in the study for both years — 15 of which implemented the enhanced after-school math program, and 12 of which offered the enhanced after-school reading program.[2] The purpose of this report is to address questions that are relevant to both years of program implementation — such as whether one-year impacts are different in the second year of program operations and whether students benefit from being offered two years of enhanced after-school academic instruction. Therefore, this report presents findings within the 27 centers that have data to address all these study questions.[3] The evaluation was conducted by MDRC in collaboration with Public/Private Ventures and Survey Research Management. A separate team at Bloom Associates, Inc., organized the process of selecting the math and reading model developers for the project and supported the implementation of the interventions in the after-school setting.

This chapter begins by providing an overview of existing evidence on the effectiveness of academic instruction in an after-school setting and a description of the enhanced after-school programs that are tested in this study, including the theory of action that underlies the interventions. It then describes the strategies used to support the implementation of these models in the study sites and the costs associated with implementing the enhanced programs. The chapter then describes the research questions and concludes with an overview of the structure of the report.

Existing Evidence

This project contributes to an ongoing body of research on after-school programs focused on comparing academic outcomes for students who participate in enhanced after-school programs with a comparable group of students who do not. In addition to the previously mentioned National Evaluation of the 21st CCLC program, this body of research, which covers elementary, middle, and high school-level programs and relies on a wide range of impact estimation methods, has recently been summarized in review articles by Lauer and others (2006), Zief and others (2006), Durlak and Weissberg (2007), Little and others (2008), and Granger (2008).

[2]Findings from all 50 centers are summarized in Appendix A and are presented in the first-year report (Black et al., 2008).

[3]Sites for this study were selected purposefully. Additionally, the 27 after-school centers that returned for the second year of the study (and which are the focus of this report) are not representative of the 50 centers that participated in the first year of the study. Thus, the findings presented in this report are not generalizable statistically to the entire group of after-school centers that participated in this study, nor are they generalizable to a larger universe of after-school programs. Appendix A presents an analysis of impacts from the first year of the study for after-school centers that returned in the second year of the study compared with centers that did not return.

One review done for the Campbell Collaboration focused exclusively on five experimental research projects (Zief and others, 2006) and did not find evidence of positive impacts on academic outcomes, such as grades and test scores. Several other reviews, which include primarily nonexperimental studies, do find a positive association between program participation and academic outcomes as measured by grades and test scores, but the studies do not all find a positive association between program participation and academic outcomes, and the studies with positive findings do not consistently find them across all measures of academic performance. Surveys of this research have attempted to understand program features that are correlated positively with academic outcomes and hypothesize that programs with "a focus on specific social and personal skills that employed sequential learning activities to develop these skills and had youth actively involved" are more likely to find positive associations between participation and academics (Durlak and Weissberg, 2007; Granger, 2008). Others have emphasized sustained participation, appropriate supervision and training for staff, and partnerships with families, schools, and other community organizations as factors related to positive academic findings (Little and others, 2008).

This project contributes to this research in two major ways. First, it relies on an experimental research design (randomized control trial) to produce impact estimates that can be confidently attributed to the strategies tested rather than other features of the program or students served. Second, this project examines the impact of specific strategies to improve student academic outcomes, contrasting structured instruction in reading or math with less formal academic support. Thus, it is not an assessment of whether participation in any after-school program improves academic outcomes.

Overview of the Intervention

The two after-school instructional models being tested were implemented in 27 study centers during two school years (2005-2006 and 2006-2007). In both years, enhanced academic instruction was to be offered four days per week following attendance-taking and a snack, during the first 45 minutes of the typical two- to three-hour after-school program schedule (a total of 180 minutes per week). In contrast, the regular (or "business as usual") after-school programs in the study would use these 45 minutes for less structured forms of academic support (e.g., homework help or tutoring). Students in both types of the after-school program (enhanced and regular) then participated in enrichment and/or recreational activities.[4] Thus, by design, the 45 minutes of daily instruction provided in the enhanced after-school program substitutes for all or a portion of the time devoted to homework completion or other academic support provided in

[4]Further details on the services provided in the regular after-school program can be found in Chapters 4 and 5 (for the math centers) and Chapters 8 and 9 (for the reading centers).

the regular after-school program. Implementation was supported by strategies related to staffing, training and technical assistance, and attendance. Thus, this evaluation is an efficacy test of an enhanced after-school program that packages several elements: an adapted curriculum, certified teachers, small class sizes, teacher support, and attendance incentives.

The Theory of Action of the Intervention

Low-achieving students often lack the fundamental skills needed to advance academically. Though students may attend after-school programs, these often provide homework help or locally assembled activities, but not structured instruction. This study's theory of action hypothesizes that formal, diagnostically driven, skill-based instruction — implemented by trained certified teachers and supported by incentives to encourage student attendance — will increase student math or reading achievement.

The Selection of the Instructional Models

In February 2004, Harcourt School Publishers (math) and Success for All (reading) were competitively selected to adapt their existing instructional materials for use in after-school programs. The development of these new reading and math models was completed by August 2004, and the models were implemented in a small number of pilot sites during the 2004-2005 school year.[5] Following the pilot year, the models were refined and then implemented in the evaluation sites during the 2005-2006 and 2006-2007 school years.

Instructional Elements of the Models

The after-school instructional models include the use of research-based instructional materials and teaching methods that are specifically designed to work in a voluntary after-school setting. They encompass the following elements:

- Materials consistent with evidence-based research on effective models for reading/math improvement
- Student diagnostic assessment integral to the model (Shepard, 2001, pp. 1066-1101)

[5]Of the 10 schools that piloted the programs, two continued to participate in the study, testing the same program they implemented during the pilot year. However, students who participated during the pilot year are not included in the study sample.

- Content geared to struggling students at multiple levels[6]
- Instruction in a small-group format (a ratio of 10 students to one teacher)
- Lessons of 45 minutes in duration, four days per week
- Lessons and exercises that are self-contained within each after-school session
- Materials that can stand alone and be used regardless of the type of instruction used during the regular school day

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 country), 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.

Below are brief descriptions of the basic structure of each of the two instructional models selected for this study.

Harcourt School Publishers adapted its existing school-day materials into 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 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, and projects, as well as 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,

[6]Although the enhanced programs can serve students from kindergarten through grade five, grades two through five are the focus of this study.

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.

Implementation Support Strategies

Implementation was supported using a set of strategies related to staffing, support for instructors, and attendance. These strategies were utilized in both years of program operations but with less intensity in the second year, as described below. Following is a description of these implementation strategies.

Staffing Strategy

During both years of program operation, sites hired certified teachers and operated the enhanced programs with a student-teacher ratio of approximately 10:1, as intended by the program developers. Three-quarters of the after-school enhanced program staff across both years were teachers who taught during regular hours in the same school; others were retired teachers or other school staff, such as special education teachers, guidance counselors, or staff from a different school within the district. Among those who did teach in that same school during the school day, more than half taught grades two through five (56 percent in the first year of implementation and 54 percent in the second year). These teachers may have taught one or more students in the enhanced after-school program during the regular school day.[7]

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. During the two years of implementation, enhanced group instructors received this training and support in a variety of ways throughout the school year:

- **Local district coordinators.** District coordinators were hired to support the enhanced program implementation. As part of their role, they observed instruction,

[7]Because some second- through fifth-grade staff did not teach the same level after school as they taught during the school day, these percentages serve as an upper bound for the amount of overlap in which students in the enhanced after-school program group were taught by the same teacher during the school day.

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. These individuals were required 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. The district coordinators served up to two centers in each site in the study. In the first year of implementation, the project funded a part-time district coordinator for 10 hours per week per school; during the second year, this was reduced to eight hours per week per school. In the second year, an effort was made to re-recruit the district coordinators from the first year of implementation; of the 22 district coordinators in the second year, 17 had been the district coordinator the year before and were thus experienced in their role.

- **Initial training.** Prior to the start of each school year, all teachers, district coordinators, and district point people — the lead staff person in each district familiar with the school district as well as the structure and operation of the existing after-school programs in their district — attended a two-day training session organized by Bloom Associates. The training sessions included an orientation to the project and training on the academic model. The curriculum developers 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, and examples of instructional approaches and classroom management techniques. They also provided guidance on how to use the assessment tools embedded in the model and offered participants the opportunity to practice instruction and the use of these materials. In the second year, sessions were designed for both experienced teachers and those new to the project, and all but four of the 130 staff providing instruction attended the training.[8]

- **Training for administrators**. In the first year of program implementation, the point person and local district coordinators received an extra day of training focused on their role in the project, management aspects of implementing the academic model, and coaching techniques. In the second year of implementation, Bloom Associates met with the point people and local district coordinators for two days during the summer to outline plans for the second year of the project.

[8]The four staff unable to attend the training were new to the enhanced program in the second year of the study; they had not been trained previously.

Together with these experienced practitioners, Bloom Associates outlined ways to strengthen implementation of the programs.

- **Midyear training.** In January 2006 (first year of implementation) and then again in January 2007 (second year of implementation), Bloom Associates organized two days of follow-up training for district coordinators, lead teachers, and point people from each site on special topics that had arisen 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.

- **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 organized by classroom, for ease of distribution. While sites were provided with the curriculum and all materials at no charge for the first year of program implementation, they were asked to pay the cost of replacing all consumable materials in the second year.

- **Paid daily preparation time.** The design of the intervention called for 30 minutes of daily paid preparation time for instructors on the days that the after-school program met. This daily preparation time was provided in both years of implementation.

- **On-site visits from representatives of the developers.** During the first year of implementation, representatives of Harcourt School Publishers and 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 occurred 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 people. In the second year, representatives of Success for All visited each site once about four to six weeks after program implementation began. Visits in the second year included observation of instruction and meetings with individual instructors for feedback and goal setting. Harcourt School Publishers chose not to visit the sites during the second year.

- **Technical assistance visits by Bloom Associates.** As part of the visits by the developers (or separately, in some cases), Bloom Associates staff visited the sites twice in each of the implementation years, four to six weeks after program implementation began and then again about four months later. During these visits,

Bloom Associates staff met with district coordinators, point people, and the lead teacher at each site (in some centers, a teacher was selected to help with administrative responsibilities). As part of these visits, Bloom Associates staff also observed classrooms, met individually with teachers after the observations, and reviewed classroom records to monitor the pace and coverage of instruction. In the first year of implementation only, Bloom Associates staff would also attend one of the weekly staff meetings conducted to discuss the implementation of the intervention and any other issues that arose.

- **Phone calls between Bloom Associates and the district coordinators.** During the first year of implementation, calls with district coordinators were held weekly; during the second year, Bloom Associates switched to biweekly calls. These phone calls covered particular problems arising in the sites as well as general issues, such as 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.

- **Teacher meetings.** District coordinators and a lead teacher in each center organized meetings for instructors to discuss problems they were encountering in instruction, to convey information from the 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. During the first year of implementation, these meetings were held weekly; during the second year, they were held biweekly.

Efforts to Support Student Attendance

Given the voluntary nature of participation in after-school programming, the project called for efforts to make the academic instruction engaging and to support student attendance through various strategies, including close monitoring of attendance, follow-up with parents and students when absences occurred, and incentives to encourage and reward good attendance.[9]

[9]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). In the 2004-2005 school year, for example, only 65 percent of students enrolled in 21st CCLC-funded programs serving elementary grades were "regular attendees" (i.e., attended for 30 days or more during that school year, which is the 21st CCLC definition of regular attendance). This is 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.

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

- **Monitoring of attendance.** In both years of implementation, weekly attendance reports were collected for students in the enhanced program group and sent to Bloom Associates. These reports were discussed with sites in the phone calls between Bloom Associates and the district coordinators, and follow-up 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 was 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.[10]

- **Incentive plans.** Each after-school center developed an incentive plan in the summer prior to the first year of implementation (summer 2005), which was then submitted to Bloom Associates for approval and announced to families and students. The local district coordinator, lead teachers, and district point person were responsible for the operation of the incentive policy, which continued through the second year of implementation. 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 high 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[11]

    - An end-of-year celebration for participating students

[10]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.

[11]A system of points and rewards is built into the enhanced reading model (Adventure Island), and points earned each week can be spent at the "Ships Store" to buy small prizes or candy. Students in the enhanced math model (Mathletics) received points for good attendance and completion of skill packs.

Key Research Questions

The overarching purpose of this evaluation is to determine whether providing students with enhanced after-school academic instruction improves their math or reading achievement above and beyond what they would have achieved had they remained in a regular after-school program. In particular, the study examines whether making the enhanced program available to students for one year improves student achievement and whether that impact differs when the program is in its second year of operation and, thus, more mature, compared to the first implementation year. Therefore, the following impact questions are examined in this report:

- **What is the impact on student achievement of offering students the opportunity to participate in the enhanced after-school program for one school year?**
- **Is this impact different in the second year of implementation than in the first year?**

The study can also examine whether making the enhanced program available to students for two school years — thereby potentially lengthening students' average level of exposure to the program — improves student achievement. Hence, the following question is also addressed in this report:

- **What is the impact of offering students the opportunity to participate in the enhanced after-school programs for two consecutive years?**

To help interpret and understand the magnitude of the impact findings, this report includes enhanced program implementation information as well as information about the contrast in services provided to treatment and control students. Specifically, the report assesses how well the enhanced after-school programs were implemented in the study centers and whether implementation differed across implementation years. In order to determine whether the enhanced program actually produced a service contrast, the report also examines the measurable differences between the services received by students assigned to the enhanced program and the services received by students assigned to the regular after-school program.

The report also examines two questions that cannot be answered based on the experimental design of the study but that may provide information that could be used to improve the design and implementation of the enhanced programs. First, since continuity of student participation across school years is particularly problematic in after-school settings, the effect of offering two years of programming (often referred to as the effect of the *intent to treat*) includes students assigned to the enhanced program who did not actually participate in the enhanced program in the second year. Thus, in order to provide information about the treatment for those who actually received it in both years, this report will present findings from an exploratory

analysis that examines the relationship between achievement and program participation for those students who participated in both years of the enhanced after-school services.

Second, the enhanced program was offered in a variety of different settings. Understanding how variation in the local school context, as well as variation in program implementation (across centers and the two implementation years), is associated with impacts on achievement can help one interpret the generalizability of the overall findings, as well as generate possible avenues for program improvement. Thus, the report also examines whether the impact of one year of enhanced services (either in the first or second implementation year) is associated with the characteristics of program implementation in the after-school center and/or with the characteristics of the local school context in which the program was implemented.

The Structure of This Report

The chapters in this report focus on the study design and implementation and impact findings of the enhanced after-school programs for the 27 after-school centers that participated in both years of the demonstration. Chapter 2 describes various issues related to the study design, including the selection of after-school centers, the recruitment and randomization of students, the data sources and measures, and the analytic approach used to estimate impacts. The following four chapters then present implementation and impact findings for the enhanced after-school math program. Chapter 3 provides context for the math impact findings by describing the implementation of the enhanced math program in both years. Chapter 4 describes how the services received by students in the enhanced program differ from what was offered in the "business as usual" after-school setting and then presents findings on the impact of offering students the opportunity to participate in the enhanced math program *for one school year* (in either the first or second implementation year). Chapter 5 examines similar issues, but in regards to the cumulative impact of offering students the opportunity to participate in the enhanced math program *for two school years*. Chapter 6 presents findings from exploratory analyses related to the enhanced math program. Chapters 7, 8, 9, and 10 then present analogous implementation and impact findings and exploratory analysis results for the enhanced reading program.

Chapter 2

Study Sample and Design

The present chapter describes the study's research design in more detail. The chapter begins by describing the recruitment and selection of after-school centers, which is followed by a discussion of the student recruitment and randomization process in each year of the study. The chapter then provides details on data collection and the measures created from these data sources, as well as the analytic methods used to assess program impacts. For the purposes of this study, a "site" is defined as the organization managing the after-school program, which in seven sites is a school district and in four sites is a community-based organization. Within each site, the after-school study is implemented in one or more after-school centers. Each center is housed in a school.

Participating After-School Centers

The first step in the site recruitment process was to identify providers of after-school programs serving the target population of students (i.e., students in grades two through five performing below grade level in math and/or reading) and to notify these programs of the study opportunity. After-school centers with these characteristics were identified through various means. First, all 21st Century Community Learning Center (21st CCLC) grantees operating elementary school programs were notified of the study opportunity. Second, through various contacts — including national organizations and research networks — the study team was able to identify other providers of after-school programs serving the target population of students and alerted them to the upcoming study. Finally, the study team contacted organizations representing networks of after-school service providers (e.g., The After-School Corporation, Public Education Network, Education Trust), who in turn advertised the study among their members. In the end, more than 300 operators of after-school programs contacted the study team to inquire about participating in the demonstration.

Because this evaluation is an efficacy study, the project team then selected after-school centers that were willing and able to implement the program with a reasonable level of fidelity, and where there would be a clear service contrast between the enhanced program and "business as usual." Sites were also selected based on the ability to meet the research requirements of the study. Specifically, the following criteria were used to select sites:

- **Serve the desired students.** Sites had to enroll students from the target population of the evaluation — namely, students from low-income families who

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 to 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.

- **Have staff able to deliver instruction.** The after-school centers were required to have or to hire staff members with experience and the ability to deliver academic instruction using structured math or reading materials, with a preference for certified elementary school teachers.[12]

- **Have adequate student attendance.** To increase the opportunity for regular and sustained student participation, after-school centers needed to have formal attendance rules in prior years of operation, creating an expectation of regular student attendance 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.

- **Provide the desired service contrast.** Sites could not use structured materials or provide direct instruction as part of their regular after-school program, so as to ensure that there would be sufficient contrast between "business as usual" and the enhanced program.

- **Able to meet research requirements.** Sites had to be willing and able to follow the research procedures as to random assignment and data collection and

[12]The staffing strategy for the enhanced after-school program calls for teachers who have experience with a structured curriculum. Because teachers' instructional experience can be difficult to assess directly, it was measured in this study using teacher certification (i.e., if a teacher was certified in elementary education, they were deemed by sites and the study team to have experience with a structured curriculum).

had to contribute at least 60 to 80 students — roughly equally distributed across the second through fifth grades — for the research sample.

Recruitment was limited to sites that were able to contribute at least two after-school centers serving children in grades two through five.[13] Whether a program implemented the reading or math program was based on a combination of local preferences, including knowledge of their student needs and sufficient contrast between current academic offerings in the subject area and the enhanced program.

When the evaluation was extended to include an additional year of program operations, the offer to participate in a second year of implementation (the 2006-2007 school year) was extended to all 50 after-school centers that implemented the program in the first year. Continuation in the study was voluntary. Using the same criteria listed above, 27 of the original 50 after-school centers agreed to and were able to participate in the study for another year (15 math centers and 12 reading centers). These after-school centers are located in 11 sites, and they provided the same type of enhanced after-school program (math or reading) as they had provided in the first year of the study.[14]

Table 2.1 shows the sites included in this report, those that implemented the enhanced program for two years (school years 2005-2006 and 2006-2007). They are geographically dispersed across the country.[15] All 27 after-school centers in these sites were housed in elementary schools, and all but six centers were operated by school district staff (as opposed to community-based organizations). Centers in all but one site received 21st CCLC funding.

[13]This additional criterion was used in order to economize on project resources (thereby increasing the funds available for supporting implementation of the programs and for data collection).

[14]The remaining 23 after-school centers — while reporting interest in the enhanced program — were unable to continue for a second year. Thirteen centers were unable to continue because they could not meet the study requirements (e.g., they did not have the funds to meet the teacher requirements for the enhanced program, or they could not meet the sample size requirements due to high student turnover rates). Eight centers were faced with leadership challenges that made implementation in the second year not feasible (e.g., a change of superintendent or staff turnover), and two declined to participate for a second year because they wanted to provide the enhanced program to all students in their after-school program.

[15]Fifty centers operated the program during the first implementation year. Appendix A provides a comparison of impacts and implementation in the 27 after-school centers that participated in both years of the demonstration and the 23 centers that participated in the first year only.

Table 2.1

Sites Implementing Mathletics and Adventure Island for Two Years

Site Name	Location
Perry County Schools	Marion, AL
Mount Diablo Unified School District	Concord, CA
The Lighthouse Program	Bridgeport, CT
School District of Palm Beach County	Palm Beach, FL
Atlanta Public Schools	Atlanta, GA
Geary County Schools	Junction City, KS
Hands Across Cultures	Espanola, NM
Builders for the Family and Youth	Brooklyn, NY
Crown Heights Beacon	Brooklyn, NY
Norristown Area School District	Norristown, PA
West Allis-West Milwaukee School District	West Allis, WI

NOTE: In one of the sites, after-school centers housed in elementary schools are attended by students in grades 2, 3, and 4. In addition to these centers, the Mathletics and Adventure Island programs were implemented in middle schools in this site, where they were offered to fifth-grade students.

Student Recruitment and Random Assignment

Target Population

The target population for this study is comprised of students in second through fifth grades who are below grade level in reading or math, but not by more than two years. At the beginning of the study, local staff members (that is, the district coordinator and teachers) were asked to identify students in need of supplemental academic support to meet local academic standards.[16] Given that instruction in these programs is provided in a small-group format of a 10:1 student-to-teacher ratio, students selected for the study were required to not have serious learning disabilities or behavioral problems and to be able to be instructed in English. All study participants were initially identified from the pool of students who were signing up for the existing after-school program and were likely to attend the program for the full school year. However, if fewer than 60 to 80 students meeting these eligibility criteria were identified, local

[16]Local staff used a variety of measures (classroom performance, performance on state or local administered tests) to recommend students for the program.

after-school center staff would then work with regular-school-day teachers and the principal to identify and recruit additional students to the after-school program.[17]

Local data collection staff, who were part of the research team, then worked with identified students and their parents to complete the study application process. After parents completed an informed consent form, enrollment form, and contact sheet, students completed 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 implemented in that center).[18] Once students had completed these steps, they were eligible for the random assignment lottery. Once a sufficient number of students in a center were eligible,[19] data collection staff submitted a roster of the eligible students to MDRC staff, and MDRC conducted the random assignment lottery using its computer system and then informed the local after-school staff of the results. Through this process, students were randomly assigned 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.[20] (The following section describes this random assignment process in greater detail, for each implementation year.) Enhanced programs in all sites were serving students by mid-October (in both program years). And throughout the school year, local district coordinators worked with the enhanced program teachers to monitor program operations and 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. Thus, among those

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

[18]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, so the use of the SAT 10 for baseline testing was prohibited. Thus, at baseline, students in this school district instead 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.

[19]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. Thus, 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. Additionally, in the second year, students were randomly assigned by their first-year random assignment status, within grade and center, with a ratio of as close to 1:1 as possible, favoring the enhanced program group. Therefore, random assignment did not produce a balanced 1:1 design ratio of enhanced program group to regular program group students in either year.

[20]In most after-school centers, all students participating in the regular after-school program were in the study. However, in some centers, students who did not apply to the study and thus were not assigned to the enhanced or regular programs groups as part of the study sample may have participated in the regular-after school program if the program at that center was large enough to accommodate more students than in the study's regular program group. But these students did not meet the eligibility requirements of the study.

who completed the study application process and were randomly assigned, there were no cases of "cross-overs" in either year.

Random Assignment

The study is based on a two-stage random assignment design. At the beginning of the first study year (first stage in fall 2005, see Stage 1 of Figure 2.1), identified low-performing students who applied to the study (as described above) were randomly assigned by grade within each after-school center to either the enhanced program group or the regular program group; they are referred to throughout this report as Cohort 1.

At the end of the first study year, IES decided to extend the study for a second study year to assess both: (1) the one-year impact of the enhanced program and whether that impact changes over time once the site and staff have experience with the program (i.e., a comparison of the one-year impact of the program between the first and second study year), and (2) the impact of extended exposure to the enhanced program (i.e., an estimate of the two-year cumulative effect of being offered the enhanced program both years compared to being offered the regular program both years). In order to address both these goals for the second study year, a second round of random assignment was conducted consisting of two groups of students, applicants and nonapplicants (second stage in fall 2006, see Stage 2 of Figure 2.1). The application process in the second year of the study was conducted the same as in the first year of the study and is as described above. Applicants in the second year consist of newly identified low-performing student applicants in Year 2 and students from Cohort 1 who voluntarily applied to the second year of the study. Both of these groups of student applicants in Year 2 were randomly assigned by grade within each after-school center to either the enhanced program group or the regular program group; applicants from Cohort 1 were also randomly assigned by their first-year treatment status (whether they were part of the enhanced or regular after-school program group) (see Stage 2 of Figure 2.1).[21] Nonapplicants are the remaining Cohort 1 students who had participated in the first year of the study but did not apply to the second year of the study. They too were randomly assigned (separately from applicants) by grade and their first-year treatment status within each after-school center.[22]

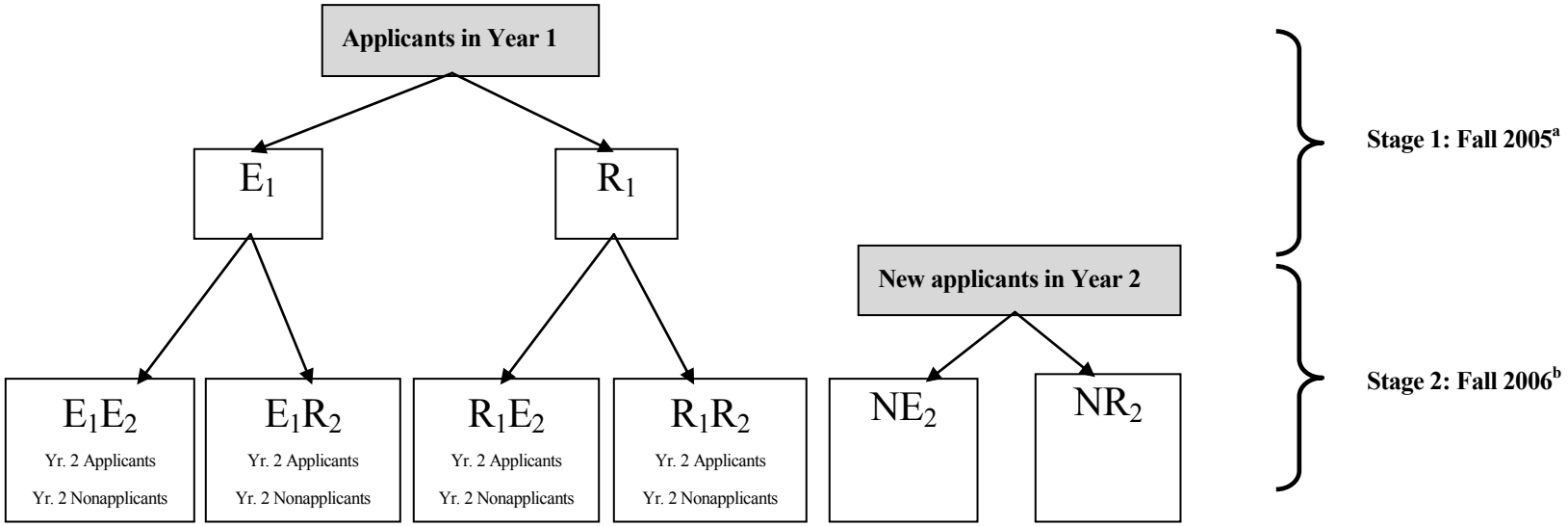
[21] Randomly assigning for a second time students who participated in the first year, rather than allowing them to maintain their initial randomly assigned grouping, ensured that those who were offered the enhanced program the first year did not receive special treatment once the study was extended. Thus, the offer of a second year of the enhanced program was fair. And, fifth-graders from the first study year (fall 2005) were excluded from the second stage of the random assignment in fall 2006 because, as sixth-graders, they were no longer eligible for the program and thus did not reapply.

[22] Randomly assigning both the applicants and nonapplicants from Cohort 1 maintains an intent-to-treat sample of Cohort 1 students who are cumulatively offered two years of the program or never offered the program. This intent-to-treat sample is described further in a subsequent analysis.

The Evaluation of Academic Instruction in After-School Programs

Figure 2.1

The Two-Stage Random Assignment Process



E_1 = Enhanced program group, Year 1 E_2 = Enhanced program group, Year 2
R_1 = Regular program group, Year 1 R_2 = Regular program group, Year 2
N = Not in Year 1 study sample (new to the study in Year 2)

NOTES:

[a]In Stage 1 of random assignment, all identified low-performing students who applied to the study were randomly assigned, stratified by grade within each after-school center, to either the enhanced after-school program or the regular after-school program.

[b]Stage 2 of random assignment consisted of two groups, applicants and nonapplicants. Applicants in the second year consisted of newly identified low-performing student applicants in Year 2 and students from Year 1 who applied to the second year of the study. Both of these groups of second year student applicants were randomly assigned, stratified by grade and their first year treatment status (whether they were part of the enhanced or regular after-school program group, or not part of the study in its first year) within each after-school center, to either the enhanced after-school program or the regular after-school program. Nonapplicants are those students from Year 1 who had participated in the first year of the study, but did not apply to the second year of the study. They too were randomly assigned (separately from applicants) by grade and their first year treatment status within each after-school center.

Analysis of Impacts

Given the random assignment design described above, this section describes the specific comparisons used to answer the key impact questions, all of which pertain to the impact of the enhanced programs on student achievement (as measured by SAT 10 scores).

Impact of offering students one year of enhanced services

The analysis begins by examining whether there is a benefit to students of having access to the enhanced program for one school year in either the first or second study year, addressing the research question:

- **What is the one-year impact on student achievement of offering students the opportunity to participate in the enhanced after-school program for one school year, and is this impact different in the second year of implementation than in the first?**

In order to answer this question, the intent-to-treat (ITT) sample includes students from both study years.[23] As mentioned earlier, Cohort 1 consists of all students randomized in the first year of implementation, within the 27 after-school centers. These students are then used to estimate the one-year impact in the first implementation year (see Figure 2.2, Cohort 1). Second, students who were not offered the enhanced program in the first year, and were applicants in the second year who were either offered the enhanced program (R_1E_2 and NE_2 applicants) or the regular program (R_1R_2 and NR_2 applicants) are used to estimate the one-year impact in the second implementation year, and are referred to throughout this report as Cohort 2 (see Figure 2.2, Cohort 2).[24]

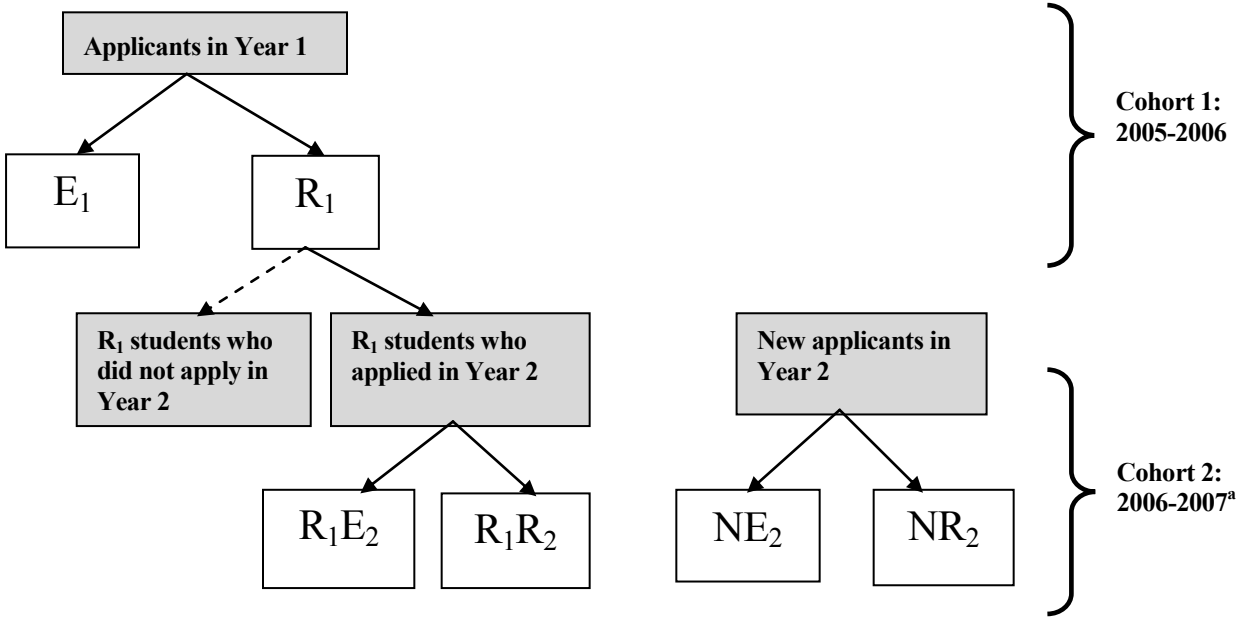
The one-year impact on student achievement is first estimated separately for Cohort 1 and Cohort 2. Because a second year of implementation may lead to greater staff experience with

[23]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.

[24]Note that the construction of the pool of students in each of the two cohorts is identical. In Cohort 1, eligible students who were interested in the enhanced program (as signaled by the application process) and had never received it before were randomly assigned. Similarly, Cohort 2 was formed by randomly assigning all eligible students who were interested in the enhanced program (as signaled by the application process in year 2) and had never received the enhanced program before Year 2. Also, note that the Cohort 2 sample is smaller than the Cohort 1 sample because by definition it excludes students who were offered the enhanced program in the first year (given that this research question pertains to the impact of access to *one year* of enhanced services). Additionally, by excluding these students, the Cohort 2 sample includes a proportionately larger percentage of students in second grade (32 percent) than other grades. Thus, estimates are weighted to ensure that second-grade students do not have a disproportionately greater weight in the Cohort 2 findings (see Appendix G for a discussion of these weights).

Figure 2.2

Analysis Samples Used to Estimate the Impact of Offering Students One Year of the Enhanced Program



Sample sizes for Cohort 1 sample		
	Math	Reading
E_1	634	504
R_1	510	401

Sample sizes for Cohort 2 sample		
	Math	Reading
R_1E_2	144	98
R_1R_1	105	74
NE_2	317	245
NR_2	226	200

E_1 = Enhanced program group, Year 1 E_2 = Enhanced program group, Year 2
R_1 = Regular program group, Year 1 R_2 = Regular program group, Year 2
N = Not in Year 1 study sample (new to the study in Year 2)

NOTES: The sample used in the analysis is limited to students with one-year follow-up data from both the evaluation-administered achievement test and the regular-school-day teacher survey.

[a]The Cohort 2 sample is students who applied to the program the second year and were either offered the enhanced program or the regular program. This includes Cohort 1 students who were not offered the enhanced program in Year 1 and new applicants. Thus, the sample sizes of the R_1 students in Cohort 2 do not sum up to the sample size of R_1 students in Cohort 1.

the programming, the one-year impacts in the second year of implementation within the 27 centers (Cohort 2) are compared with the one-year impacts in the first year of implementation within the same 27 centers (Cohort 1).[25] This comparison provides information about whether the impacts differed between the two implementation years. However, it should be noted that students in Cohort 1 and Cohort 2 may differ in their level of prior exposure to regular after-school services. While some Cohort 1 students may have attended the after-school program in the year prior to the study, it is not known how many. Within Cohort 2, 31 percent of the math sample and 27 percent of the reading sample were part of the regular program group study sample in the first year and did attend the regular after-school program, and some new students may also have attended prior to entering the study. If differences in motivation exist between students who attended the regular after-school program in the year prior to participation in the study and those that did not, then the differences in impacts between cohorts could be influenced.

Impact of offering students two years of enhanced services

An ongoing enhanced program would provide students with access to the program over multiple years. Therefore, the next research question examines the ITT impact of providing students with access to the enhanced program for two consecutive school years:

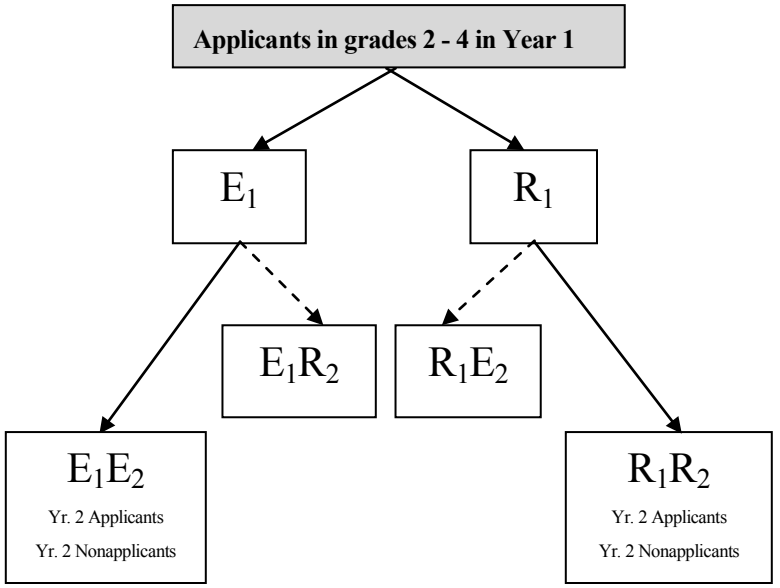
- **What is the impact on student achievement of offering students the opportunity to participate in the enhanced after-school program *for two consecutive school years*?**

This question can be answered by comparing the outcomes of students who, through the two-stage random assignment design, were randomly assigned to the enhanced program in both implementation years (E_1E_2 group in Figure 2.3) to the outcomes of students assigned to the regular program in both years (R_1R_2 group). These two groups of students (E_1E_2 and R_1R_2) will be referred to as the two-year sample. As mentioned above, to maintain the experimental design, all Cohort 1 students were randomly assigned (both those Cohort 1 students who reapplied in the second year — applicants — and those Cohort 1 students who did not — nonapplicants). Thus, this intent-to-treat analysis provides impact estimates of a two-year enhanced after-school program in which 42 percent of students in the math sample and 43 percent in the reading sample who were offered two years of the enhanced program did not reapply for, and did not receive, the second year of the program services. Details on the statistical model that underlies these findings are presented in Appendix H.

[25]When comparing impact estimates between implementation years, standard errors are adjusted to account for student-level clustering caused by the fact that some students appear in both Cohort 1 and Cohort 2. For math, 246 of the 792 observations in the Cohort 2 sample (R_1E_2 or R_1R_2) are students that are part of the Cohort 1 sample (R_1). For reading, 166 of the 626 observations in the Cohort 2 sample are students that are also part of the Cohort 1 sample.

Figure 2.3

Analysis Samples Used to Estimate the Impact of Offering Students Two Years of the Enhanced Program



Sample sizes for two-year sample		
	Math	Reading
E_1E_2	227	169
R_1R_1	140	101

E_1 = Enhanced program group, Year 1 E_2 = Enhanced program group, Year 2
R_1 = Regular program group, Year 1 R_2 = Regular program group, Year 2

NOTES: The sample used in the analysis is limited to students with two-year follow-up data from both the evaluation-administered achievement test and the regular-school-day teacher survey.

[a]This sample includes the two-year intent-to-treat sample; students who were randomly assigned to the enhanced program for both years of the study and students who were randomly assigned to the regular program for both years of the study through a two-stage random assignment process. The sample includes all Year 1 students in grades 2-4, whether or not they reapplied to the center for the second year of the study. Random assignment was stratified by grade, Year 1 treatment status (that is, the enhanced program or the regular program), and whether they reapplied to a second year at the center. Randomizing those first-year students who did not reapply is necessary so that the impact of offering students two consecutive years of the enhanced program could be estimated experimentally. Test and survey data were collected at the end of Year 2. Missing data information can be found in Appendixes C and D.

Data Sources and Measures

The evaluation draws on multiple data sources — some used exclusively for the analysis of program impacts, some used exclusively for the implementation and service contrast analysis, and some used for both aspects of the study. Table 2.2 describes the available data for this study, listing the sources, the samples used, the time of collection, and the type of information provided. This section first describes the data sources for the core impact research question and then describes data used for the implementation and service contrast analysis.

Outcome Measures

Table 2.3 lists the outcome measures used in the impact analysis. Note that all outcomes are measured at the level of individual students. Follow-up data were collected in the spring of each implementation year. Response rates for the one-year sample (math and reading) are between 91 and 100 percent on all measures except the state assessment, which is between 81 and 94 percent. Response rates for the two-year sample in math are between 71 and 82 percent and in reading, between 59 and 79 percent. (See Appendices C and D for additional information about response rates on the outcome measures.)

The primary tool for gauging student achievement is the SAT 10 abbreviated battery test for reading or math.[26] The key outcome measure is the "total" score for the subject that was implemented in the center, but impacts on the subcomponents of the total — vocabulary, 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 more specific types of skills. *Scaled scores* on the SAT 10 are used to allow the comparison of scores across grades.[27]

Because reading fluency is an important skill in the early grades, fluency was measured (in the reading centers) using two subscales of a standard fluency test, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS): the oral reading fluency scale and the nonsense word fluency scale. In year one, second- and third-grade students in the reading sites were administered the DIBELS. In the second year, DIBELS was administered to all study grades in the reading sites.

[26]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.

[27]A secondary measure of academic achievement is the student performance on district-administered standardized tests, given the policy relevance of these test scores. Not all districts in the study test second-grade students, so impacts on this measure are based on a subset of the analysis sample. Additionally, because each district uses a different test, scores are rescaled. Appendix F describes the scaling of this measure.

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 enhanced and regular program group students for the 2005-2006 and 2006-2007 school years.	Daily attendance was collected for all days when the enhanced instruction was offered.
Harcourt School Publishers' Class Record Forms	Data are available for enhanced program group classrooms for the 2006-2007 school year.	Data on the number of skills assigned during the school year, collected from Harcourt School Publishers, were used to assess whether staff were spending the intended amount of time on instruction in the Mathletics program.
After-school staff surveys	Data are available for all after-school staff providing academic support in the study sites both years; includes data for approximately 230 staff serving the enhanced program group and 180 staff serving the regular program group; data were collected from February to April 2006 and February to April 2007.	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 interviews with after-school instructors	Research staff interviewed half the instructors serving the enhanced program group (randomly sampled). 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.
Structured interviews with regular after-school program group staff	Data collection coordinators interviewed two randomly sampled instructors serving the regular after-school program group at each center. Interviews were conducted from March to April 2007.	Questions cover issues around the academic focus of the after-school activity, the content covered each day, the use of assessments, and where materials are drawn from. Specifically, staff were asked about the activity's main method of helping students with academic work. Response categories were as follows: assistance on homework assignments, formal instruction using a published after-school curriculum, practice or review of academic material covered during the school day, something else.

(continued)

Table 2.2 (continued)

Data Source	Sample and Time Collected	Description of Data
Structured interviews with after-school district coordinators	Research staff interviewed the district coordinators from March to April 2007.	Open-ended questions to district coordinators included, but were not limited to, their perspectives on the strengths and weaknesses of the enhanced program, how the 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 about whether challenges identified during the first implementation year continued to be challenges during the second year, whether new challenges surfaced in the second year, and what supports were given in the second year to new teachers.
Structured protocol observations of the implementation of Mathletics and Adventure Island	Data are available for all instructors serving the enhanced program group during both implementation years. 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 years 2005-2006 and 2006-2007.	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 workout components (such as skill packs) appropriately, and inclusion 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 spring 2006 and spring 2007.	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.

(continued)

Table 2.2 (continued)

Data Source	Sample and Time Collected	Description of Data
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 and spring 2007.	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 and 2006 (pre-random assignment) and in spring 2006 and 2007 (follow-up). (Students who were in the first year of the study and returned in the second year were not administered the fall 2006 test. Their baseline test score for the second program year is their spring 2006 score.)	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.
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 in the first year, and for students in all grades in the second year. Fielded in spring 2006 and 2007.	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 and 2006-2007 school years.	Data include test scores on local or state standardized tests.
School or district employees	In spring 2007 and 2008, phone calls were made to school or district employees at schools housing the after-school centers. Also, research staff interviewed point people from March to April 2007.	Employees were asked the name of the reading and math curricula, and the duration of reading and math instruction, during the 2005-2006 and 2006-2007 school years.
Common Core of Data (CCD)	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 and 2006-2007 school years are available for the schools housing the after-school centers.	Data include Adequate Yearly Progress (AYP) status of the schools housing after-school centers.

The Evaluation of Academic Instruction in After-School Programs

Table 2.3

Key Outcome Measures for the Impact Analysis

Outcome Domain	Math Outcome	Reading Outcome

Student achievement[a]	Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery
	• Math total scaled scores
	  ◦ Problem-solving (all grades)
	  ◦ Procedures (all grades)
		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)
		• Oral reading fluency (all grades)[b]
		• Nonsense word fluency (grades 2-3)

Student academic behavior[c]	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

NOTES:

[a]For reliability data on the student achievement outcomes, see Appendix Table F.1.

[b]In the first year of the study, the oral reading fluency measure was administered to just second- and third-grade students.

[c]Each of these measures is based on only one survey item, thus there is no associated reliability data.

Impacts on three measures of student academic behavior — homework completion, attentiveness, and disruptiveness in class — are also examined. These measures are drawn from the survey of the sites' regular-school-day teachers and are included in order to assess whether the enhanced after-school program affected students' behavior in any way.[28] All three measures

[28]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, if the enhanced program improves academic performance, it might help students complete their homework. There are also theories associating students' behavior in the classroom with their academic perfor-

(continued)

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.* Note though that impacts on these three measures should be interpreted with caution because all three variables were measured with a single survey item, thus compromising the reliability of the measures.

Further description of the outcome measures can be found in Appendix F.

Implementation Measures

To understand how the interventions were implemented, and whether implementation differed from the first to the second study year, the project team collected data on the use of the instructional models and on the strategies that were used to support the implementation of the models. These measures are briefly described below; greater detail is provided in the chapters that present the implementation findings (Chapter 3 for the math centers and Chapter 7 for the reading centers).

Use of Special Instructional Models

Three different aspects of teachers' implementation of the after-school instructional models were assessed:

- **Use of instructional elements.** In order to examine whether teachers used all the intended materials and instructional methods, information on the use of instructional elements was obtained both years from structured protocol observations of implementation conducted by local district coordinators.[29] Factors recorded on a check-off list by the district coordinators indicate to what extent teachers covered specific core content and instructional strategies of the enhanced program.

- **Pacing of daily lesson plans.** In order to assess whether teachers were able to keep up with the intended pace of the enhanced program model during a class period, measures of the prevalence of pacing problems were collected by and created by the research team. In the first study year, measures of pacing were collected from structured protocol interviews of half the teachers (randomly sampled) in the enhanced after-school program. In the second

mance. One 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). A competing hypothesis is that lengthening academic instruction introduces fatigue and induces a student to act out during class.

[29]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. No formal measure of reliability was computed for these data. (See Appendix E, Boxes E.1 and E.2.)

year, structured protocol interviews of district coordinators were conducted by the research team to learn whether, in the second year, teachers were better able to get through the material in each session.[30]

- **Pacing of the instructional content.** In the second year of the study, data from Class Record Forms created by Harcourt School Publishers were used to assess whether staff were spending, on average, three days instructing students on the same skill, as intended by the program developers. In particular, a measure of the average "Instruction days per skill assigned" was calculated given the total number of days a student attended the program and the total number of skills assigned to that student during the school year.

Strategies Used to Implement the Models

Data were also collected during both implementation years on the strategies used to support the implementation of instructional models:

- **Staffing and support for instructors.** The staffing strategy and support for instructors in the enhanced program are evaluated using data primarily drawn from the survey of the after-school program staff (enhanced program teachers). These data are used to examine whether sites hired certified teachers and operated the programs with the intended small groups of students (approximately 10 students per instructor). These data are also used to assess whether instructors received upfront training, continued support, and daily paid preparation. Additionally, data gathered by Bloom Associates, Inc., are used to report on teacher turnover.

- **Amount of instruction offered.** To measure the intensity of the program, responses from the survey of after-school staff were used to calculate how many minutes of instruction were offered each week. Additionally, in order to assess the amount of instruction being offered over the course of the school year, a measure was created that combined the number of days over the course of the school year that the enhanced program was offered, with the number of minutes of instruction offered each week.

Service Contrast Measures

To measure the differences between the services received by students randomly assigned to the enhanced program group and the services received by students assigned to the

[30]Recall that district coordinators were responsible for supporting staff members in the enhanced program.

regular program group, the project team collected data on various aspects of the service contrast during both program years:

- **Service offerings.** The survey of after-school program staff is used to describe the characteristics of staff in each type of after-school program (enhanced and regular), in terms of their qualifications and experience, as well as the support provided to them. The responses of regular program staff are also used to evaluate the nature of the services offered in the "business as usual" setting (i.e., whether the regular after-school program focused on math, reading, or mixed subjects and whether the help came in the form of homework help, tutoring, or structured academic support). In addition, to further evaluate the "business as usual" setting in the regular-program group, two randomly selected regular-program-group teachers in each after-school center were interviewed in the second year.

- **Overall attendance in the after-school programs.** Attendance data were collected from students in the enhanced and regular program groups for the days on which the enhanced program met, in order to determine whether the enhanced program encouraged students to attend the after-school program more frequently than those in the regular after-school program.

- **Hours of academic instruction received.** The difference in hours of academic instruction received by students in the enhanced and regular program groups lies at the heart of the designed strategy and underlies the enhanced program's impacts. This key aspect of the service contrast is measured by combining two data sources: (1) the attendance of students (enhanced and regular) on the days that enhanced after-school support was provided and (2) survey responses from the regular after-school program staff about whether they provided academic instruction in the subject being tested, rather than homework help, tutoring, or some other approach. For the enhanced program group, all of the time spent in the enhanced program was focused on academic instruction. For the regular program group, hours were counted as "instructional hours" if regular program staff reported on the survey providing academic instruction in the subject being tested.[31]

[31]Total hours for students in the enhanced program group is calculated by multiplying each student's total days of attendance by the length of the enhanced program session (in the first year of implementation: 45 minutes in 14 centers and 60 minutes in one center; in the second year of implementation: 45 minutes in 11 centers and 60 minutes in four centers). Total hours for students in the regular program group is calculated by multiplying the total number of days attended by the length of the enhanced program session (45 or 60 minutes, depending on the

(continued)

- **Other sources of academic support.** Surveys of students and regular-school-day teachers were used to collect information on any additional sources of academic support that students might have received during the regular school day, or outside the regular school day, but not during the enhanced or regular after-school program. The purpose of this data collection effort is to assess whether the service contrast was diluted by any supplemental services that students in the regular program group sought out in response to not having been selected for the enhanced after-school program.

Analytic Methods and Procedures

The experimental impact estimates presented in this report are of the effect of the *intent to treat* students with one year or two years of enhanced services. For this reason, in order to estimate the impact of the enhanced programs on student achievement, it is necessary to compare the experiences of a group of students who were offered the after-school enhanced program with a similar group of students who were offered the regular program. As discussed earlier in this chapter, random assignment was used to determine who would be offered the enhanced program. This creates the expectation that students assigned to the enhanced and regular program are similar on observed and unobserved characteristics prior to the intervention. Because of random assignment, students assigned to the regular program can serve as a benchmark, or "counterfactual," for how students selected for the enhanced program would have performed *had they remained in the regular program*. Thus, any subsequent differences between the outcomes of students in the enhanced and regular program can be fairly attributed to the effect of offering the enhanced program. (For a detailed explanation of how the outcome levels of students in the enhanced and regular program groups are calculated and presented throughout this report, see Box 2.1.)

This section discusses the technical issues related to estimating the impact of offering the enhanced programs on student achievement and other outcomes. First, it discusses the statistical model used to estimate impacts. It then reviews the sample sizes for each analysis and the implications for statistical power.

---

center), 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 students in that center is zero. Note that staff reports of academic instruction are subject to recall and other biases.

Box 2.1

Description of the Calculation and Presentation of Outcome Levels

Throughout the report, when a table is presented to report estimated program impacts, the mean outcome levels for the enhanced and the regular program groups are reported, to provide context for interpreting the estimated differences. Program impacts are estimated using an impact regression model that uses 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 impact regression model. In other words, means for *both* groups are "regression-adjusted" using a 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, and
- 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, which makes it possible to compare these actual values with those for other reference groups or for the same group of students over time. The reported mean outcome level for the regular after-school program group also has a straightforward interpretation: it provides an unbiased estimate of how the enhanced program group students would have performed had they not been assigned to the enhanced program. In other words, it represents the "counterfactual."

Throughout the text of this report, when presenting these outcome levels, the *observed* mean level for the enhanced program group is referred to as the ''enhanced program group" mean. The mean value for the counterfactual, or the *regression-adjusted* mean for the regular program group, is referred to as the "regular program group" mean. In addition, observed means (adjusted only for randomization strata) for both the enhanced program group and the regular program group are included in Appendix G and Appendix H, Tables G.3, G.6, H.3, and H.6.

Primary Impact Analyses

Statistical Model and Presentation of Impacts

All of the impact analyses use ordinary least squares (OLS) regression to estimate the difference in outcomes between students in the enhanced and regular program group, adjusted for random assignment strata. In order to improve the precision of the impact estimates, the analysis also controls for differences between the enhanced and regular group in their prior achievement levels and the following student characteristics: individual-level pretest measures, 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. Because centers were selected purposefully and are not a random sample of a larger population of centers, the analyses do not attempt to statistically generalize the results beyond the 27 after-school centers in the study. Details on the statistical model can be found in Appendix G and Appendix H. For the purposes of this report, statistical significance is indicated in the tables by an asterisk (*) when the p-value of the impact estimate is less than or equal to 5 percent.

In order to help the reader interpret the findings, impact estimates are presented both in their original metric and in effect-size units. Effect sizes provide an indication of the magnitude of the impact estimates relative to the overall variation in the outcome of interest for students in the study sample. For the purposes of the impact analysis, effect sizes are calculated as a proportion of the standard deviation of the outcome for students in the regular program group at follow-up. The standard deviation for the regular program group reflects the expected variability in the outcome of interest that one would find in the absence of the enhanced program. The impact effect size, therefore, provides an indication of how much the enhanced program moved students along this variability in expected performance.

Where there are multiple outcomes for the same sample of students, a multiple comparisons adjustment will be applied using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). In particular, this adjustment will be applied to the two-year reading sample that examines two reading outcomes, SAT 10 scores and DIBELS oral reading fluency. Note that the SAT 10 total score is the qualifying measure for the subtests so the subtests are not included in this test of multiple comparisons. Additionally, no adjustments are made for any of the math samples as the SAT 10 is the only academic outcome.

Secondary and Exploratory Analyses

Impacts on several secondary outcomes are also examined, using the same samples and statistical models described above. This includes impacts on students' homework completion and other in-school behaviors, as well as impacts on locally administered standardized tests.

In addition, the report presents findings from two sets of non-experimental exploratory analyses that were conducted for the purpose of examining questions that cannot be answered within the randomized experiment. Note that these two sets of exploratory analyses are not based on the experimental design of the study and may not reflect true causal relationships.

The first analysis examines the association between *receiving* two consecutive years of enhanced after-school services and student achievement. Recall that some students who were assigned to two years of the enhanced program did not participate in the program for a second year. Note, however, that the number of years of enhanced services that students receive could be related to their experience in the enhanced program in the first year of the study. For example, students who chose to receive enhanced services for two school years (i.e., applicants in the E_1E_2 group) may be those who felt that they particularly benefited from the enhanced program in the first year. Conversely, students who chose to receive only one year of enhanced services (i.e., nonapplicants in the E_1E_2 group) could be students who felt that they did not benefit at all from the enhanced program in the first year. In other words, students *self-select* themselves into different amounts of enhanced instruction. As a result of this self-selection, students in the R_1R_2 group (who did *not* receive enhanced services) may no longer provide the right counterfactual for what would have happened to students who received two years (or one year) of enhanced services in the absence of the enhanced program. Nor is it possible to identify which students in the R_1R_2 group would have made similar participation decisions had they been invited to enroll in the enhanced after-school program in the first year. Thus, using an instrumental variables approach, the first exploratory analysis makes adjustments for enhanced program students in the two-year sample who did not attend the program during any of the second year. Details on the analysis are provided in Appendix I.

The second exploratory analysis uses both Cohort 1 and Cohort 2 samples and examines whether the impact of offering students the opportunity to participate in one year of enhanced services, either during the first or second study year, is associated with particular school or implementation characteristics. *A priori*, impacts were hypothesized to be greater in centers where: the staff turnover is less, the service contrast is greater, the program's instructional approach to the subject was similar to that use during the school day, students were receiving fewer hours of school-day instruction in the subject, the student-teacher ratio after school was smaller than that during the school day, the students were needier, and when the quality of the school day instruction was not sufficient to allow it to meets its Adequate Yearly Progress (AYP) goals.

The three measures of *program implementation* included in the correlational analyses are: whether one or more teachers teaching the enhanced program left during the school year (included as a measure of disruption in instruction); the number of days over the course of the school year that the enhanced math program was offered (included as a measure of program

dosage); and the difference between the total hours of after-school academic instruction received by students in the enhanced program group relative to students in the regular program group (a measure of service contrast). Five measures of the *local school context* capture the characteristics of the regular school day, as well as the characteristics of the school's student body: the instructional approach of the school-day curricula (available for the math sample but not for the reading sample);[32] how much time is spent in the regular school day on instruction in math or reading;[33] whether the school meets its Adequate Yearly Progress (AYP) goals;[34] whether the in-school student-to-teacher ratio is greater than the student-teacher ratio in the after-school program (13:1);[35] and what proportion of students in the school receive free or reduced-price lunch (a measure of the students' neediness).[36,37]

The analysis is based on an impact model that includes a set of interactions between treatment status (i.e., whether a student was assigned to the enhanced or regular program) and various school and program characteristics. The coefficient on these interactions represents the

[32]Students who are struggling during the school day may benefit from an alternative instructional approach after school. This information is not available for the reading sample because 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.

[33]Additional time in math or reading may have a greater benefit for students who spend less time on this topic during the school day.

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

[35]The planned student-teacher ratio was 10:1; however, up to 13 students were randomly assigned to each class, in order to account for the possibility that some students might not attend on a given day.

[36]Data on the student-teacher ratio and the proportion of student receiving free or reduced-price lunch come from the National Center for Education Statistics' Common Core of Data (CCD), which compiles school-level demographic information. At the time of writing, 2006-2007 data (corresponding to the second year of the study) were not yet available. Given that these two characteristics are unlikely to have changed substantially in one year, schools in the second year of the study were assigned their value from the prior year (2005-2006).

[37]Three additional school-level measures were available for the second year of program implementation. The first is the average yearly achievement gain of students in the school, which serves as a proxy for the level and quality of instruction and leadership at the school.

The second measure is the percentage of enhanced program teachers in the second year of the study who also taught during the first year (i.e., "returning" teachers). This measure is intended to gauge program implementation strength, since one would expect returning teachers to be better able to deliver the enhanced curriculum than new teachers.

The analysis based on math centers also includes a third additional measure: an indicator of whether, on average, students in the enhanced program spent fewer than four days on each math skill pack assigned by the teacher (where four days is the center-level average in the sample). This indicator serves as a measure of teachers' instructional pacing.

Given the availability of these additional measures, a separate analysis was conducted focusing on the second year of the study only (i.e., 15 center-level impacts in the Cohort 2 sample for math and 12 center-level impacts in the sample for reading) and using all available school-level characteristics in the second year of the study.

association between impacts and the school and program characteristics. Details on the statistical model and measures of school and program characteristics are provided in Appendix J.

Sample Sizes and Statistical Power

An important goal of the study design was to ensure that the sample size would be sufficient to enable the study to detect program effects of reasonable magnitude (if they exist). The number of students in the sample is a crucial factor that determines the degree to which the impacts on student achievement and other outcomes can be estimated with enough precision to reject with confidence the hypothesis that the program had no effect. In general, larger sample sizes provide more precise impact estimates. A common way to represent statistical precision is through the "minimum detectable effect size" (MDES). Formally, the MDES is the smallest true program impact (scaled as an effect size) that can be detected with a reasonable degree of power (80 percent) for a given level of statistical significance (5 percent).

The MDES for each analysis sample used in the impact analyses are presented below, with additional details on these MDES calculations provided in Appendix B. These analysis samples are limited to students with data on both the follow-up SAT 10 assessment and the regular-school-day teacher survey.[38] Analysis that eliminates this second inclusion criterion, thereby increasing each sample by between one and 18 students, are presented in Appendix G.

Impact of offering students one year of enhanced services

In the math centers, the sample for the analysis includes 1,144 students in Cohort 1 and 792 students in Cohort 2. For the Cohort 1 sample, the study can detect one-year impacts of 0.10 standard deviation or larger and, for the Cohort 2 sample, 0.15 standard deviation or larger. This translates into an impact of 3.9 and 5.9 scaled score points on the SAT 10 total math test for Cohorts 1 and 2, respectively. For Cohort 1, this is equivalent to 22 percent, and, for Cohort 2, 33 percent of the expected improvement of students in grades two through five nationally.[39]

[38]These instruments were administered at the end of each implementation year. See Appendix C (math) and Appendix D (reading) for details on response rates and the characteristics of students in the analysis samples.

[39]The expected annual growth in average SAT 10 total math scores for a nationally representative sample of students (based on normed data from the test developers) with the same grade composition as the one-year samples is 18 scaled score points (this expected growth is weighted to reflect the distribution of students across grades in the cohort samples combined). Specifically, a weighted average of fall scores of nationally representative second-, third-, fourth-, and fifth-graders is calculated where the weights are the proportion in the one-year sample that were in these grades at baseline. This weighted average is subtracted from the weighted average of spring scores of nationally representative second-, third-, fourth-, and fifth-graders (the weights are the same as before).

In the reading centers, the sample for the analysis includes 905 students in Cohort 1and 626 students in Cohort 2. For the Cohort 1 sample, the study can detect one-year impacts of 0.11 standard deviation or larger, and, for the Cohort 2 sample, it can detect impacts of 0.14 standard deviation or larger. This translates into an impact of 4.3 and 5.5 scaled score points on the SAT 10 total reading test for Cohort 1 and 2, respectively. For Cohort 1, this is equivalent to 45 percent, and, for Cohort 2, 57 percent of the expected improvement of students in grades two through five nationally.[40]

Impact of offering students two years of enhanced services

The two-year sample for the analysis includes 367 students in the math centers and 270 students in the reading centers. Thus, the study is equipped to detect two-year impacts of 0.21 standard deviation or larger for the math program and 0.23 standard deviation or larger for the reading program, approximately double the impact in the first year in each subject area. To put these findings in context, the test score growth for a nationally representative sample of students with the same grade composition in each period as the two-year sample is also presented. However, no systematic statistical analysis was performed to test the significance of differences between the study sample and the nationally representative sample.

[40]The expected annual growth in average SAT 10 total reading scores for a nationally representative sample of students (based on normed data from the test developers) with the same grade composition as the one-year samples is 9.6 scaled score points. Again, as stated above, this expected growth is weighted to reflect the distribution of students across grades in the samples.

Chapter 3

Implementation of the Enhanced After-School Math Program

This chapter begins by describing the 15 after-school centers that implemented the enhanced math instruction for both years of the evaluation. It then presents the intended design of the enhanced math instruction and the implementation findings for both the structural and instructional elements of the program.

Centers in the Math Study Sample

Table 3.1 presents the characteristics of schools in school year 2005-2006 that house the 15 after-school centers that implemented the enhanced math program over two school years. As shown in this table, six schools are located in a large or midsize city, five are within the urban fringe of a large or midsize city, and four are in a large or small town or rural area. Four of the 15 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 2006-2007.[41] Slightly less than 40 percent of the students in the schools are black (38 percent), approximately one-third (35 percent) are white, 22 percent are Hispanic, 3 percent are Asian, and approximately 1 percent are American Indian.[42] While the types of communities surrounding these centers vary, 69 percent of all students in these schools come from low-income families.[43] The average student-to-teacher ratio in these schools is 15:1.

During the regular school day, students in 10 of the 15 schools received 60 minutes or less of math instruction, with five schools offering more than 60 minutes (see Table 3.2).[44] In all of these schools, the school-day instructional approach varies. Eight schools in the study sample use an instructional approach during the day that has a format of math topic sections within

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

[42]Rounding may cause slight discrepancies in calculating sums and differences.

[43]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).

[44]School administrators were asked how many minutes teachers spend per day teaching math 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 of offering 60 minutes.

The Evaluation of Academic Instruction in After-School Programs

Table 3.1

Characteristics of Schools Housing After-School Centers Implementing the Enhanced Math Program

Characteristic	
Number of schools	
School setting[a]	
Large or midsize city	6
Urban fringe of a large or midsize city	5
Large or small town, or rural area	4
Schools not making Adequate Yearly Progress (AYP) goals	4
Composition of student body	
Race/ethnicity of students (%)	
Black	38.04
White	35.41
Hispanic	21.73
Asian	2.53
American Indian	0.52
Low-income students[b] (%)	69.21
Average student-to-teacher ratio	15:1
Sample size (total = 15)	

SOURCES: All school-level characteristics were collected from the Common Core of Data (CCD) Web site, except for AYP status, which was collected from each state's Department of Education Web site. CCD data reflect the 2005-2006 school year (the first year of implementation), which is the most recent year for which data are available. AYP status data reflect the 2006-2007 school year.

NOTES: The composition of the student body is calculated by averaging the proportion of students within each school 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.

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 the Enhanced Math Program

Regular-School-Day Characteristic	Number of Schools
Minutes of math instruction offered	
Number of schools with 60 minutes or less	10
Number of schools with more than 60 minutes	5
Math materials/curricula[a]	
Everyday Mathematics (Wright Group/McGraw-Hill)	
Harcourt	
Houghton Mifflin Math	
McGraw-Hill	
Saxon	
Scott Foresman-Addison Wesley Mathematics	
Sample size (total = 15)	

SOURCES: Data were collected from research staff interviews with point persons and phone calls made to schools and districts in spring 2007 in regard to the 2005-2006 school year (the first year of implementation).

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.

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

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 either 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) or that employs a direct instructional approach organized by lessons with spiraled curriculum (for example, Saxon).

The Enhanced After-School Program Instructional Model

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: numbers and operations, measurement, geometry, algebra and functions, and data analysis and probability. 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. 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 two through five. 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. 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. 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. In each 15-minute workout rotation, the teacher works in a small group with two to three students on a specific math topic or skill to begin a skill pack, 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.

Implementation Findings

This section presents the implementation findings for both the structural and instructional elements of the program and the implementation challenges encountered. As described in Chapter 2, it draws on surveys of after-school program staff involved in its operation, conducted by the research staff; structured protocol observations of implementation of Mathletics, conducted by district coordinators; interviews with district coordinators and teachers of the enhanced after-school program, conducted by the research staff; and attendance records.

Implementation findings are presented by implementation year in Table 3.3. Additionally, as after-school teachers and centers became more experienced with the delivery of the intervention, program implementation may have improved. Thus, this section also examines whether implementation differed between the two years of the study. In instances where implementation did not differ between the two years and findings for each year are presented in Table 3.3, only *first* implementation year findings are discussed in the text.

The Evaluation of Academic Instruction in After-School Programs

Table 3.3

Characteristics of and Support for Enhanced Math Program Staff

Service Offering	Year 1	Year 2	Difference	P-Value for the Estimated Difference
Structural Elements				
Staffing				
Certified in elementary education (%)	98.36	95.59	2.77	0.18
Years of elementary school teaching experience (%)				
No experience	0.00	4.41	-4.41	
1-2 years	9.84	7.35	2.48	
3-4 years	13.11	11.76	1.35	
More than 4 years	77.05	76.47	0.58	
			chi-square	0.94
Staff-youth ratio (youth enrolled)	8.67	9.03	-0.36	0.20
Staff-youth ratio (actually attended)	8.09	8.58	-0.49	0.08
The Amount of Instruction Offered				
Hours of instruction offered	75.13	72.49	2.64	0.40
Support for Staff				
High-quality training to carry out activity (%)				
Very true	66.67	81.16	-14.49	
Sort of true, not very true, or not at all true	33.33	18.84	14.49	
			chi-square	0.13
Had enough materials and equipment to carry out work (%)				
Very true	67.21	75.36	-8.15	
Sort of true, not very true, or not at all true	32.79	24.64	8.15	
			chi-square	0.23
Amount of paid preparation time to carry out activity (%)				
No minutes to less than 30 minutes per day	8.47	10.29	-1.82	
30 or more minutes per day	91.53	89.71	1.82	
			chi-square	0.73
Ongoing support from district for how to teach children in activity (%)				
Very true	75.00	88.24	-13.24	
Sort of true, not very true, or not at all true	25.00	11.76	13.24	
			chi-square *	0.04

(continued)

Table 3.3 (continued)

Service Offering	Year 1	Year 2	Difference	P-Value for the Estimated Difference
Instructional Elements				
Teachers' Assessment of the Content of the Program				
Materials were appropriate for students (%)	91.80	98.55	-6.75	0.09
Material difficulty (%)				
At about the right level of difficulty	85.71	91.30	-5.59	
Too easy	8.93	4.35	4.58	
Too challenging	5.36	4.35	1.01	
			chi-square	0.13
Sample size (total = 130)	61	69		

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

NOTE: Percentages are based on the number of staff who responded to the question.

Structural Elements[45]

The implementation of Mathletics was supported using a set of strategies related to staffing, instructional hours, and support for instructors. These strategies were utilized in both years of the study as intended, but some were provided with less intensity in the second year. Following is a description of these implementation strategies and how they were implemented.[46]

[45]Findings in this section are largely drawn from the After-School Staff Survey, which was completed at the midpoint of both school years by all staff providing academic support to students in the participating after-school centers to gain information about instructors' impressions of and interactions with the intervention. The staff surveys were given to all teachers in the second year, regardless of whether it was their first or second time teaching in the enhanced after-school program. In the first year, 90 percent of staff (61 of 68) responded to the survey; in the second year, 99 percent of staff (69 of 70) responded to the survey. Among the staff responding to the survey, not all staff answered every question. Throughout this section, percentages are out of the 61 staff in the first year or 69 staff in the second year who responded to the survey, unless indicated otherwise.

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

Staffing

There are two key staffing strategies: (1) hiring certified teachers as instructors, with a preference for experienced teachers who also are able to make a full-year commitment to the program, and (2) establishing 10:1 student-to-teacher ratios for instruction. Additionally, when the study was extended to include a second year of program operations, every effort was made to recruit back staff from the first program year.

Based on responses to the survey of after-school staff, certified teachers with experience were hired as intended. And centers across both years did not statistically differ in the proportion of certified staff and staff with varying degrees of experience. Specifically, in the first year, 98 percent of Mathletics instructors were certified teachers, and 77 percent of teachers had more than four years of elementary school teaching experience.

In both implementation years, random assignment was conducted in a manner to produce enhanced program groups of 10 to 13 students per grade, which allowed for some attrition and absences and still maintain an average class size of 10 students. When surveyed, Mathletics instructors in both years 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 again reported an average of nine students in both years.

While there was teacher turnover within each of the implementation years, comparatively more teacher turnover occurred across implementation years. Specifically, of the 68 teachers hired at the beginning of the first school year, there were three instances of teachers leaving before the end of the school year. In the second year, of the 70 teachers hired, 10 staff from eight centers left before the end of the school year.[47] Thus, at least 85 percent of the teachers remained teaching in the program within a given program year (4 percent left in the first year and 15 percent left in the second year).[48] However, at the beginning of the second school year, of the 70 teachers hired, 40 staff were returning to the program for a second year, while the other 30 second-year staff were new to the program. Thus, about 60 percent of staff in the first year (40 out of 68) returned to teach in the program for a second year.

The Amount of Instruction Offered

The intended amount of instruction was 180 minutes per week, either in four 45-minute lessons or in three 60-minute lessons. On average, the program was implemented each year

[47]Among the 13 who left, reasons for leaving included: to get a masters degree; conflict with their supervisor; did not work well with the math curriculum; and personal reasons.

[48]The difference between the number of teachers that left within the first year and the number of teachers that left within the second is statistically significant (p-value = 0.048).

with, at a minimum, this intended amount of instruction. In the first year of implementation, the after-school program staff teaching Mathletics reported on the staff survey that they offered an average of 189 minutes of instruction per week, a statistically significantly greater amount than the 180 minutes intended (p-value = 0.00). In the second year, the program staff reported offering an average of 171 minutes of instruction per week, not statistically significantly different from the amount as intended (p-value = 0.45)

Across the entire school year, the total hours of enhanced after-school instruction offered does not statistically differ between the two implementation years (p-value = 0.40). Specifically, in the first year the program was offered on average for 75 hours, whereas in the second year it was offered on average for 72.5 hours.

Support for Staff

Enhanced program instructors received the intended training and support in a variety of ways throughout both school years. In both years, all the instructors (68 in the first year and 70 in the second year) were hired in time to attend the summer training on Mathletics prior to the start of the school year, and the training was repeated in the following January for new staff. In the first year, four new math instructors were trained in January during the midyear conference (one replacement for a teacher who left and three new substitute teachers). In the second year, 11 new math instructors were trained (five replacements for teachers who left throughout the year and six new substitutes).[49]

When surveyed, instructors' responses across both years about whether they received high-quality training to carry out their activities did not statistically differ. In the first year, 67 percent of Mathletics instructors reported that it was "very true" that they received high-quality training to carry out their activities.

In the first year, a component of the implementation strategy was to provide staff with all materials needed to teach Mathletics, so they would not be burdened by purchasing supplies. In the second year, this strategy was modified, and sites were asked to pay the cost of replacing all consumable materials. Despite this modification, when asked if the instructors had enough materials and equipment to carry out their work, more than two-thirds of the instructors' responses indicate that enough materials were provided, as intended. And responses did not statistically differ across the two implementation years. In the first year, 67 percent of the instructors reported that it was "very true" that they had enough materials and equipment to carry out their work. The implementation plan also called for 30 minutes of paid daily prepara-

[49]Although 10 teachers left throughout the second year, only five replacement teachers were trained at the midyear conference. The other five either were replaced by substitutes or they did not leave during the fall, so replacements were brought in after the January training.

tion time, and, again, reports indicate this was provided as intended, with no statistically significant differences across the two years. Specifically, 92 percent of instructors in the first year reported that they had 30 minutes or more of paid preparation time each day.

However, interviews with teachers conducted just in the first study year suggest that the 30 minutes of prep time was not always sufficient for developing individual plans for each child and deciding which children should be grouped together for the following day's 15-minute rotations. Specifically, because assignments are determined daily, 17 of the 30 teachers interviewed reported that it was difficult "at least some of the time" for them to accomplish the necessary preparation within the 30-minute paid preparation period during the afternoon prior to instruction.[50] 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.

The project also provided ongoing, on-site technical assistance. As outlined in Chapter 1, in the first year this consisted of 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 (two on-site visits during the school year and weekly conversations by phone). In the second year, on-site technical assistance was provided less intensively. A district coordinator continued to support implementation, and Bloom Associates provided assistance through two site visits and biweekly phone calls. However Harcourt School Publishers chose not to visit the sites. Despite this lessening in support, when asked whether they received ongoing support on how to teach children in Mathletics, second-year staff were more likely than first-year staff to report receiving ongoing support (p-value = 0.04). In the first year, of the 60 instructors responding to the survey question, 75 percent reported that it was "very true" that they received ongoing support on how to teach children in Mathletics. In the second year, of the 68 instructors responding to the question, 88 percent said that it was "very true."

Instructional Elements

The project team collected data on the teachers' assessments of the content of the program and on three different aspects of teachers' implementation of the Mathletics program: use of instructional elements, the pacing of daily lesson plans, and the pacing of the instructional content of the program.

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

Teachers' Assessment of the Content of the Program

Staff were asked in both years whether the Mathletics materials were appropriate for their students. Across the two implementation years, overall staff responses did not statistically differ. In the first implementation year, 92 percent of staff reported it was "true" that materials were appropriate for their students, and 86 percent 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 5 percent saying "too challenging."

Use of Instructional Elements

Under the guidance of Bloom Associates staff, local district coordinators conducted structured protocol observations of implementation of Mathletics classes in each center three times, on average, over both the first and second school years. The protocol included a checklist of the following six components: use of the Mathletics materials throughout the instructional period, establishment of routines that allow for smooth transitions between the parts of the instructional session, use of workout components (such as skill packs) appropriately, provision of direct and differentiated instruction during the workout, inclusions of all the components in the allocated times, and inclusion of a teacher-led warm-up and cool-down for all students. Each year, researchers obtained from the district coordinators overall scores that consisted of the total number of Mathletics components present during that observation. In order to create an aggregated rating for the class, the scores of each class's observations were averaged. Across the two implementation years, aggregated ratings did not statistically differ (p-value = 0.32).

In the first year, on average, 98 percent of the observed classes' aggregated ratings showed that the instructor implemented, on average, between five or six of the six components. In the second year, all of the observed classes' aggregated ratings showed that the instructor implemented, on average, between five or six of the six components. In addition to each observation's overall score, researchers in the second year received the component implementation checklist from the district coordinator's observation records (for more details see Appendix F). This component checklist shows that of the 182 individual classroom observations conducted by district coordinators, all six components were implemented 91 percent of the time (165 observations).

Pacing of Daily Lesson Plans and Instructional Content

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. In the first year of the study, as part of the field research, two randomly selected teachers in each center (half of all math teachers in the evaluation) were interviewed and asked about pacing

issues. In the second year of the study, research staff conducted interviews with district coordinators about implementation challenges.

As part of the teacher interview in the first year, each teacher was asked, "Can you get through all the material you need to in each session?" Fourteen of the 29 teachers responding to the question indicated experiencing some challenges related to pacing (one teacher did not respond to the question). Their responses were categorized as follows: four (14 percent of 29) described pacing as a "consistent problem" and said that, as a rule, they had trouble completing the daily lesson in the allotted time. Seven (or 24 percent) indicated that pacing was "sometimes a challenge," whereas three (10 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. The remaining 52 percent of the teachers indicated that they were able to cover the material in the allotted time and that pacing was "rarely a problem" for them. Among the 14 teachers who reported that pacing was a challenge at least at some point throughout the first year, the most frequently cited challenge was the instructional rotation time.[51]

In the second year of the study, all nine district coordinators were interviewed and were asked whether "finishing direct instruction in one rotation" continued to be a problem for staff. Of the eight district coordinators responding to the question, four said finishing instruction on time was a challenge for staff again in the second year, and four said that teachers found it less challenging this year to complete instruction in the 15 minutes of a rotation (one district coordinator did not answer the question).

Additionally, in their training, instructors were told that the program developers recommended students spend approximately three days on the same skill, then move onto a new skill.[52] In the second year of the study, to determine whether students moved through the academic content of the Mathletics program at this recommended pace, the average number of instructional days per skill assigned was calculated, given the total number of instructional days a student attended and the total number of skills assigned to that student throughout the school year.[53]

[51]The 14 teachers were asked to identify what, in particular, they found challenging. Seven reported that the 15-minute rotation time did not always allow enough time for students to master the skill or concept. Three of these seven pointed out that the rotation time was especially insufficient for the "struggling" students (that is, students who were characterized by teachers as lower performers).

[52]Teachers were encouraged to move students on to the next skill, after trying multiple instructional methods, rather than getting bogged down for weeks on one skill that a student might not be developmentally ready to master.

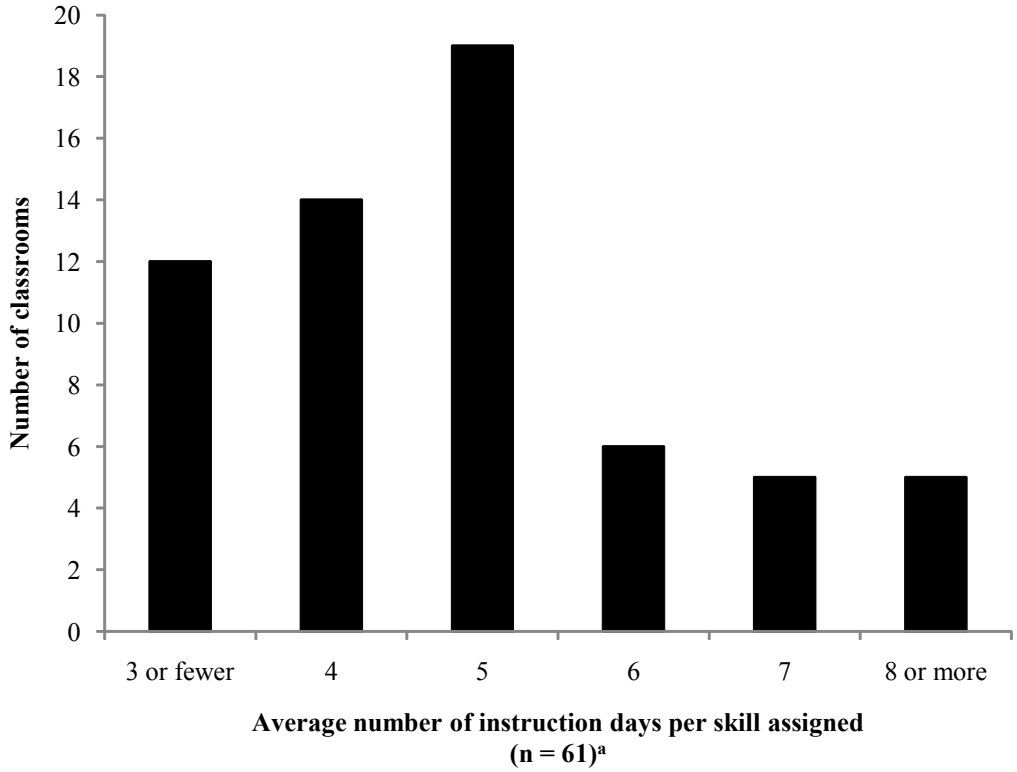
[53]Harcourt School Publisher created a "Class Record Form" as a management tool to help teachers track student progress through the skills. As part of this form, teachers document which skills are assigned to each student over the course of the year (see Appendix E).

As shown in Figure 3.1, within each classroom, the average number of days students spent working on an assigned skill ranged, with about half (33 classrooms) spending four or five days on an assigned skill. This included time spent receiving direct instruction from the teacher (individually or in small groups of two or three), completing practice activities in the skill packs, utilizing the computer-assisted instruction or computer games that reinforced the skill, or playing board games that offered students more time to practice the math skills. Across 63 classrooms, the median number of days that students spent on a skill was 4.5 (mean of five days), which is 1.5 days longer than the three days per skill recommended by the program developers.

With the available data, it is not clear whether the slower pace arose from the educational needs of the children served or from teachers who were reluctant to have students leave one skill area without achieving mastery, but over time this difference aggregated to slower than intended progress through the material.

Figure 3.1

Average Number of Math Instruction Days per Skill Assigned, by Classroom (Second Year of Implementation)



SOURCES: All classroom-level characteristics were collected from the Harcourt School Publishers' Class Record Forms regarding the second implementation year.

NOTES: The number of instruction days per skill assigned assumes that a "day" is 45 minutes of instruction. If a class met for over 60 minutes, each "day" was adjusted by (4/3).

[a]Two classrooms are not included in these calculations because data was not available on average number of instruction days per skill assigned.

Chapter 4

Analysis of the Offer of One Year of Service in Math: Sample Characteristics, Service Contrast, and Impacts

The primary focus of the Evaluation of Enhanced Academic Instruction in After-School Programs is to assess the impact of the enhanced after-school programs on student achievement. The present chapter focuses on the first two research questions for the 15 centers implementing the enhanced math program for two years:

- What is the impact of offering students the opportunity to participate in the enhanced math program *for one school year*?
- Is this impact different in the second year of program implementation than in the first year?

These two questions are answered by comparing the outcomes of students who were randomly assigned to participate in the enhanced after-school math program for one school year with the outcomes of students who were randomly assigned to remain in the regular after-school program during that same school year. Impacts are estimated for each year of implementation separately and then compared.

Before presenting the impact findings, however, the chapter begins by providing two key pieces of background information. First, the chapter provides a brief description of the sample of students included in this analysis. Then, in order to contextualize the magnitude of the impact findings, the chapter provides a comparison of the academic services received by students in the enhanced after-school math program relative to students in the regular after-school program — that is, the service contrast.

Characteristics of Students in the Math Sample

As explained in Chapter 2, the analysis uses students from two cohorts to examine the impacts of one-year of the enhanced program: Cohort 1 includes students from the first implementation year; Cohort 2 includes students from the second implementation year (see Figure 2.2). Analysis is limited to students with one-year follow-up data from both the evaluation-administered achievement test and the regular-school-day teacher survey.

Table 4.1 presents the baseline characteristics of students in the Cohort 1 and Cohort 2 samples, separately showing students in the enhanced and regular program groups. As seen in this table, except for those in Cohort 1 who didn't provide information about whether or not

The Evaluation of Academic Instruction in After-School Programs

Table 4.1

Baseline Characteristics of Students in the Math Analysis Sample (One Year of Service)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Cohort 1[a]						
Enrollment						
2nd grade	281	156	125			
3rd grade	300	162	138			
4th grade	290	160	130			
5th grade	273	156	117			
Total	1,144	634	510			
Race/ethnicity (%)						
Hispanic		29.06	24.80	4.26	0.09	0.08
Black, non-Hispanic		38.84	41.64	-2.80	-0.05	0.21
White, non-Hispanic		25.59	26.07	-0.48	-0.01	0.82
Asian		1.26	2.15	-0.88	-0.06	0.26
Other		5.21	5.29	-0.08	0.00	0.95
Gender (%)						
Male		47.16	43.12	4.04	0.07	0.17
Average age (years)		8.63	8.65	-0.01	-0.02	0.65
Overage for grade[b] (%)		16.88	17.03	-0.15	0.00	0.94
Free/reduced-price lunch (%)						
Eligible (among information providers)		77.24	75.32	1.92	0.04	0.36
No information provided		3.15	1.30	1.85 *	0.11	0.02
Average household size		1.96	1.87	0.09	0.08	0.14
Single-adult household (%)		33.00	35.15	-2.15	-0.04	0.43
Mother's education level (%)						
Did not finish high school		16.72	15.06	1.66	0.04	0.46
High school diploma or GED certificate		31.23	32.08	-0.85	-0.02	0.76
Some postsecondary study		44.48	46.01	-1.53	-0.03	0.60
No information provided		7.57	6.85	0.73	0.03	0.61
SAT 10 baseline math total scaled scores		567.24	565.55	1.70	0.04	0.34
Problem solving		573.09	570.94	2.14	0.05	0.26
Procedures		560.18	558.83	1.36	0.03	0.55
Sample size (total = 1,144)		634	510			

(continued)

Table 4.1 (continued)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Cohort 2[c]						
Enrollment						
2nd grade	256	153	103			
3rd grade	184	105	79			
4th grade	177	100	77			
5th grade	175	103	72			
Total	792	461	331			
Race/ethnicity (%)						
Hispanic		29.34	26.30	3.04	0.06	0.31
Black, non-Hispanic		37.50	38.62	-1.12	-0.02	0.67
White, non-Hispanic		24.80	28.21	-3.41	-0.07	0.21
Asian		1.63	1.40	0.23	0.02	0.81
Other		6.60	5.39	1.20	0.05	0.47
Gender (%)						
Male		42.08	45.84	-3.76	-0.07	0.31
Average age (years)		8.64	8.65	-0.01	-0.02	0.77
Overage for grade[b] (%)		13.60	15.65	-2.04	-0.05	0.44
Free/reduced-price lunch (%)						
Eligible (among information providers)		76.07	75.32	0.74	0.02	0.79
No information provided		3.15	3.01	0.14	0.01	0.91
Average household size		1.97	1.88	0.09	0.09	0.23
Single-adult household (%)		30.11	35.83	-5.72	-0.11	0.08
Mother's education level (%)						
Did not finish high school		17.75	16.43	1.32	0.03	0.63
High school diploma or GED certificate		31.79	30.08	1.70	0.03	0.62
Some postsecondary study		45.40	49.18	-3.79	-0.07	0.30
No information provided		5.06	4.30	0.76	0.03	0.61
SAT 10 baseline math total scaled scores		571.39	570.94	0.45	0.01	0.85
Problem solving		577.89	577.24	0.65	0.02	0.80
Procedures		563.53	562.32	1.21	0.02	0.69
Sample size (total = 792)		461	331			

(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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Table 4.1 (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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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 average observed mean for members randomly assigned to the enhanced program group. The regular program group values in the next column are the average 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined.

F-tests were calculated for the analysis 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.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

[c]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

they receive free or reduced-price lunch, there are no statistically significant differences between the two program groups' baseline characteristics for either of the cohort samples. Additionally, an overall F-test indicates that there is no systematic difference in the background characteristics of students in the enhanced and regular program groups in either of the cohort-specific samples. This supports the notion that, after limiting the sample used for analysis to those with follow-up data from both the evaluation-administered achievement test and the regular-school-day teacher survey, the statistical equivalence of the two research groups is preserved in the sample used for the analysis.

As seen in the first panel of Table 4.1, the majority of students in the enhanced program group within the Cohort 1 sample are black (39 percent) or Hispanic (29 percent). About half of students (47 percent) are male; 17 percent are overage for grade; 77 percent are eligible for free or reduced-price lunch; and 33 percent lived in a household with a single adult. Seventeen percent of students had a mother who did not finish high school, while 31 percent had a mother with a high school diploma or a General Educational Development (GED) certificate. Additionally, students in Cohort 1 are approximately equally distributed across grades (25 percent in

second grade, 26 percent in third grade, 25 percent in fourth grade, and 24 percent in fifth grade). Finally, at their enrollment in the study, 79.5 percent of the students in the Cohort 1 sample were performing at a level defined by the publisher of the achievement test used in this study as below proficient in math.[54]

Characteristics of the students in Cohort 2 are presented in the second panel of Table 4.1. Again, 38 percent of the enhanced program students are black and 29 percent are Hispanic. A little less than half of students (42 percent) are male; 14 percent are overage for grade; 76 percent are eligible for free or reduced-price lunch; and 30 percent lived in a household with a single adult. Eighteen percent of students had a mother who did not finish high school, while 32 percent had a mother with a high school diploma or a GED certificate. However, students in Cohort 2 are not equally distributed across grades (32 percent in second grade, 23 percent in third grade, 22 percent in fourth grade, and 22 percent in fifth grade). This is because Cohort 2 excludes students who were randomly assigned in the second year but were offered the enhanced program in the first year (given that this sample is used to estimate impacts after access to one year of enhanced services) and, by excluding these students, includes a proportionately larger percentage of students in second grade than other grades.[55] Finally, at their enrollment in the study, 73 percent of the students in the Cohort 2 sample were performing at a level defined by the publisher of the achievement test used in this study as below proficient in math.[56]

The Academic Service Contrast Between the Enhanced and Regular After-School Programs

The service contrast, the extent to which the academic support services received by students in the enhanced program group differ from the "business as usual" services received by students in the regular after-school program group, is what produces the estimated impact on student outcomes. Therefore, this section describes the academic support services offered to and

[54]As mentioned in Chapter 2, local staff used a variety of measures to recommend students for the program. However, because performance standards for these measures may differ from those of the study-administered baseline test, 20.5 percent of students in Cohort 1 identified by local staff as in need of supplemental support and randomly assigned into either the enhanced or regular program group tested at or above the proficient level on the study-administered baseline test (SAT 10).

[55]Estimates are weighted to ensure that the second-grade students do not have a disproportionately greater weight in the Cohort 2 findings (see Appendix G for a discussion of these weights).

[56]Again, local staff used a variety of measures to recommend students for the program. However, because performance standards for these measures may differ from those of the study-administered baseline test, 27 percent of students in Cohort 2 identified by local staff as in need of supplemental support and randomly assigned into either the enhanced or regular program group tested at or above the proficient level on the study-administered baseline test (SAT 10).

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 that underlies the impacts is described through five interrelated findings: the content of the service offerings, the experience and training of staff members, overall student attendance in the after-school program, the extent of academic instruction in math, and finally, student academic support from other sources. The following sections present detailed findings on each of these topics, drawing on data from surveys of after-school program staff, attendance records, and surveys of students and regular-school-day teachers.

Differences in Content of the Service Offering

Whether the nature of the content offered to students in the regular program group was different from the support for students in the enhanced program group is explored using responses to the surveys of after-school program staff.[57]

Regular after-school program staff reported providing different types of academic supports to students. Figure 4.1 describes the reported academic services and highlights the type of support that is most similar to the enhanced after-school program — academic instruction in math. In the first year, 47 staff taught the regular after-school program, and, among them, 40 percent reported activities focusing on math. However, 17 percent (eight instructors) of the 47 staff reported providing some form of math instruction beyond tutoring or homework help. Among these eight instructors, six formally assessed students' progress at least monthly, seven indicated using student assessments to guide their instruction, and three indicated providing math instruction using a daily lesson plan and supporting materials — for example, the school-day math curricula, math games and activities, or math books.

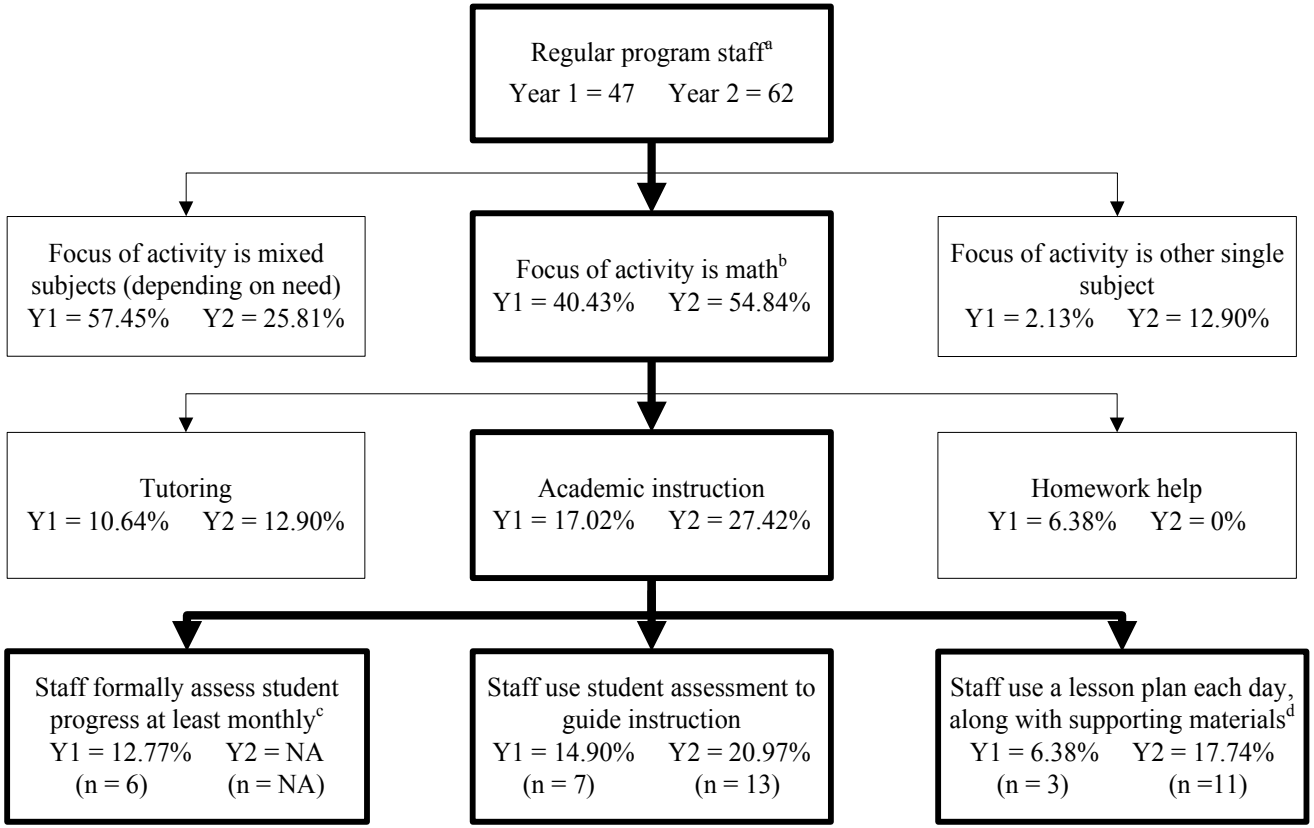
In the second year, 55 percent of the 62 staff teaching the regular after-school program reported activities focusing on math, with 27 percent (or 17 instructors) providing some form of math instruction beyond tutoring or homework help. Among the 17 instructors, 13 indicated using student assessments to guide their instruction, and 11 indicated providing math instruction using a daily lesson plan and supporting materials.

Interviews with a random sample of regular program staff were used to further explore the nature of the academic services provided by these regular program instructors in the second

[57]In the regular after-school program, some staff members provided academic support to students, while other staff members were primarily involved in enrichment or recreational activities. The results presented in this section are based on staff in the former group only. Percentages are based on the number of staff who responded to the survey.

Figure 4.1

Academic Services Offered by Regular After-School Program Staff at Centers Implementing the Enhanced Math Program



(continued)

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

NOTES: Percentages are calculated based on the total number of regular program staff in each year: 47 staff in Year 1 and 62 staff in Year 2.

[a]For Year 2, of the 62 staff who filled out the survey, four (6.45 percent) did not respond to any of these questions.

Figure 4.1 (continued)

[b]For Year 1, of the 19 staff who focus on math, three (6.38 percent) do not provide academic instruction, tutoring, or homework help. They responded that they use another method of helping students. Of the 34 staff in Year 2, nine (15.52 percent) did not indicate providing academic instruction, tutoring, or homework help.

[c]This question was only on Year 1's Evaluation of Academic Instruction in After-School Programs after-school staff survey; thus, values for Year 2 are not applicable (NA).

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

year of implementation.[58] Specifically, of the 17 second-year instructors who reported providing math-focused instruction, 13 were part of the randomly selected staff to be interviewed. As part of the interview, each instructor was asked, "What is the activity's main method of helping students with academic work?" Seven of the 13 instructors interviewed indicated that they provide practice or review of academic material covered during the school day,or help students using assessments; and six (all from the same school district) said they provide formal instruction using a published after-school curriculum, such as "Knowing Mathletics" or "After-School Kidz Math."

Responses to the after-school staff survey as well as the interviews with regular program staff indicate that, when staff reported providing academic instruction in math, they were providing at least one key element of the enhanced afterschool math program — use of a structured after-school math curriculum, frequent assessments to guide instruction, and/or use of a daily lesson plan. Hence, the math instruction that the 17 percent of regular after-school staff in Year 1 and 27 percent in Year 2 indicated they provided was likely similar in nature to the enhanced program, thus dampening the service contrast in the study.[59]

Differences in Staff Providing Academic Support Services

Differences in the staffing strategy and support provided to staff for those offering academic support in the enhanced program group compared with those in the regular program group are also illustrated in the responses to the surveys of after-school program staff.[60]

Characteristics of Staff

Table 4.2 presents information on the characteristics of staff members in the enhanced and regular after-school programs, based on the survey of after-school program staff. As shown in this table, staff members in the two types of program differ on several dimensions.

The top panel of Table 4.2 — which presents the characteristics of staff in the first implementation year — shows that staff members in the regular after-school program were less likely to be certified teachers. Sixty-six percent of regular program staff members were certified

[58]As part of the field research in the second year, two randomly selected regular program instructors in each after-school center were interviewed.

[59]These reports across the two implementation years do not statistically differ (i.e., 17 percent is not statistically different from 27 percent, p-value = 0.14).

[60]In the regular after-school program, some staff members provided academic support to students, while other staff members were primarily involved in enrichment or recreational activities. The results presented in this section are based on staff in the former group only (which includes 47 staff from the first year and 62 from the second year). Percentages are based on the number of staff who responded to the survey.

The Evaluation of Academic Instruction in After-School Programs

Table 4.2

Characteristics of After-School Staff at Centers Implementing the Enhanced Math Program

Service Offering	Enhanced Program	Regular Program	Estimated Difference		P-Value for the Estimated Difference
First implementation year					
Certified in elementary education (%)	98.36	65.96	32.40	*	0.00
Years of elementary school teaching experience (%)					
No experience	0.00	17.39	-17.39		
1-2 years	9.84	21.74	-11.90		
3-4 years	13.11	8.70	4.42		
More than 4 years	77.05	52.17	24.88		
			chi-square	*	0.00
Staff-youth ratio (youth enrolled)	1:9	1:12	-3.27	*	0.01
Sample size (total = 115)	68	47			
Second implementation year					
Certified in elementary education (%)	95.59	64.91	30.68	*	0.00
Years of elementary school teaching experience (%)					
No experience	4.41	7.02	-2.61		
1-2 years	7.35	19.30	-11.95		
3-4 years	11.76	14.04	-2.27		
More than 4 years	76.47	59.65	16.82		
			chi-square		0.19
Staff-youth ratio (youth enrolled)	1:9	1:11	-1.62		0.06
Sample size (total = 132)	70	62			

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

NOTES: This table reflects staff in the first and second year of the study in the 15 centers that implemented the program in both years. 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 any given characteristic varies by as much as 9 for the enhanced program group and 9 for the regular program group due to nonresponse on particular survey items. Staff for whom values are missing are not included in the calculations.

teachers, compared to 98 percent of enhanced program staff. This difference is statistically significant at the 5 percent level.

Regular program staff also had less teaching experience. Fifty-two percent of regular program staff had more than four years of elementary teaching experience (compared with 77 percent of enhanced program staff), while 17 percent had no prior elementary school teaching experience (compared with no enhanced program staff). The overall difference in teaching experience between the two types of program is statistically significant.

The regular after-school program was also characterized by a higher staff-to-youth ratio. The staff-to-youth ratio was 1:12 on average in the regular after-school program, while the enhanced after-school program had an average staff-to-youth ratio of 1:9. This difference is also statistically significant.

The bottom panel of Table 4.2 shows characteristics of staff in the enhanced and regular program for the second implementation year. The differences in characteristics of staff members between the two implementation years are not statistically significant.[61] In the second implementation year, however, staff in the enhanced and regular program statistically differed in terms of their certification levels, but not in terms of their years of experience and the staff-to-youth ratio.

Support for Staff

The top panel of Table 4.3 presents information on the support provided to staff in the first implementation year. As shown in this table, staff in the regular after-school program were less likely than staff for the enhanced program to report having received high-quality training to carry out their work (50 percent and 95 percent, respectively, p-value = 0.00) or to report receiving ongoing support for how to teach children in their Mathletics activity (69 percent and 97 percent, respectively, p-value = 0.00).

Regular program staff members were also less likely to report receiving paid daily preparation time. Eighty-three percent of regular program staff reported getting less than 30 minutes per day, and 17 percent reported getting 30 minutes or more. In comparison, 92 percent of enhanced program staff in the first year received 30 minutes or more of paid preparation time — a difference of 74 percentage points. The overall difference in paid preparation time between the two types of after-school program is statistically significant.

[61] P-values for the test of the difference of service offering measures across implementation years are 0.18, 0.88, and 0.20, respectively, for certification, years of experience, and the staff-to-youth ratio.

The Evaluation of Academic Instruction in After-School Programs

Table 4.3

Support for After-School Staff at Centers Implementing the Enhanced Math Program

Service Offering	Enhanced Program	Regular Program	Estimated Difference		P-Value for the Estimated Difference
First implementation year					
High-quality training to carry out activity[a] (%)	95.00	50.00	45.00	*	0.00
Ongoing support from district for how to teach children in activity[a] (%)	96.67	68.89	27.78	*	0.00
Amount of paid preparation time to carry out activity (%)					
No minutes to less than 30 minutes per day	8.47	82.61	-74.13		
30 or more minutes per day	91.53	17.39	74.13		
			chi-square	*	0.00
Sample size (total = 115)	68	47			
Second implementation year					
High-quality training to carry out activity[a] (%)	98.55	70.37	28.18	*	0.00
Ongoing support from district for how to teach children in activity[a] (%)	100.00	61.82	38.18	*	0.00
Amount of paid preparation time to carry out activity (%)					
No minutes to less than 30 minutes per day	10.29	68.63	-58.33		
30 or more minutes per day	89.71	31.37	58.33		
			chi-square	*	0.00
Sample size (total = 132)	70	62			

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

NOTES: This table reflects staff in the first and second year of the study for the 15 centers that implemented the program in both years. 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 11 for the enhanced program group and 12 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 presents percentages of after-school staff who responded "sort of true" or "very true" when surveyed.

The bottom panel of Table 4.3 shows that this pattern of differences between staff in the enhanced and regular program in the first year is consistent with what occurred in the second year. And the differences between the two years of implementation — with respect to the support provided to staff members — are not statistically significant.[62]

Differences in Attendance and Hours of Academic Instruction in the After-School Program

Table 4.4 presents information on student attendance on the days that the enhanced program operated during the school year, as well as the yearly amount of after-school math instruction received by students. In both years, nearly all students assigned to the enhanced program for one year participated in the enhanced services (fewer than five students attended zero days and received zero hours of instruction). The top panel presents yearly attendance and hours of instruction for the first implementation year (students in Cohort 1) while the bottom panel present this information for the second implementation year (Cohort 2).

In the first implementation year, students in the enhanced program group were offered the Mathletics program for 98 days and attended 78 days, while students in the regular program group attended 61 days of the regular after-school program. This difference of 18 days is statistically significant at the 5 percent level. Attendance for students in the enhanced program group was also statistically higher in the second implementation year (by 17 days), and the difference in days attended between implementation years is not statistically significant (p-value = 0.95).

In the first implementation year, students in the enhanced math program group received 60 hours of after-school math instruction during the school year, while students in the regular program group received 11 hours of after-school math instruction. This yearly difference of 48 hours between the enhanced and regular program group — which is approximately 64 sessions of 45-minutes each — is statistically significant and represents an estimated 30 percent increase in math instruction over and above what is received during the regular school day.[63]

[62]P-values for the test of the difference in measures of support provided to staff across implementation years are 0.27, 0.16, and 0.66, respectively, for "received high quality training," "ongoing support," and "paid preparation time."

[63]This percentage increase is based on information about the number of minutes of school-day math instruction. 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. Hence, the 48 additional hours of math instruction received by students in the enhanced program group represents a 30 percent increase in instructional time in math.

Table 4.4

Attendance of Students in the Math Analysis Sample (One Year of Service)

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]					
Attendance in after-school program[b]					
Number of days attended	78.36	60.79	17.56 *	0.46	0.00
Total hours of math instruction received[c]	59.76	11.41	48.36 *	2.08	0.00
Math support from other sources					
Out-of-school math class or tutoring[d]					
Students receiving instruction (%)	29.65	21.85	7.80 *	0.17	0.00
Number of days per week[e]	1.05	0.60	0.46 *	0.27	0.00
Regular school day[f]					
Students receiving special support (%)	23.92	23.85	0.06	0.00	0.98
Minutes per week of individualized help	49.38	43.36	6.02	0.12	0.50
Sample size (total = 1,144)	634	510			
Cohort 2[g]					
Attendance in after-school program[b]					
Number of days attended	72.10	54.68	17.42 *	0.46	0.00
Total hours of math instruction received[c]	58.40	16.56	41.84 *	1.80	0.00
Math support from other sources					
Out-of-school math class or tutoring[d]					
Students receiving instruction (%)	36.05	26.01	10.04 *	0.22	0.00
Number of days per week[e]	1.30	0.85	0.45 *	0.27	0.00
Regular school day[f]					
Students receiving special support (%)	20.80	26.24	-5.44 *	-0.12	0.05
Minutes per week of individualized help	25.62	28.73	-3.11	-0.06	0.25
Sample size (total = 792)	461	331			

(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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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

Table 4.4 (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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

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

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

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 or 60, 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.

[d]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 five students who did not complete a survey.

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

[f]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.

[g]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

Though the difference in hours received is also statistically significant in the second implementation year (p-value = 0.00), it was larger in the first year of implementation than in the second year (p-value = 0.00). Specifically, for Cohort 1, students in the enhanced program group received 48 more hours of math instruction than students in the regular program group, while for Cohort 2, this difference is 42 hours. This six-hour difference between implementation years is equivalent to the amount of math instruction offered during two weeks of the enhanced after-school program.[64]

The reduction in the difference of instructional hours between the enhanced and regular program groups in the second year of implementation is consistent with reports from regular program staff in the second year of the study providing some form of academic instruction in math, as noted in the prior section on service offerings. Given how hours of instruction is calculated, this had the effect of increasing the number of hours of academic instruction received by students in the regular program group relative to students in the enhanced program group, thereby reducing the service contrast for instructional hours in the second year.[65]

Differences in Academic Support from Other Sources

If students in the regular after-school program group sought out other supplemental math programs outside of school — or received additional help from their school-day teachers — in response to not having been selected for the enhanced after-school program, it would undermine the service difference created in the after-school program. Thus, the second section in each panel of Table 4.4 presents findings for academic support from other nonschool sources and during the regular school day, based on student surveys, as well as surveys of regular-school-day teachers.

On the follow-up student survey, students were asked whether they attended a math class or math-related activity outside of the regular school day that was not part of the after-school program and how many days per week they attended this class or activity.[66] Within each year of implementation, students in the enhanced program group reported a statistically significantly greater amount of participation in a math class or activity outside of school, and the difference between implementation years in participation in a math class or activity outside of

[64]Students in the enhanced program are offered three hours of after-school math instruction per week (4 days * 45 minutes). Thus, six hours of math instruction are offered during two weeks of the program.

[65]As seen in Table 4.4, the reduction in the service contrast for instructional hours in the second year is not explained by a smaller second-year difference between the two program groups in the number of days attended (since this difference was approximately 17 days in both years of the study).

[66]These data are student self-reports of academic support received and are subject to bias inherent in such a method of data collection; however, there is no reason to believe that such bias would differ for enhanced program students compared to regular program students.

school, and days per week of participation, is not statistically significant (p-values are 0.54 and 0.97, respectively). Specifically, in Cohort 1, 30 percent of students in the enhanced program group reported such participation compared to 22 percent of students in the regular program group. The enhanced program group participated in this type of activity 1.05 days per week, on average, while the regular program group participated an average of 0.60 day per week.

Additionally, surveys of regular-school-day teachers of students in the sample asked whether each of these students received "any special support in math during the school day, such as pull-out tutoring, a computer lab, or a special class." Teachers were also asked to report the number of minutes of individualized instruction that they or an aide provided each student in the sample in math during the prior week. For Cohort 1, there are no statistically significant differences in the amounts of individualized instruction received by students in the enhanced and regular program groups, nor is there a statistically significant difference in the percentage of students in each program group who received special in-school support. These findings do not differ by a statistically significant amount across the two years of implementation.[67] However, for Cohort 2, a statistically significantly greater percentage of students in the regular program group compared with the enhanced program group received special in-school support (p-value = 0.05).

Impacts on Student Achievement and Other Outcomes

This section examines whether one year of access to the enhanced after-school math program improves student achievement and investigates whether this impact differs across the first and second year of program implementation. In addition to examining impacts on math achievement, the effect of the enhanced program is also estimated for three teacher-reported academic behaviors for the study sample students: homework completion, attentiveness, and disruptiveness in class. When interpreting these impact findings, the key service contrast finding from the previous section to bear in mind is that students in the enhanced program group received 48 more hours of math instruction than students in the regular program group in the first year, and 42 more hours in the second year.

Impacts on Student Achievement

In the spring of each study year, the Stanford Achievement Test, Tenth Edition (SAT 10), abbreviated battery in math was administered to all students in the sample.[68] Total scores on the math test — as well as scores on two subtests, problem-solving and procedures — are used to measure individual students' academic achievement in math.

[67]P-values for the test of the difference in in-school support services across implementation years are 0.10 and 0.34, respectively, for the percentage of students receiving special support and minutes of individualized help.

[68]Spring 2006 for Cohort 1 and Spring 2007 for Cohort 2.

The top panel of Table 4.5 presents the impact on SAT 10 math scores for students in the Cohort 1 sample. As seen in this table, one year of access to the enhanced after-school math program improved the total math achievement of students by a statistically significant amount. Specifically, the average total math score for students in the enhanced program group is 3.5 scaled score points higher than that of their counterparts in the regular program group, which translates into an effect size of a 0.09 standard deviation.

The first two bars in the top graph of Figure 4.2 places this impact estimate within the context of the actual and expected growth in total math scores for students in the enhanced program group. The dark bar in the graph represents the actual growth of students in the enhanced program group, which was 39.77 scaled score points over the school year. The light bar in the graph represents the growth in test scores for the regular program group; this growth of 36.28 points provides the best indication of what the enhanced program group would have achieved had they not had access to the enhanced after-school math program.[69] Thus, the improvement in test scores that is attributable to the enhanced after-school math program is 3.5 scaled score points. This impact represents a 10 percent improvement over and above what the enhanced program group would have achieved had they not participated in the enhanced program.[70] Assuming that learning is equally distributed across the school year, 10 percent of a nine-month school year is equivalent to 0.9 months of additional learning.

To investigate whether specific types of math knowledge are affected by the enhanced math program, impacts were also examined for the two subtests embedded in the SAT 10 (problem-solving and procedures). As seen in Table 4.5 (and Figure 4.2), students in the enhanced program group had higher scores on average on both of these subtests than students in the regular program group, and the difference is statistically significant for the procedures subtest. Specifically, the enhanced program improved students' procedures scores by 5.8 scaled score points (0.11 standard deviation, p-value = 0.00).

After-school teachers and centers potentially became more experienced with the delivery of the intervention in the second year. Thus, to determine whether the impact of offering students the opportunity to enroll in the enhanced after-school program differed from the first to the second study year, the bottom panel of Table 4.5 presents the impacts of access to one year of the enhanced math program on student achievement for students in Cohort 2. The estimated impact of the enhanced math program on SAT 10 total math scores is not statistically significant

[69]The fall-to-spring growth in test scores for students in the sample is 36 scaled score points, based on the abbreviated SAT 10 test, whereas the fall-to-spring average growth for a nationally representative sample of students in grades 2 through 5 is 18 scaled score points, based on the full-length SAT 10 test. However, note that the study sample has a high proportion of low-performing students. (At the beginning of the program, 79.5 percent of the students in the Cohort 1 math program sample were performing "below proficient" in math.)

[70]This is calculated as 3.5 points (impact) divided by 36 points (regular program group growth).

The Evaluation of Academic Instruction in After-School Programs

Table 4.5

Impact of the Enhanced Math Program on Student Achievement in the Math Analysis Sample (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]					
SAT 10 math total scaled scores	607.01	603.52	3.49 *	0.09	0.02
Problem solving	607.88	605.38	2.50	0.06	0.11
Procedures	607.63	601.80	5.82 *	0.11	0.00
Sample size (total = 1,144)	634	510			
Cohort 2[b]					
SAT 10 math total scaled scores	606.72	603.35	3.37	0.09	0.07
Problem solving	608.80	606.24	2.57	0.06	0.18
Procedures	605.20	600.73	4.47	0.09	0.10
Sample size (total = 792)	461	331			

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

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: 389 to 796, 414 to 776, and 413 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.

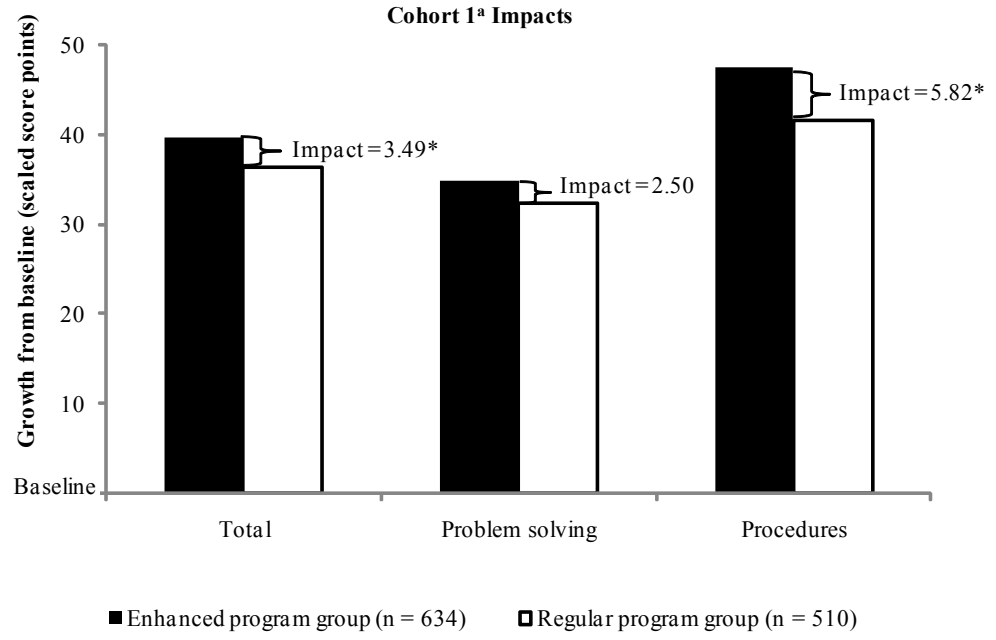
For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

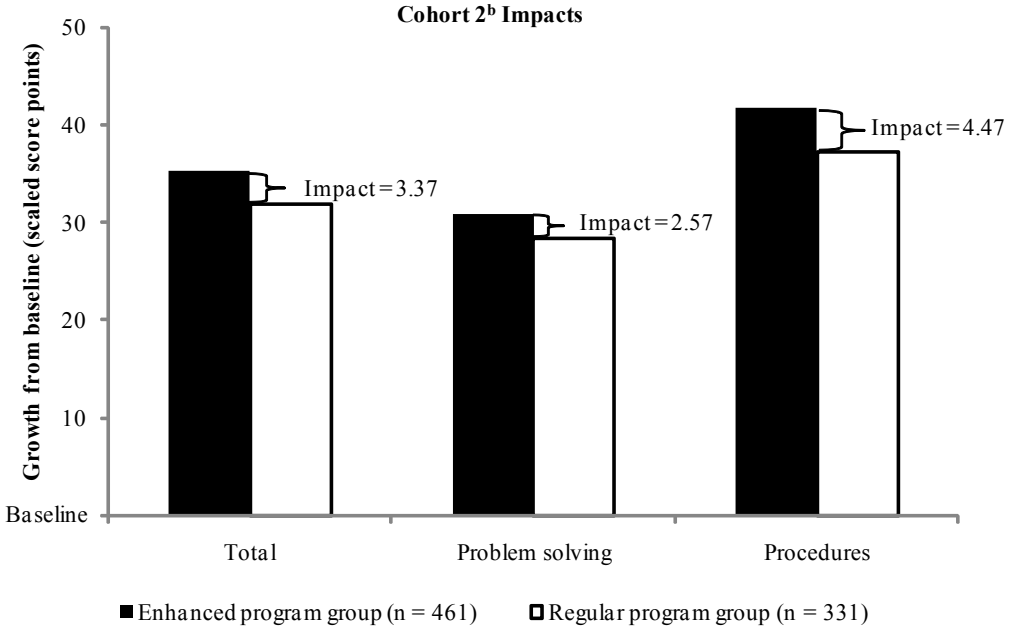
[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

Figure 4.2

SAT 10 Math Test Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Math Program (One Year of Service)





(continued)

Figure 4.2 (continued)

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

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. For Cohort 1, these effect sizes are 0.09, 0.06, and 0.11 for the math total, problem solving, and procedures scores, respectively. For Cohort 2, these effect sizes are 0.09, 0.06, and 0.09 for the math total, problem solving, and procedures scores, respectively.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

for students in the second year of implementation (p-value = 0.07). However, the difference in impacts between implementation years (Cohort 1 and Cohort 2 samples) is not statistically significant.[71] Thus, it cannot be concluded that the enhanced after-school math program was more effective in one implementation year than the other.

Another achievement measure of policy interest is the school district's locally administered standardized test, since it is typically tied to local accountability provisions. Thus, student scores on locally administered (state) tests were collected, and impacts on these test scores are examined. Because the locally administered tests were not available for second-grade students in some of the centers,[72] the impact analysis on locally administered tests is confined to students in grades 3 through 5. Since the scale of the locally administered test differs by site, all test scores were standardized within each study site by grade, and all estimated impacts on these tests are expressed in effect sizes. (See Appendix F for details on these outcome measures.)

[71]P-values for the difference in impacts between cohorts are 0.96, 0.98, 0.68 for the total, problem-solving, and procedures test scores, respectively.

[72]Tests for second-grade students were not available for nine of the 15 centers in the first year and seven of the 15 centers in the second year.

Appendix Table G.1 presents the results of this analysis for students in grades three to five. The impact of the enhanced math program on the locally administered math test is positive and statistically significant for Cohort 2 (0.18 standard deviation, p-value = 0.01).[73] And the difference in one-year impacts across cohorts is not statistically significant (p-value = 0.16), so it cannot be concluded that the impact of the enhanced program on locally administered tests differed from one implementation year to the other.

Impacts on Academic Behaviors

As explained in the theory of action outlined in Chapter 1, the impact of the enhanced after-school math program on student academic behaviors is uncertain in terms of its magnitude and direction. On the one hand, if students become better able to complete their school work, their classroom behavior may improve as a result of the enhanced math program. On the other hand, the additional formal instruction that students receive in the after-school program may cause "fatigue" and, therefore, negatively affect their behavior during the regular school day. Furthermore, the enhanced program replaces time spent on homework help, which could adversely affect students' homework completion.

To assess whether the enhanced after-school program changed students' behavior in any way, impacts on three measures of academic behaviors — homework completion, attentiveness, and disruptiveness in class — were examined. These measures are drawn from the survey of regular-school-day teachers. All three measures 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 4.6 shows that one year of access to the enhanced math program did not interfere with or improve homework completion, nor did it have a statistically significant effect on the two classroom behavior measures in either of the two years of program implementation (Cohort 1 or Cohort 2 samples). Nor is the difference in impacts across implementation years (cohorts) statistically significant. However, these findings should be interpreted with caution because all three variables were measured with a single survey item, thus compromising the reliability of the measures.

[73]State test results are available for the study sample students located in eight states. Two of these eight states use norm-referenced tests similar to the SAT 10. The other six states used criterion-referenced tests, which are typically linked to specific content in the curricula that is used during the regular school day. (See Appendix F for a detailed description of the state tests.)

The Evaluation of Academic Instruction in After-School Programs

Table 4.6

Impact of the Enhanced Math Program on Student Academic Behavior in the Math Analysis Sample (One Year of Service)

Student Academic Behavior Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]					
Student does not complete homework	2.25	2.24	0.00	0.00	0.96
Student is disruptive	2.10	2.14	-0.03	-0.03	0.54
Student is attentive	3.30	3.26	0.03	0.04	0.44
Sample size (total = 1,144)	634	510			
Cohort 2[b]					
Student does not complete homework	2.35	2.44	-0.10	-0.08	0.19
Student is disruptive	2.06	2.03	0.03	0.03	0.59
Student is attentive	3.35	3.33	0.03	0.03	0.57
Sample size (total = 792)	461	331			

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

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: homework = 1.15; disruptive = 1.09; attentive = 0.85.

The sample sizes reported represent the number of students from the analysis sample in each cohort. The sample size for each outcome varies by the number of regular-school-day teachers who responded to any given question.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

Chapter 5

Analysis of the Offer of Two School Years of Service in Math: Sample Characteristics, Service Contrast, and Impacts

This study also examines whether making the enhanced program available to students for two school years — thereby potentially lengthening students' average level of exposure to the program — improves student achievement above and beyond what they would have achieved had they remained in a regular after-school program. Hence, the present chapter focuses on the third key research question:

- What is the impact of offering students the opportunity to participate in the enhanced after-school math program for *two consecutive school years*?

As explained in Chapter 2, the impact of offering students the opportunity to participate in the enhanced program for two consecutive years is estimated by comparing the outcomes of students who were randomly assigned to either the enhanced after-school program (enhanced program group) or the regular after-school program (regular program group) for two consecutive school years. Not all students received the treatment to which they were randomly assigned. Thus, this analysis includes students assigned to two years of the enhanced program, whether or not they attended both years. In fact, 42 percent of the students assigned to the enhanced program in the fall of 2006 and then again in 2007 did not attend the afterschool program for a second year.[74] And 41 percent of the students assigned to the regular after-school program in the fall of 2006 and then again in 2007 did not attend the regular afterschool program for a second year. Hence, the impact findings presented later in this chapter are of a two-year offer of services (an intent- to-treat analysis), rather than the impact of two years receipt of the enhanced program. This latter relationship is addressed nonexperimentally in Chapter 6.

Before presenting the impact findings, however, the chapter describes the sample of students included in the analysis and provides a comparison of the academic services offered to students in each of the two program groups across both implementation years.

[74]The most common reason for students not reenrolling in the enhanced program was that they no longer had physical access to the program, either because they had moved away or did not have a means of transportation to/from the program. This second-year nonparticipation rate of 42 percent is lower than the student turnover seen in the prior national study of 21st Century Community Learning Center programs (James-Burdumy et al., 2005), in which 60 percent of treatment group students did not return for the second year of the program.

The Analysis Sample

The two-year sample used for the analysis includes 367 students; 227 (62 percent) were randomly assigned to the enhanced after-school program in both years of the study, and 140 (38 percent) were randomly assigned to remain in the regular after-school program in both years of the study. This sample is limited to students with follow-up data from both the evaluation-administered achievement test and the regular-school-day teacher survey.[75]

Table 5.1 presents the characteristics of students in the two-year sample, for each of the two program groups. As seen in this table, there is a statistically significant difference on the SAT 10 problem-solving achievement measure between students in the enhanced and regular after-school program groups. However, an overall F-test indicates that there is no systematic difference in the background characteristics of students in the enhanced and regular program groups. As seen in Table 5.1, the majority of students in the enhanced program group within the two-year sample are either Hispanic (30 percent) or black (36 percent); about half of the students in the sample (49 percent) are male; 16 percent are overage for grade; 74 percent were eligible for free or reduced-price lunch; and about one-third (33 percent) lived in a household with a single adult. Eighteen percent of students in the sample had a mother who did not finish high school. In addition, all students were enrolled in grades two through four in the first year of the study, given the two-year nature of the treatment.[76] At the beginning of the first implementation year, 64 percent of the students in the two-year sample were performing at a level defined by the publisher of the achievement test used in this study as below proficient in math.[77]

[75]Among those in the two-year sample who did not apply to the second year of the study and did not receive the second year of program services, follow-up data were collected for 67 students in the enhanced after-school program group (E_1E_2) and 38 students in the regular after-school program group (R_1R_2).

[76]A student enrolled in grade five in the first year of the study typically could not have been offered the opportunity to participate in the enhanced after-school program in the second year of the study because the enhanced-after school program is only available to students in grades two through five. Ten students enrolled in grade five in the first year of the study were retained in the second year of the study; however, these students were excluded from the analysis because, assuming that the enhanced program has an impact on grade promotion, retained students in the regular program group may no longer have a counterpart in the enhanced program group.

[77]As mentioned in Chapter 2, local staff used a variety of measures to recommend students for the program. However, because 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 the proficient level on the study-administered baseline test (SAT 10).

The Evaluation of Academic Instruction in After-School Programs

Table 5.1

Baseline Characteristics of Students in the Math Analysis Sample (Offer of Two Years of Service)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Enrollment						
2nd grade	121	79	42			
3rd grade	134	75	59			
4th grade	112	73	39			
Total	367	227	140			
Race/ethnicity (%)						
Hispanic		29.60	32.05	-2.45	-0.05	0.61
Black, non-Hispanic		35.81	39.05	-3.24	-0.06	0.47
White, non-Hispanic		26.69	22.51	4.18	0.09	0.30
Other		5.12	6.95	-1.83	-0.07	0.52
Gender (%)						
Male		49.13	46.03	3.10	0.06	0.60
Average age (years)		8.06	8.11	-0.04	-0.08	0.46
Overage for grade[a] (%)		16.15	16.40	-0.25	-0.01	0.95
Free/reduced-price lunch (%)						
Eligible (among information providers)		74.44	79.54	-5.10	-0.11	0.20
No information provided		4.52	1.56	2.96	0.18	0.13
Average household size		1.93	1.89	0.04	0.03	0.75
Single-adult household (%)		32.85	39.92	-7.07	-0.13	0.19
Mother's education level (%)						
Did not finish high school		18.25	15.37	2.87	0.07	0.53
High school diploma or GED certificate		26.88	35.41	-8.53	-0.17	0.10
Some postsecondary study		45.82	39.01	6.82	0.12	0.21
No information provided		9.05	10.21	-1.16	-0.05	0.65
SAT 10 baseline math total scaled scores		551.95	546.66	5.29	0.14	0.10
Problem solving		559.54	552.14	7.40 *	0.18	0.03
Procedures		542.30	538.31	4.00	0.08	0.34
Sample size (total = 367)		227	140			

(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 enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the

Table 5.1 (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 enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 effect size for each characteristic is calculated as a proportion of the standard deviation for students in the two-year sample regular program group.

An F-test was calculated 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-value is not significant.

[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.

Finally, recall from Chapter 2 that given the size of the two-year analysis sample (367 students), the study is equipped to detect a two-year impact of the enhanced program of 0.21 standard deviation or larger. This translates into an impact of 8.2 scaled score points on the SAT 10 total math test, which is equivalent to 20 percent of the expected growth in test scores for a nationally representative sample of students in grades two through four.[78,79]

[78]The growth from "fall one year" to "spring the next school year" in average SAT 10 total math scores for a nationally representative sample of students (based on normed averages for each grade from the test developers) with the same grade composition as the two-year sample is 41 scaled score points. Specifically, a weighted average of fall scores of nationally representative second-, third-, and fourth-graders is calculated where the weights are the proportion in the two-year sample that were in the second, third, and fourth grade at baseline. This weighted average is subtracted from the weighted average of spring scores of nationally representative third-, fourth-, and fifth-graders (the weights are the same as before) and derives the 41-point difference. Therefore, an 8.2 scaled score point impact is equivalent to 20 percent of the expected two-year improvement of nationally representative students in the same grade levels.

[79]Note that the minimum detectable effect size (MDES) for the test of the *difference* between the impact on students of their *first* year of access vs. the impact on students of being offered the opportunity to participate for two years is also 0.21 standard deviation, since both impacts are based on the same sample of students.

The Academic Service Contrast Between the Enhanced and Regular After-School Programs

This section describes the extent to which the academic support services received by students in the enhanced program group during both years of implementation differ from the "business as usual" services received by students in the regular program group. This cumulative two-year service contrast is what produces the impact of offering the enhanced after-school math program to students in both years of the study.

As seen in Chapter 4, the services received by the enhanced and regular program group differed as intended with respect to instructional offerings and the qualifications and experience of staff, in both years of implementation. For the purposes of understanding the impact of offering the student the opportunity to participate in two years of enhanced services, however, the other aspects of the service contrast discussed in Chapter 4 — i.e., student attendance in the after-school program, hours of after-school math instruction, and student academic support from other sources — are less useful because they reflect the service contrast over the course of only one year. Hence, the remainder of this section examines the cumulative difference between students assigned to the enhanced and regular program groups (across both years of program implementation), for these three aspects of the service contrast, drawing on data from surveys of after-school program staff, attendance records, and surveys of students and regular-school-day teachers.

Differences in Attendance and Hours of Academic Instruction in the After-School Program

Table 5.2 presents information on student attendance on the days that the enhanced program was operating, as well as the amount of after-school math instruction received by students in each program group. The top panel presents average attendance and instructional hours across both years of the study, while the bottom two panels present this information separately for each year of the study.

Cumulatively across both study years, students assigned to the enhanced program were offered the Mathletics program for 187 days and attended, on average, 122 days (for an average of 95 hours), whereas students in the regular program attended for 101 days (for an average of 24 hours) over the two-year span.[80] For days attended, the difference of 21 days (difference of 70 hours) is statistically significant at the 5 percent level and represents an estimated 22 percent

[80] Attendance for the regular program group is only counted for the days during which the enhanced math program was operating.

The Evaluation of Academic Instruction in After-School Programs

Table 5.2

Attendance of Students in the Math Analysis Sample (Offer of Two Years of Service)

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cumulative across both study years					
Attendance in after-school program[a]					
Number of days attended	121.67	100.63	21.04 *	0.56	0.00
Total hours of math instruction received[b]	94.51	24.10	70.42 *	3.03	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	37.20	33.21	3.99	0.09	0.44
Number of days per week[d]	1.01	0.65	0.36 *	0.22	0.01
Regular school day[e]					
Students receiving special support (%)	32.63	39.53	-6.90	-14.52	0.19
Minutes per week of individualized help	31.86	32.29	-0.43	-0.01	0.91
Sample size (total = 367)	227	140			
Study year					
First year (2005-2006 school year)					
Attendance in after-school program[a]					
Number of days attended	79.71	63.92	15.79 *	0.42	0.00
Total hours of math instruction received[b]	60.51	11.38	49.13 *	2.12	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	30.68	21.66	9.02 *	0.19	0.05
Number of days per week[d]	1.14	0.54	0.60 *	0.35	0.00
Regular school day[e]					
Students receiving special support (%)	21.03	23.04	-2.01	-4.24	0.64
Minutes per week of individualized help	39.71	37.86	1.85	0.04	0.76
Sample size (total = 367)	227	140			

(continued)

Table 5.2 (continued)

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Second year (2006-2007 school year)					
Attendance in after-school program[a]					
Number of days attended	41.96	36.70	5.25	0.14	0.23
Total hours of math instruction received[b]	34.00	12.72	21.28 *	0.92	0.00
Math support from other sources					
Out-of-school math class or tutoring[c]					
Students receiving instruction (%)	22.50	22.21	0.29	0.01	0.95
Number of days per week[d]	0.89	0.75	0.13	0.08	0.43
Regular school day[e]					
Students receiving special support (%)	21.91	25.28	-3.36	-7.08	0.47
Minutes per week of individualized help	24.00	26.72	-2.72	-0.05	0.46
Sample size (total = 367)	227	140			

(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 enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample 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 on the days they were present, or 60 minutes in centers that met only three days a week (one center in the first year and four centers in the second year). Total hours is calculated for these students by multiplying each student's total days of attendance by 45 (or 60).

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 or 60, 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.

[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?"

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

Table 5.2 (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?"

[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.

measured in days or hours occurred during the first year of the program (see Table 5.2, first- and second-year findings). This statistically significant decrease between implementation years in attendance (both in total days, p-value = 0.02, and hours of instruction, p-value = 0.00) is not surprising given that 42 percent of students assigned to the enhanced program for two consecutive years did not actually participate in a second year of enhanced services (and therefore attended zero days of the enhanced after-school program during that year and received zero hours of instruction).[82]

Differences in Academic Support from Other Sources

The second section of the first panel in Table 5.2 presents findings on the supplementary academic support services received by each program group over both years of the study, whether from non-school sources or during the regular school day. The regular program group received out-of-school math services 0.7 day per week on average, while the enhanced program group received such services 1 day per week on average, which is a statistically significant difference

[81]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, or 324 hours across two school years. Therefore, the 70 additional hours of after-school math instruction received by students in the enhanced program group represents a 22 percent increase in instructional time over the two-year period.

[82]The exploratory analysis in Chapter 6 will examine the association between *receiving* two years of enhanced services and the amount of instruction received, for students who actually participate in the enhanced program in the second year of the study.

and increases the service contrast. However, there is no statistically significant difference in the percentage of students in each program group who received out-of-school math support.[83]

Table 5.2 also shows that, across both years of the study, there are no statistically significant differences in the percentage of students in the enhanced and regular program groups receiving special support during the regular school day or in the amount of individualized help received.

Impacts on Student Achievement and Other Outcomes

This section examines whether being offered the opportunity to participate in the enhanced after-school math program for two consecutive years improves student achievement. Specifically, this intent-to-treat analysis indicates what the impact may be when a school offers a program to students for two consecutive years and when approximately 42 percent of the students do not return to the program after the first year. In addition to examining impacts on math achievement, the effect of the enhanced program is also estimated for three academic behaviors: homework completion, attentiveness, and disruptiveness in class.

Impacts on Student Achievement

In the spring of each study year, the Stanford Achievement Test, Tenth Edition (SAT 10), abbreviated battery in math was administered to students.[84] Total scores on the math test — as well as scores on two subtests, problem-solving and procedures — are used to measure individual students' academic achievement in math. The results presented in Table 5.3 show that the estimated impact of offering students the opportunity to participate in the enhanced after-school program for two consecutive years is not statistically significant (2.0 scaled score points on the SAT 10 or 0.05 standard deviation, p-value = 0.52). Additionally, there are no statistically significant differences on either of the subtests.[85]

[83]These findings are based on the follow-up student survey, administered in the spring of each school year. The survey asked students whether they attended a math class or activity outside the regular school day that was not part of the after-school program. (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.

[84]Spring 2006 for Year 1 and Spring 2007 for Year 2.

[85]The robustness of the impact findings presented in this section was tested by estimating program impacts based on the full sample instead of the analysis sample (i.e., students who have SAT 10 total test scores rather than students who have both SAT 10 scores *and* a regular-school-day teacher survey) and by using an alternative estimation model that includes only the random assignment block indicators as covariates. (In other words, the impact estimates are unadjusted except for the random assignment strata.) These sensitivity tests yield similar results to those reported in this chapter (see Appendix H).

The Evaluation of Academic Instruction in After-School Programs

Table 5.3

Impact of the Enhanced Math Program on Student Achievement in the Math Analysis Sample (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
SAT 10 math total scaled scores	618.27	616.30	1.97	0.05	0.52
Problem solving	620.09	617.15	2.94	0.07	0.34
Procedures	617.10	616.59	0.51	0.01	0.91
Sample size (total = 367)	227	140			

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

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: 428 to 796, 444 to 776, and 466 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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

-

hanced program was also estimated and compared to their cumulative two-year impact. Their first-year impact is not statistically significant (5.2 scaled score points on the SAT 10 or 0.13 standard deviation, p-value = 0.07). And the estimated impact of assigning students to two years of enhanced services is not statistically different from the impact on these students of their first year of access to the program (p-value = 0.28).

Figure 5.1 places these impact estimates in the context of the actual and expected two-year achievement growth of students in the enhanced program group. The figure plots the two-year growth in SAT 10 total math scores for students in the enhanced program group, as well as the expected growth that these students would have achieved had they not been assigned to the enhanced program for two consecutive years (as represented by the growth of students in the regular program group). As another frame of reference, the figure also plots the test score growth for a nationally representative sample of students with the same grade composition in each period as the two-year sample. As shown in this figure, the SAT 10 total scores of students in the *enhanced program group* grew by 66.3 points across both years of the study (44.5 points in the first year and another 21.8 points in the second year). However, the test scores of students in the *regular program group* also grew — by 64.3 points across both years of the study (39.4 points in the first year and another 24.9 points in the second year). The difference in growth rates between the two program groups produces the estimated impacts (not statistically significant) mentioned above, a difference of 5 points between the two groups after one year and a difference of 2 points after two years. This means that the test score growth of students in the enhanced program group cannot be attributed to the impact of the enhanced program because their scores would have grown by a similar amount had they not been assigned to the enhanced program for two consecutive years. Note that the average test score growth exhibited by students in *both* program groups may represent a closing of the achievement gap, but it could also be partially attributable to regression to the mean.[86]

Impacts on locally administered (state) tests were also examined, given the policy-relevance of these test scores.[87] Though not statistically significant, the estimated impact on locally administered standardized test scores of offering students the opportunity to participate in the enhanced program for two consecutive school years is 0.15 standard deviation (p-value = 0.09).[88] Appendix Table H.1 presents the results of this analysis.

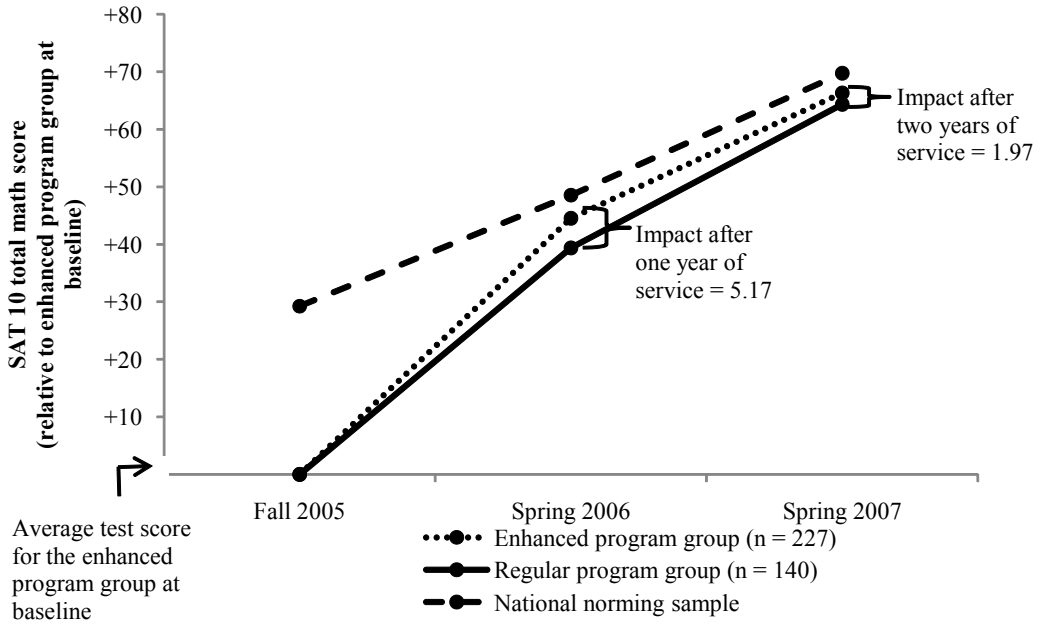
[86]Regression to the mean is a statistical artifact that makes random variation in longitudinal data look like true growth. Specifically, even in the absence of true growth, students with below-average SAT 10 scores at baseline (such as the students in this sample) would score closer to the national mean on the follow-up test than they did on the baseline test, due to measurement error in the SAT 10 assessment.

[87]Because the scale of the locally administered tests differs by site, all test scores were standardized within study site by grade, and all estimated impacts on these tests are expressed in effect sizes. (See Appendix F for details on these outcomes measures.). State test results are available for students in eight states. Two of these eight states use norm-referenced tests similar to the SAT 10. The other six states used criterion-referenced tests, which are typically linked to specific content in the curricula that is used during the regular school day. (See Appendix F for a detailed description of the state tests.)

[88]Because locally administered tests are not available for students in grade two, it is not possible to determine the impact on local tests for this particular sample of students.

Figure 5.1

SAT 10 Total Math Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Math Program After One Year and Two Years of Service



SOURCES: MDRC calculations are from baseline and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery. National norming sample calculations are from the SAT 10 (2002 norming sample): Stanford Achievement Test Series: Tenth Edition: Technical Data Report (Harcourt Assessment, 2004, pp. 312-338).

NOTES: The growth line for the enhanced program group is based on the observed mean baseline and follow-up test scores of students assigned to the enhanced after-school program for two consecutive years (baseline is Fall 2005; follow-ups are Spring 2006 and Spring 2007). The growth line for the regular program group represents the test scores that students in the enhanced program group would have obtained had they not been assigned to the enhanced program (calculated as the mean test score for the enhanced program group minus the estimated impact at a given time point). The growth line for the national norming sample is based on the average SAT 10 total math scores for a nationally representative sample of students with the same grade composition in each period as the two-year sample. Specifically, at each point in time (the fall baseline, the first spring, and the second spring), the SAT 10 national norm scores for second-, third-, and fourth-graders are averaged weighting each grade average score according to their proportion in the two-year study sample at baseline. This creates an expected two-year improvement of nationally representative students at the same grade levels as this study's sample. The baseline for the national norming sample is set relative to the average baseline score of the enhanced program group.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data was not collected. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

As noted earlier, however, the estimated impact of assigning students to the enhanced program for two consecutive years must be interpreted in light of the fact that 42 percent of students in the enhanced program group did not actually attend the program for a second year. This means that the results presented in Table 5.3 are a weighted average of the impact for students who attended both years of the enhanced program and the impact for students who attended the enhanced program in the first year only. Thus, the results discussed in this section represent the impact of *offering* the enhanced program to the same students in two consecutive years (an "intent-to-treat" analysis), rather than the impact of *receiving* two years of enhanced after-school services (an analysis of the impact of the "treatment on the treated"). Because the association between receiving two years of enhanced services and student outcomes cannot be estimated within the experimental framework of the study design, this question will be examined in the next chapter, which presents findings from some non-experimental exploratory analyses.

Impacts on Academic Behaviors

Offering students the opportunity to participate in the enhanced math program for two school years did not have a statistically significant impact on students' academic behaviors. Table 5.4 shows that being assigned to the enhanced after-school program in two consecutive years had no statistically significant impacts on homework completion or the two classroom behavior measures. However, as mentioned in the previous chapter, these findings should be interpreted with caution because all three variables were measured with a single survey item, thus compromising the reliability of the measures.

The Evaluation of Academic Instruction in After-School Programs

Table 5.4

Impact of the Enhanced Math Program on Student Academic Behavior in the Math Analysis Sample (Offer of Two Years of Service)

Student Academic Behavior Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Student does not complete homework	2.23	2.43	-0.20	-0.18	0.08
Student is disruptive	2.16	1.99	0.16	0.15	0.14
Student is attentive	3.31	3.38	-0.07	-0.08	0.44
Sample size (total = 367)	227	140			

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

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: homework = 1.15; disruptive = 1.09; attentive = 0.85.

The sample sizes reported represent the number of students from the analysis sample. The sample size for each outcome varies by the number of regular-school-day teachers who responded to any given question.

Chapter 6

Exploratory Analyses of the Impact of the Enhanced After-School Math Program

This chapter reports on two exploratory analyses whose purpose is to provide information that may inform the design and implementation of the enhanced math program. However, because these analyses are nonexperimental, they should be viewed as hypothesis-generating since they may not reflect true causal relationships.

As discussed in Chapter 5, not all students assigned to the enhanced program both years participated in the second year. In order to provide information about the treatment for those who actually received it in both years, and to examine whether longer exposure to the program is associated with improved student outcomes, the first exploratory analysis examines the relationship between achievement and program participation for those students who participated in both years of the enhanced after-school services.

Additionally, the enhanced program was offered in a variety of different settings. Understanding how variation in the local school context, as well as variation in program implementation (across centers and the two implementation years), is associated with impacts on achievement can help one interpret the generalizability of the overall findings, as well as generate possible avenues for program improvement. Thus, the second exploratory analysis examines whether the impact of one year of enhanced services is associated with the characteristics of program implementation in the after-school center and/or with the characteristics of the local school context in which the program was implemented.

The Association Between Receiving Two Years of Enhanced After-School Math Instruction and Student Achievement

This section examines the association between receiving enhanced after-school services for two consecutive years and math achievement by focusing on the students in the enhanced program group who were randomly assigned to — and participated in — the enhanced after-school math program in both years of the study.

Estimating the two-year impact for these students is challenging, however, because students who received two years of enhanced after-school services chose to attend a second year, perhaps based on factors related to their experience in the enhanced program during the first year of the study. Because these students' decision processes are not known, it is not possible to identify students in the regular program group who would have made the same choice had they

been given the option to participate. In other words, it is not clear which students in the regular program group provide the appropriate counterfactual for returning students in the enhanced program group who received two years of enhanced services.

Thus, the association between receiving two years of enhanced services and math achievement is estimated from nonexperimental methods, using an instrumental variables analysis. This technique identifies who among the regular program group are most like those in the two-year enhanced program group and essentially compares outcomes of like individuals.[89]

Table 6.1 shows that the association between students receiving two years of the enhanced after-school program and achievement is not statistically significant (3.7 scaled score points for SAT 10 total math scores, p-value = 0.36). Additionally, the nonexperimental estimate of receiving two years of enhanced after-school services does not statistically differ from the estimated impact of receiving one year of enhanced services (p-value = 0.40).

Taken together, the experimental findings for Cohort 1 from the previous chapter and the above nonexperimental findings suggest that for this population of struggling students, a second year of the enhanced after-school services — whether offered or received — does not improve math achievement, over and above the achievement gains already made in the first year.[90]

[89]Specifically, estimated comparisons are based on students who were randomly assigned to one of three conditions: two years of enhanced services, two years of regular services, or enhanced services in the first year of the study but not the second. Based on this sample of students, impact estimates were obtained from an instrumental variable analysis in which the two treatment conditions (that is, two years of enhanced services and enhanced services in the first year but not the second) are used as instrumental variables for the number of years of enhanced services that were actually received (one year or two years). This model was fitted using two-stage least squares. Estimated associations 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. Appendix I further describes the conceptual underpinnings of the analysis and the statistical model in greater detail, as well as the sample of students included in the analysis.

[90]In order to interpret the two-year associations in Table 6.1, it is important to understand the extent to which the services received by students in the enhanced program group who applied in the second year differ from the services received by their counterparts in the regular program group who also applied in the second year. For this reason, the association between receiving two years of enhanced services and the *hours of math instruction received* by students was estimated (see Appendix I for details). As seen in the service contrast section in the previous chapter, *offering* students the opportunity to participate in enhanced services for two years increases the amount of math instruction that they receive by 70 hours across both years of the study. Based on an instrumental variables analysis (see Appendix I), *receiving* two years of enhanced services increases the amount of instruction by 86 hours (p-value = 0.00).

The Evaluation of Academic Instruction in After-School Programs

Table 6.1

Association Between Receiving Two Years of the Enhanced Math Program and Student Achievement

Student Achievement Outcome	Students Who Received Two Years of Services	Estimated Counterfactual	Estimated Comparison	Standardized Estimated Comparison[a]	P-Value for the Estimated Comparison
SAT 10 math total scaled scores	617.37	613.69	3.68	0.09	0.36
Problem solving	621.10	618.40	2.70	0.07	0.53
Procedures	612.96	607.43	5.54	0.11	0.33
Sample size (total = 534)[b]	NA	NA			

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: 428 to 796, 444 to 776, and 466 to 768.

Estimated comparisons are based on students who were randomly assigned to one of three conditions: two years of enhanced services, two years of regular services, or enhanced services in the first year of the study but not the second. Based on this sample of students, impact estimates were obtained from an instrumental variable analysis in which the two treatment conditions (that is, two years of enhanced services; enhanced services in the first year but not the second) are used as instrumental variables for the number of years of enhanced services that were actually received (one year or two years). This model was fitted using two-stage least squares. Estimated associations 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 "Students Who Received Two Years of Services") are the observed means for students who were assigned to and received two years of enhanced services. The values in column 2 (labeled "Estimated Counterfactuals") are the estimated outcomes that these students would have obtained had they not received two consecutive years of enhanced services. Rounding may cause slight discrepancies in calculating sums and differences.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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.

[a]The standardized estimated comparision for each outcome is calculated as a proportion of the standard deviation for students in the regular program group. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

[b]Group-specific sample sizes are not presented because the analysis is not based on a direct comparison of students who received two years of enhanced services to students who did not receive two years of enhanced services.

Linking the Impact of One Year of Enhanced Services on Math Achievement with School and Program Characteristics

As discussed in Chapter 4, the estimated impact effect size of one year of access to the enhanced program on total math scores is a 0.09 standard deviation (or 3.5 scaled score points in Cohort 1 and 3.4 in Cohort 2). However, each year, not all centers in the study sample experienced this exact gain.[91] Understanding how variation in the local school context, as well as variation in program implementation, is linked to impacts on achievement may suggest settings or implementation features that may be associated with different impacts. Thus, this section explores whether the impact of one year of enhanced services on SAT 10 total math scores in an after-school center (in either implementation year) is associated with (1) the characteristics of the school that housed the after-school center and (2) the characteristics of a center's implementation of the enhanced program. Using both study years allows these characteristics to vary both within centers over time and across centers within a given implementation year.[92]

The analysis was conducted by using a linear interaction model to estimate the association between these center characteristics and program impacts on SAT 10 total scores in the participating after-school centers in both study years (i.e., the 30 center-level impacts).[93] Because students were not randomly assigned to programs with different school characteristics, this analysis is exploratory rather than experimental; as such, these results should be viewed as hypothesis-generating rather than as establishing causal inferences.

Three measures of *program implementation* are included in the analysis: the number of days over the course of the school year that the enhanced math program was offered (included as a measure of program dosage), whether one or more teachers teaching the enhanced program left during the school year (included as a measure of disruption in instruction), and the difference between the total hours of after-school academic instruction received by students in the enhanced program group relative to students in the regular program group (a measure of service contrast). The analysis also includes five measures of the *local school context* that capture the characteristics of the regular school day, as well as the characteristics of the school's student body. These measures are: whether the school met its Adequate Yearly Progress (AYP) goals,

[91]Center-by-year impact estimates on SAT 10 total math scores range from -10.1 scaled score points to 18.8 scaled score points. An F-test indicates that the overall variation in impacts across centers and implementation years is not statistically significant at the 5 percent level (p-value = 0.07). Nonetheless, statistically significant associations between school-level predictors and impacts may still be found, thus providing information that can be used to improve the design and implementation of the program. See Appendix J for a more detailed discussion of variation in impacts across centers and implementation years.

[92]Variation in each of the program implementation and local school context measures across centers and years is statistically significant (p-value for variation of each measure is 0.00).

[93]Fifteen centers * two implementation years = 30 center-level impacts.

whether the in-school student-to-teacher ratio is greater than in the enhanced after-school program (13:1),[94] the amount of math instruction that students received during the regular school day,[95] an indicator for the instructional approach of the math curriculum used during the school day,[96] and the proportion of students receiving free or reduced-price lunch. Details on these measures are provided in Appendix J.

Table 6.2 presents the estimated association between program impacts on SAT 10 total math scores and these school-level characteristics. Program impacts were larger in after-school centers that offered the enhanced program for a greater number of days during the school year (p-value = 0.00), where one or more teachers of the enhanced program left during the school year (p-value = 0.04), and in schools that made adequate yearly progress (p-value = 0.00). Given the unexpected direction of some of these findings, it is not possible to explain the reasons for these relationships.[97]

---

[94]As noted in Chapter 2, the planned student-teacher ratio was 10:1; however, up to 13 students were randomly assigned to each class, in order to account for the possibility that some students might not attend on a given day.

[95]School administrators were asked how many minutes teachers spend per day teaching math 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. Specifically, across both cohorts, 30 percent of schools offer 50 to 60 minutes; 43 percent offer 60 minutes; 13 percent offer 60 to 90 minutes; and the remaining 13 percent offer 90 minutes or more (rounding may cause slight discrepancies in calculating sums and differences). 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.

[96]Based on their instructional approaches, school-day curricula were categorized into two groups. The first group 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). The reference 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.

[97]Three additional school-level measures were available for the second year of program implementation in math centers. The first is the average yearly achievement gain of students in the school, which serves as a proxy for the level and quality of instruction and leadership at the school.

The second measure is the percentage of enhanced program teachers in the second year of the study who also taught during the first year (i.e., "returning" teachers). This measure is intended to gauge program implementation strength, since one would expect returning teachers to be better able to deliver the enhanced curriculum than new teachers.

The third additional measure is an indicator of whether, on average, students in the enhanced program spent fewer than four days on each math skill pack assigned by the teacher (where four days is the center-level average in the sample). This indicator serves as a measure of teachers' instructional pacing.

Given the availability of these additional measures, a separate analysis was conducted focusing on the second year of the study only (i.e., 15 center-level impacts) and using all available school-level characteristics in the second year of the study. None of the individual school context or implementation characteristics were associated with program impacts by a statistically significant amount.

The Evaluation of Academic Instruction in After-School Programs

Table 6.2

Associations Between School and Program Characteristics and the Impact of the Enhanced Math Program on Student Achievement After One Year of Service

Interaction Characteristic	Estimated Coefficient	P-Value for the Estimated Coefficient
School context		
Curriculum group 1[a]	-6.83	0.12
More than 60 minutes of math instruction	2.43	0.52
Student-to-teacher ratio greater than that in the enhanced program[b]	1.24	0.68
Did not meet Adequate Yearly Progress (AYP) goals	-11.31 *	0.00
Percentage of student body that is low-income	-0.02	0.68
Program implementation		
Total days enhanced program was offered	0.44 *	0.00
Service contrast between enhanced and regular program groups[c]	-0.07	0.29
Enhanced teacher left the program during the school year	6.57 *	0.04
F-test of all characteristics	*	0.01
F-test of school context characteristics	*	0.01
F-test of program implementation characteristics	*	0.01
Size of student sample (total = 1,936)		
Size of school sample (total = 15 schools times 2 years = 30)		

(continued)

SOURCES: MDRC calculations 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 (CCD) Web site, http://nces.ed.gov/ccd/. Program implementation characteristics are from the Evaluation of Academic Instruction in After-School Programs attendance data and from Bloom Associates. These data reflect the 2005-2006 and 2006-2007 school years.

NOTES: The estimated coefficients represent how the impact of the math program on SAT 10 math total scaled scores varies with each school characteristic. These estimates were obtained by fitting an impact model that includes an indicator of treatment status, as well as a set of interaction terms between the treatment indicator and each of the school characteristics listed above; the findings reported in the table are the coefficients of the interaction between treatment status and the school characteristics. The model also controls for random assignment strata, students' baseline math total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The F-test tested whether the coefficients of the school characteristic variables are jointly equal to zero. Within each center, the analysis sample includes, on average, 65 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 two 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). The left-out group contains curricula that have a format with math topic sections within chapters. Each section contains guided practice problems, numerous computational

Table 6.2 (continued)

[a]Based on their instructional approaches, school-day curricula were categorized into two 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). 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.

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

[c]Service contrast is measured as the difference between the total hours of after-school academic instruction received by students in the enhanced program group relative to students in the regular program group. This difference is obtained from a regression model that estimates the impact of the enhanced program on the number of hours of after-school academic instruction received by students, controlling for random assignment strata and student characteristics. This regression model is estimated for each center in each year of the study.

Chapter 7

Implementation of the Enhanced After-School Reading Program

This chapter begins by describing the 12 after-school centers that implemented the enhanced reading instruction for both years of the evaluation. It then presents the intended design of the enhanced reading instruction and the implementation findings for both the structural and instructional elements of the program.

Centers in the Reading Study Sample

Table 7.1 presents the characteristics of schools in school year 2005-2006 that house the 12 after-school centers that implemented the enhanced reading program over two school years. As shown in this table, six schools are located in large or midsize cities, and six are located in the urban fringe of a city or in a town. Five of the 12 schools (42 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.[98] Students in the schools are black (37 percent), white (23 percent), Hispanic (35 percent), Asian (4 percent) or American Indian (1 percent), and 71 percent of all students in these schools come from low-income families.[99] The average student-to-teacher ratio in these schools is 15:1. During the regular school day, students in five of the 12 schools receive more than 90 minutes of reading instruction each day, with students in seven schools receiving 90 minutes or less (see Table 7.2).[100] As shown in Table 7.2, the school-day reading instructional approach varies, and schools may use different reading curricula across grades two through five.

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

[99]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).

[100]School administrators were asked how many minutes teachers spend per day teaching 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 of offering 90 minutes.

The Evaluation of Academic Instruction in After-School Programs

Table 7.1

Characteristics of Schools Housing After-School Centers Implementing the Enhanced Reading Program

Characteristic	
Number of schools	
School setting[a]	
Large or midsize city	6
Urban fringe of a large or midsize city or large or small town	6
Schools not making Adequate Yearly Progress (AYP) goals	5
Composition of student body	
Race/ethnicity of students (%)	
Black	36.88
White	23.17
Hispanic	35.34
Asian	3.77
American Indian	0.80
Low-income students[b] (%)	70.81
Average student-to-teacher ratio	15:1
Sample size (total = 12)	

SOURCES: All school-level characteristics were collected from the Common Core of Data (CCD) Web site, except for AYP status, which was collected from each state's Department of Education Web site. CCD data reflect the 2005-2006 school year (the first year of implementation), which is the most recent year for which data are available. AYP status data reflect the 2006-2007 school year.

NOTES: The composition of the student body is calculated by averaging the proportion of students within each school 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.

Table 7.2

Characteristics of the Regular School Day in Schools Housing After-School Centers Implementing the Enhanced Reading Program

Regular-School-Day Characteristic	Number of Schools
Minutes of reading instruction offered	
Number of schools with 90 minutes or less	7
Number of schools with more than 90 minutes	5
Reading materials/curricula[a]	
Balanced Literacy	
Guided Reading Model	
Houghton Mifflin Reading: A Legacy of Literacy	
Open Court Reading (SRA/McGraw-Hill)	
Scholastic	
Scott Foresman	
Sample size (total = 12)	

SOURCES: Data were collected from research staff interviews with point persons and phone calls made to schools and districts in spring 2007 in regard to the 2005-2006 school year (the first year of implemenation).

NOTES: Data reflect grades 2 through 5 only. School and district staff were asked for the names and publishers of the reading 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.

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

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.[101]

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 and brief video segments, which are embedded into the daily lessons and 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

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

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.

Implementation Findings

This section presents the implementation findings for both the structural and instructional elements of the program, as well as the implementation challenges encountered. As described in Chapter 2, it draws on surveys of after-school program staff involved in its operation, conducted by the research staff; structured protocol observations of implementation of Adventure Island, conducted by district coordinators; interviews with district coordinators and teachers of the enhanced after-school program, conducted by the research staff; and attendance records.

Implementation findings are presented by implementation year in Table 7.3. Additionally, as after-school teachers and centers became more experienced with the delivery of the intervention, program implementation may have improved. Thus, this section also examines whether implementation differed between the two years of the study. In instances where implementation did not differ between the two years, findings for each year are presented in Table 7.3 and only *first* implementation year findings are discussed in the text.

Structural Elements[102]

The implementation of Adventure Island was supported using a set of strategies related to staffing, instructional hours, and support for instructors. These strategies were utilized in both years of the study, but some were provided with less intensity in the second year. Following is a description of these strategies, and reports on how they were implemented.[103]

[102]Findings in this section are largely drawn from the After-School Staff Survey, which was completed at the midpoint of both school years by all staff providing academic support to students in the participating after-school centers to gain information about instructors' impressions of and interactions with the intervention. The staff surveys were given to all teachers in the second year, regardless of whether it was their first or second time teaching in the enhanced after-school program. In the first year, 93 percent of staff (52 of 56) responded to the survey; in the second year, 83 percent of staff (50 of 60) responded to the survey. Among the staff responding to the survey, not all staff answered every question. Throughout this section, percentages are out of the 52 in the first year or 50 of staff in the second year who responded to the survey, unless indicated otherwise.

[103]Sites trained substitute teachers to teach Adventure Island, but these individuals are not included in the findings of this section unless they replaced a regular teacher prior to the time that the after-school staff survey was fielded.

The Evaluation of Academic Instruction in After-School Programs

Table 7.3

Characteristics of and Support for Enhanced Reading Program Staff

Service Offering	Year 1	Year 2	Difference	P-Value for the Estimated Difference
Structural Elements				
Staffing				
Certified in elementary education (%)	98.08	100.00	-1.92	0.32
Years of elementary school teaching experience (%)				
No experience to 2 years	11.54	8.00	3.54	
3-4 years	11.54	20.00	-8.46	
More than 4 years	76.92	72.00	4.92	
			chi-square	0.65
Staff-youth ratio (youth enrolled)	9.44	9.31	0.13	0.76
Staff-youth ratio (actually attended)	8.69	8.60	0.09	0.84
The Amount of Instruction Offered				
Hours of instruction offered	76.15	79.17	-3.02	0.38
Support for Staff				
High-quality training to carry out activity (%)				
Very true	73.08	79.17	-6.09	
Sort of true, not very true, or not at all true	26.92	20.83	6.09	
			chi-square	0.45
Had enough materials and equipment to carry out work (%)				
Very true	88.46	90.00	-1.54	
Sort of true, not very true, or not at all true	11.54	10.00	1.54	
			chi-square	0.80
Amount of paid preparation time to carry out activity (%)				
No minutes to less than 30 minutes per day	25.49	34.69	-9.20	
30 or more minutes per day	74.51	65.31	9.20	
			chi-square	0.31
Ongoing support from district for how to teach children in activity (%)				
Very true	82.69	89.58	-6.89	
Sort of true, not very true, or not at all true	17.31	10.41	6.90	
			chi-square	0.58

(continued)

Table 7.3 (continued)

Service Offering	Year 1	Year 2	Difference	P-Value for the Estimated Difference
Instructional Elements				
Teachers' Assessment of the Content of the Program				
Materials were appropriate for students (%)	100.00	100.00	0.00	NA[a]
Material difficulty (%)				
At about the right level of difficulty	92.31	100.00	-7.69	
Too easy or too challenging	7.69	0.00	7.69	
			chi-square	0.73
Sample size (total = 102)	52	50		

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

NOTES: Percentages are based on the number of staff who responded to the question.
[a]A statistical test cannot be conducted for this difference because there is no variability in teachers' responses to the survey question.

Staffing

There are two key staffing strategies: (1) hiring certified teachers as instructors, with a preference for experienced teachers who also are able to make a full-year commitment to the program and (2) establishing 10:1 student-to-teacher ratios for instruction. Additionally, when the study was extended to include a second year of program operations, every effort was made to recruit back staff from the first program year.

Based on responses to the survey of after-school staff, centers in both years did not statistically differ in the proportion of certified staff and staff with varying degrees of experience, nor did they differ in the number of students per staff member. Specifically, in the first year, 98 percent of Adventure Island instructors were certified teachers, and 77 percent of teachers had more than four years of elementary school teaching experience.

In both implementation years, random assignment was conducted in a manner to produce enhanced program groups of 10 to 13 students per grade, which allowed for some attrition and absences and still maintained an average class size of 10 students. When asked midyear on the After-School Staff Survey, Adventure Island instructors in both years reported an average of nine students were enrolled in their classes per staff member. When asked, "How many students

actually attend this activity on a typical day?" instructors in both years reported that an average of nine students per staff member were present.

While there was teacher turnover within each of the implementation years, comparatively more teacher turnover occurred across implementation years. Specifically, of the 56 teachers hired at the beginning of the first school year, there were four instances of teachers (in two different centers) leaving before the end of the school year. In the second year, of the 60 teachers hired, 10 staff from seven centers left before the end of the school year.[104] Thus, at least 83.4 percent of the teachers remained teaching in the program within a given program year (7 percent left in the first year and 16.6 percent left in the second year).[105] However, at the beginning of the second school year, of the 60 teachers hired, 21 staff were returning to the program for a second year, while the other 39 second-year staff were new to the program Thus, about 38 percent (21 out of 56) returned to teach in the program for a second year.

The Amount of Instruction Offered

The intended amount of instruction is 180 minutes per week, either in four 45-minute lessons or in three 60-minute lessons. On average, the program was implemented each year with, at a minimum, this intended amount of instruction. In the first year of implementation, the after-school program staff teaching Adventure Island reported on the staff survey that they offered an average of 177 minutes of instruction per week, which is not statistically significantly different from the amount intended (p-value= 0.61). In the second year, the program staff teaching Adventure Island reported that they offered an average of 175 minutes of instruction per week, which is also not statistically significantly different from the intended amount of instruction (p-value= 0.46).

Across the entire school year, the total hours of enhanced after-school instruction offered does not statistically differ between the two implementation years (p-value= 0.38). Specifically, in the first year the program was offered on average for 76 hours, whereas in the second year it was offered on average for 79 hours.

Support for Staff

Enhanced program instructors received training and support in a variety of ways throughout both school years. In both years, all the instructors (56 in the first year and 60 in the second year) were hired in time to attend the summer training on Adventure Island prior to the

[104]Eight of the teachers who left had not taught in the first year. Of those who left, reasons for leaving included: becoming an assistant principal, becoming pregnant, and not working well with the reading curriculum.

[105]The difference between the number of teachers who left within the first year and the number of teachers who left within the second is statistically significant (p-value = 0.05).

start of the school year, and the training was repeated in the following January for new staff. In the first year, seven new reading instructors were trained during the midyear conference (four replacements for teachers who left and three new substitute teachers). In the second year, five new reading instructors were trained (three replacements for teachers who left throughout the year and two new substitutes).[106]

When surveyed, instructors were asked if they received high-quality training to carry out their activities. Teachers' responses in both years did not statistically differ. In the first year, 73 percent of Adventure Island instructors reported that it was "very true" that they received high-quality training to carry out their activities.

In the first year, a component of the implementation strategy was to provide staff with all materials needed to teach Adventure Island so they would not be burdened by purchasing supplies. In the second year, this strategy was modified and sites were asked to pay the cost of replacing all consumable materials. Despite this modification, when asked if the instructors had enough materials and equipment to carry out their work, the instructors' responses did not statistically differ across the two implementation years. In the first year, 88 percent of the instructors reported that it was "very true" that they had enough materials and equipment to carry out their work. The implementation plan also called for 30 minutes of paid daily preparation time, and, again, reports on how much time was received did not statistically differ across the two years. Specifically, 75 percent of instructors in the first year reported that they had 30 minutes or more of paid preparation each day.

However, interviews with teachers conducted just in the first study year suggest that the 30 minutes of prep time was not always sufficient. As part of the structured interviews (following the classroom observation of half the instructors), the teachers were asked open-ended questions to identify what challenges they encountered implementing Adventure Island and how the program might be improved. When asked specifically about their preparation time, 21 percent of teachers (five of the 24) volunteered that they did not feel the preparation time allotted was sufficient. In the second year, researchers conducted structured interviews with district coordinators. A specific question asked of district coordinators was "Is the amount of preparation time sufficient?" Out of the 10 Adventure Island district coordinators, seven said that "it is sufficient" or "yes." The other three said that preparation time was insufficient in the beginning of the year. These three felt that as the year progressed the preparation time became adequate once teachers got used to the program or that the preparation time felt sufficient to them but that teachers do not agree.

[106]Although 10 teachers left throughout the second year, only three had replacements trained at the midyear conference. The other seven either were replaced by substitutes or did not leave during the fall.

The project also provided ongoing, on-site technical assistance. As outlined in Chapter 1, in the first year this consisted of Success for All representatives visiting each reading site twice during the school year; a project-funded, part-time district coordinator to support implementation; and frequent technical assistance from Bloom Associates (two on-site visits during the first intervention year and weekly conversations by phone). In the second year, on-site technical assistance was provided less intensively. A district coordinator continued to support implementation. However, Bloom Associates provided assistance through two site visits and bi-weekly phone calls, and Success for All representatives visited the sites only once. Despite this lessening in support, when asked whether they received ongoing support on how to teach children in their Adventure Island activity, responses from teachers across the two years did not statistically differ. In the first year, 83 percent of the instructors said that it was "very true" that they received ongoing support on how to teach children in Adventure Island.

Instructional Elements

The project team collected data on the teachers' assessment of the content of the program and on four different aspects of teachers' implementation of the Success for All program: use of instructional elements; the use of assessments to guide instruction; student placement and progression through the skill levels; and the pacing of the instructional content of the program.

Teachers' Assessment of the Content of the Program

In both years staff were asked whether the Adventure Island materials were appropriate for their students. Across the two implementation years, staff responses did not statistically differ. In both years, all staff reported it was "true" that materials were appropriate for their students. In the first year, 92 percent of the instructors reported that the materials and exercises were at "about the right level of difficulty," while the remaining 8 percent felt that the materials were "too easy" or "too challenging."

Use of Instructional Elements

In both years of the study, under the guidance of Bloom Associates staff, 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 core elements of the material that were identified by the developer as being key to intended implementation.

Observers of the two lower-level Adventure Island classes (Alphie's Lagoon and Captain's Cove) used a protocol with six components, including three procedural factors (use of SFA materials, cooperative learning, and awarding of points to student teams for performance) and three key topics to be covered (phonics, fluency, and completion of lesson plan). Across the

two implementation years, staff observation scores did not statistically differ (p-value = 0.90). In the first implementation year, 77 percent of the two lower-level Adventure Island classes (17 classes) included between three and five of the six components, and 23 percent (five classes) included between five and six.

Among staff teaching the two lower levels of Adventure Island, returning staff in the second year were more likely to receive a higher implementation observation score than new staff (p-value = 0.00). When observed, out of a total possible score of six components, half of the returning eight teachers received a score between four and five, and half received a score between five and six. Among 14 new teachers, three teachers received a score between three and four and 11 teachers received a score between four and six.

Since phonics was emphasized in the lower levels of Adventure Island but not in the upper levels, observers of the two higher levels (Discover Bay and Treasure Harbor) received a different protocol, which included five components. Staff in the first year were more likely to receive a higher observation score (p-value = 0.00). In the first year, 80 percent of the 20 classes included between three and four of the five components, and 20 percent included between four and five. In the second year, all 21 classes included between two and four of the five components.

The lower scores of staff in the second year were driven by the lower scores of new teachers.[107] When observed, out of a total possible score of 5 components, all of the new teachers received a score between two and four components, while all of the returning teachers received a score between three and four components.

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

[107]The difference between new and returning teachers' observation scores was statistically significant (p-value = 0.00).

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).

Student Placement and Progression Through the Skill 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). To illustrate this, Figure 7.1 shows for the first implementation year (Cohort 1) how students in each grade were initially placed in the Adventure Island levels in the fall, based 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, 84 percent of second-graders (or 107 students) were placed as "beginning readers" in Alphie's Lagoon; 93 percent of third-graders (or 113 students) were placed below the third-grade-level Discovery Bay; and all fourth- (129 students) and fifth-graders (126 students) 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.[108] Starting with the second semester, 63 percent of the second-graders (or 81 students) were placed in Captain's Cove; 32 percent of third-graders (or 38 students) were placed in Discovery Bay or Treasure Harbor; and 24 percent of fourth-graders (31 students) and 52 percent of fifth-graders (64 students) were placed in Treasure Harbor.

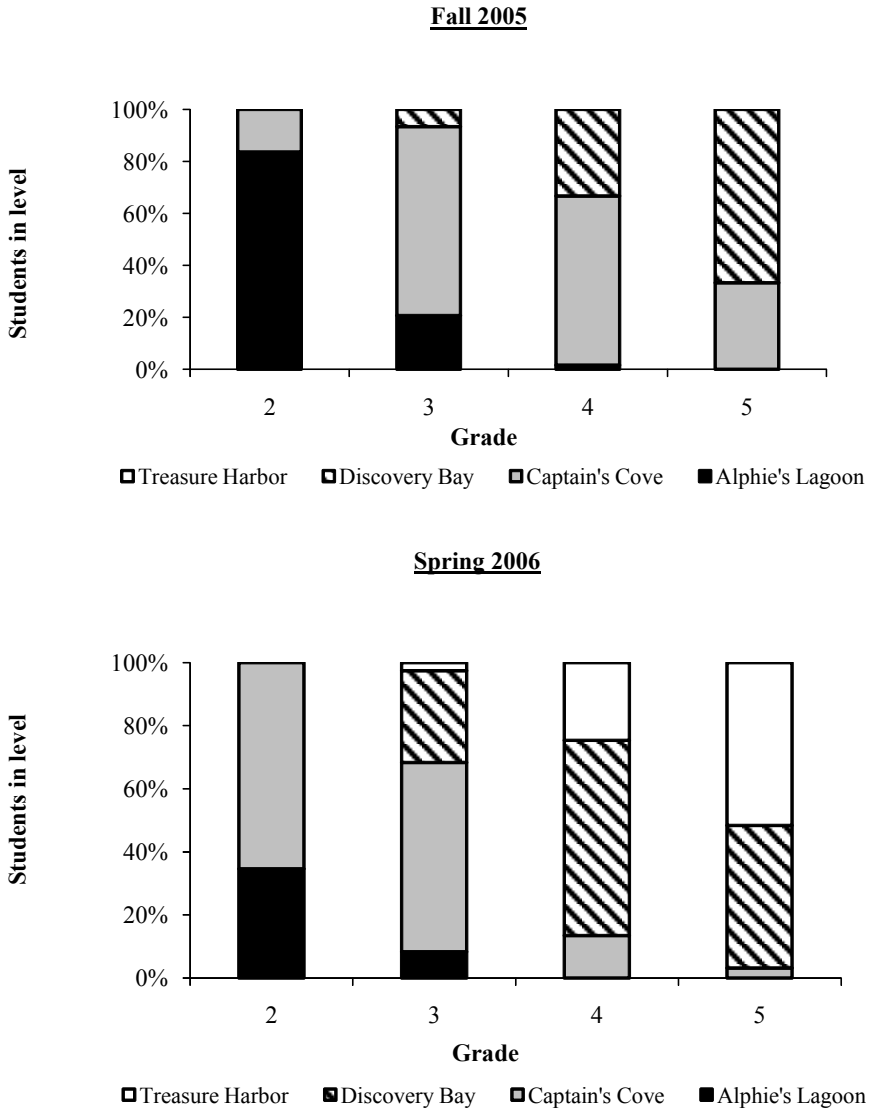
Pacing of Instruction

The Adventure Island daily lesson plans contain multiple instructional methods (such as direct instruction and cooperative learning) and specific topics, like phonics. In the first year, 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?" Nineteen of the 24 teachers interviewed indicated experiencing some challenges

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

Figure 7.1

The Percentage of Students in Each Adventure Island Level for Cohort 1, by Grade



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

NOTE: The Fall 2005 sample consists of the 504 students who received enhanced reading instruction in the fall, and the Spring 2006 sample consists of the 500 students who received enhanced reading instruction in the spring.

related to pacing. The 24 teachers' responses were categorized as follows: 46 percent (11 of the 24) described pacing as a "consistent problem" and said that, as a rule, they had trouble completing the daily lesson in the allotted time. Another 33 percent (8 of the 24) 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, 21 percent (5 of the 24) reported that they were generally able to cover the material in the allotted time and that pacing was "rarely a problem" for them.

In the second year of the study, research staff conducted interviews with district coordinators about implementation challenges. District coordinators were specifically asked whether "pacing" continued to be a problem for staff in the second year. Of all 10 district coordinators responding to the question, four said that pacing was a problem in the second year, four said it was not, and two did not answer the question.

Chapter 8

Analysis of the Offer of One Year of Service in Reading: Sample Characteristics, Service Contrast, and Impacts

The primary focus of the Evaluation of Enhanced Academic Instruction in After-School Programs is to assess the impact of the enhanced after-school programs on student achievement. The present chapter focuses on the first two research questions for the 12 centers implementing the enhanced reading program for two years:

- What is the impact of offering students the opportunity to participate in the enhanced reading program *for one school year*?
- Is this impact different in the second year of program implementation than in the first year?

These two questions are answered by comparing the outcomes of students who were randomly assigned to participate in the enhanced after-school reading program for one school year with the outcomes of students who were randomly assigned to remain in the regular after-school program during that same school year. Impacts are estimated for each year of implementation separately and then compared.

Before presenting the impact findings, however, the chapter begins by providing two key pieces of background information. First, the chapter provides a brief description of the sample of students included in this analysis. Then, in order to contextualize the magnitude of the impact findings, the chapter provides a comparison of the academic services received by students in the enhanced after-school reading program relative to students in the regular after-school program — that is, the service contrast.

Characteristics of Students in the Reading Sample

As explained in Chapter 2, two cohorts of students were randomly assigned to enroll in either the enhanced after-school reading program for one school year (enhanced program group) or to remain in the regular after-school program during that time (regular program group). Students who were randomly assigned in the first implementation year comprise the "Cohort 1" sample; this sample is used to estimate the impact of the enhanced program in the first year of implementation. Students who applied for the opportunity to be randomly assigned in the second year of the study — and who were not enrolled in the enhanced program in the first year of the study — comprise the "Cohort 2" sample (see Figure 2.2); this sample is used to estimate the impact of the enhanced program in the second year of implementation. The analyses

presented in this chapter are based on data from both of these cohort-specific samples and are limited to students with one-year of follow-up data from both the evaluation-administered achievement test and the regular school-day teacher survey.

Table 8.1 presents the characteristics of students in the Cohort 1 and Cohort 2 samples. As seen in this table, within each cohort sample, there are statistically significant differences on individual characteristics between students in the enhanced and regular after-school program groups. Additionally, an overall F-test indicates that there is a systematic difference in the background characteristics of students in the enhanced and regular program groups, for the two cohort-specific samples. This means that, taken together, individual differences between the enhanced and regular program group are greater than what would be predicted by chance.[109] This difference is primarily driven by a difference between the enhanced and regular program groups in terms of household composition in the Cohort 2 sample (students in the enhanced program group are more likely to come from a single-adult household) and a difference in baseline reading test scores in the Cohort 1 sample (students in the enhanced group have lower baseline scores on average). The difference in baseline test scores is especially important because reading achievement is also a key outcome measure in this evaluation. In order to address this issue, measures of student characteristics were included in the impact model (among them students' fall pretest score) in order to control for observed differences between the enhanced and the regular program group at baseline. (See Appendix G for a detailed description of the statistical model and sensitivity tests that were used to validate this approach.)

Characteristics of students in Cohort 1 presented in the top panel of Table 8.1 indicate that the majority of students in the enhanced program group are black (39 percent) or Hispanic (38 percent). Approximately half of these students (48 percent) are male; 16 percent are overage for grade; 83 percent are eligible for free or reduced-price lunch; and 29 percent lived in a household with a single adult. Twenty-five percent of students in the analysis sample had a mother who did not finish high school, while 31 percent had a mother with a high school diploma or a General Educational Development (GED) certificate. And students in the analysis sample are approximately equally distributed across grades. Finally, at their enrollment in the study, 89 percent of students in Cohort 1 were performing at a level defined by the publisher of the achievement test used in this study as below proficient in reading.[110]

[109]Note that baseline differences between the enhanced and regular program group were also found in the first report for the 25 after-school reading centers that participated in the first year of the study (Black et al., 2008). The Cohort 1 sample represents a subset of the students included in the sample for the first-year report.

[110]As mentioned earlier in the report, local staff used a variety of measures to recommend students for the program. However, because performance standards for these measures may differ from those of the study-administered baseline test, 11 percent of students in Cohort 2 identified by local staff as in need of supplemental support and randomly assigned into either the enhanced or regular program group tested at or above the proficient level on the study-administered baseline test (SAT 10).

The Evaluation of Academic Instruction in After-School Programs

Table 8.1

Baseline Characteristics of Students in the Reading Analysis Sample, by Cohort (One Year of Service)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Cohort 1[a]						
Enrollment						
2nd grade	231	128	103			
3rd grade	219	121	98			
4th grade	231	129	102			
5th grade	224	126	98			
Total	905	504	401			
Race/ethnicity (%)						
Hispanic		38.44	41.50	-3.06	-0.06	0.19
Black, non-Hispanic		39.14	35.99	3.15	0.06	0.10
White, non-Hispanic		14.54	14.99	-0.45	-0.01	0.82
Asian		2.19	2.83	-0.63	-0.04	0.54
Other		5.58	4.59	0.99	0.05	0.48
Gender (%)						
Male		48.02	45.46	2.55	0.05	0.45
Average age (years)		8.61	8.55	0.05	0.09	0.13
Overage for grade[b] (%)		16.27	13.11	3.16	0.08	0.19
Free/reduced-price lunch (%)						
Eligible (among information providers)		82.88	83.04	-0.16	0.00	0.94
No information provided		4.56	4.90	-0.33	-0.02	0.82
Average household size		2.11	2.01	0.09	0.08	0.23
Single-adult household (%)		29.22	28.35	0.87	0.02	0.77
Mother's education level (%)						
Did not finish high school		25.20	19.87	5.33	0.11	0.06
High school diploma or GED certificate		31.35	25.87	5.48	0.11	0.07
Some postsecondary study		38.29	47.30	-9.00 *	-0.17	0.01
No information provided		5.16	6.96	-1.80	-0.07	0.27
SAT 10 baseline reading total scaled scores		565.66	571.03	-5.37 *	-0.16	0.00
Vocabulary/word reading[c]		556.71	563.73	-7.01 *	-0.16	0.01
Reading comprehension		566.12	572.87	-6.76 *	-0.18	0.00
Word study skills[d]		575.99	577.77	-1.78	-0.04	0.45
Sample size (total = 905)		504	401			

(continued)

Table 8.1 (continued)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Cohort 2[e]						
Enrollment						
2nd grade	199	117	82			
3rd grade	149	79	70			
4th grade	133	74	59			
5th grade	145	82	63			
Total	626	352	274			
Race/ethnicity (%)						
Hispanic		38.71	41.00	-2.29	-0.04	0.43
Black, non-Hispanic		36.35	38.42	-2.07	-0.04	0.38
White, non-Hispanic		17.90	14.41	3.49	0.10	0.17
Asian		2.18	2.98	-0.80	-0.05	0.51
Other		4.95	3.34	1.60	0.08	0.32
Gender (%)						
Male		56.19	48.05	8.14	0.15	0.05
Average age (years)		8.58	8.53	0.04	0.07	0.32
Overage for grade[b] (%)		14.40	13.60	0.80	0.02	0.79
Free/reduced-price lunch (%)						
Eligible (among information providers)		81.82	83.17	-1.35	-0.04	0.61
No information provided		4.91	3.10	1.81	0.08	0.30
Average household size		1.99	2.20	-0.21 *	-0.17	0.02
Single-adult household (%)		32.41	21.60	10.81 *	0.23	0.00
Mother's education level (%)						
Did not finish high school		22.32	26.32	-4.01	-0.09	0.26
High school diploma or GED certificate		27.30	29.50	-2.20	-0.05	0.56
Some postsecondary study		44.82	39.27	5.55	0.10	0.18
No information provided		5.57	4.91	0.66	0.02	0.73
SAT 10 baseline reading total scaled scores		570.96	572.84	-1.87	-0.06	0.42
Vocabulary/word reading[c]		562.08	562.79	-0.71	-0.02	0.83
Reading comprehension		571.86	574.05	-2.19	-0.06	0.42
Word study skills[d]		579.82	580.96	-1.14	-0.03	0.69
Sample size (total = 626)		352	274			

(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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school

Table 8.1 (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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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 average observed mean for members randomly assigned to the enhanced program group. The regular program group values in the next column are the average 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined.

F-tests were calculated for the analysis 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. The F-value for the Cohort 1 analysis sample (F = 2.83) and the Cohort 2 analysis sample (F = 2.07) are significant at the 5 percent level.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

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

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

[e]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

As seen in the bottom panel of Table 8.1, the majority of Cohort 2 students in the enhanced program group are black (36 percent) or Hispanic (39 percent). Just over half of these students (56 percent) are male; 14 percent are overage for grade; 82 percent are eligible for free or reduced-price lunch; and 32 percent lived in a household with a single adult. Twenty-two percent of students in the analysis sample had a mother who did not finish high school, while 27 percent had a mother with a high school diploma or a General Educational Development (GED) certificate. Additionally, because Cohort 2 excludes students who were randomly assigned in the second year but were offered the enhanced program in the first year (given that this sample is used to estimate one-year impacts of enhanced services), the sample includes a proportionate-

ly larger percentage of students in grade two (32 percent) than in other grades.[111] Finally, at their enrollment in the study, 92 percent of students in Cohort 2 were performing at a level defined by the publisher of the achievement test used in this study as below proficient in reading.[112]

The Academic Service Contrast Between the Enhanced and Regular After-School Programs

This section describes the extent to which the academic support services received by students in the enhanced program group differ from the "business as usual" services received by students in the regular after-school program group. This service contrast is what underlies the impact on student outcomes of being enrolled in the enhanced after-school reading program for one year, which will be reported later in this chapter.

The service contrast that underlies the impacts is described through five interrelated findings: the content of the service offerings, the experience and training of the staff members, overall student attendance in the after-school program, the extent of academic instruction in reading, and, finally, student academic support from other sources. The following sections present detailed findings on each of these topics, drawing on data from surveys of after-school program staff, attendance records, and surveys of student and regular-school-day teacher surveys.

Differences in Content of the Service Offerings

On the survey of after-school program staff, instructors in the regular after-school program were asked about the nature of the academic services offered in the regular after-school program to assess whether the nature of the content offered was different from support for students in the enhanced program group.[113] Figure 8.1 describes the reported academic services provided by regular program staff and highlights the type of support that is most similar to the enhanced after-school program — academic instruction in reading.

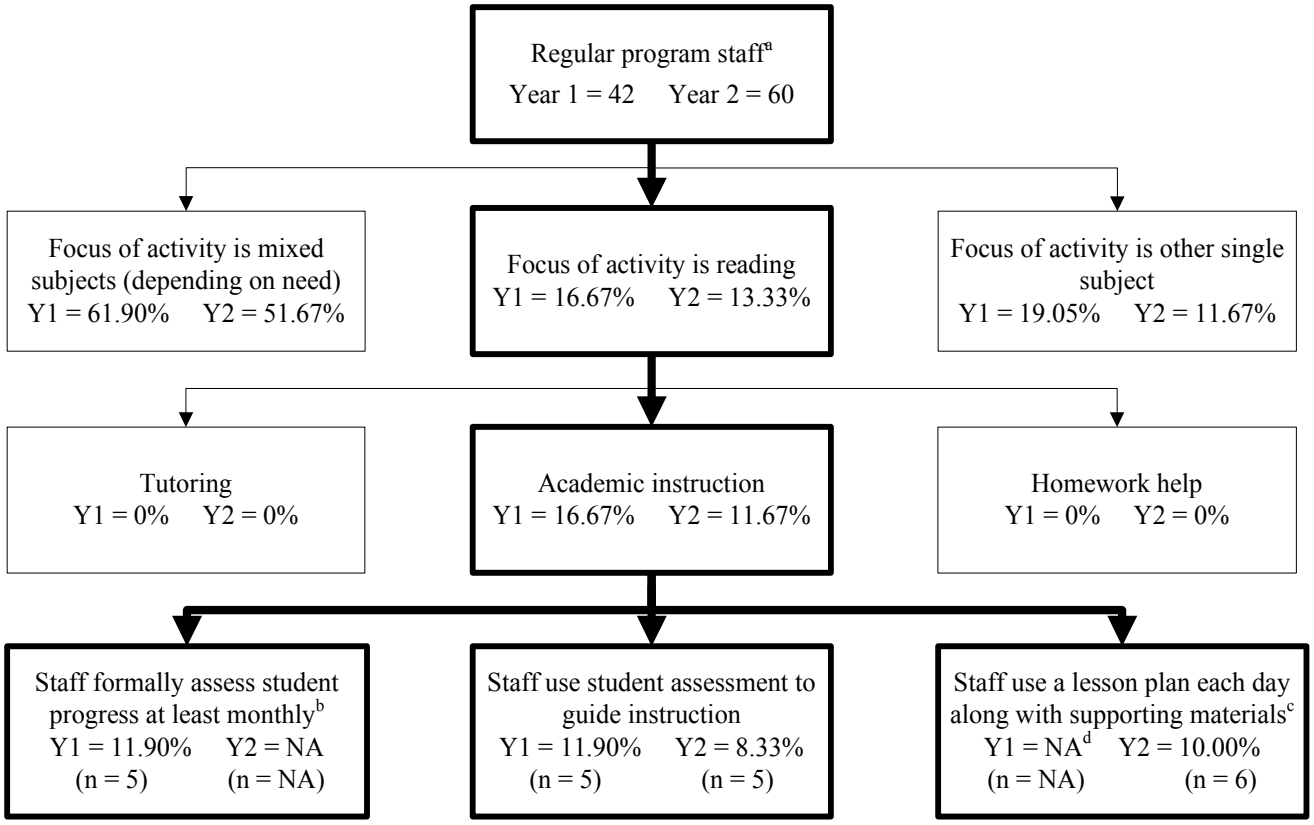
---

[111]Weights are used to ensure that grade two students do not have a disproportionately greater weight in the Cohort 2 sample findings (see Appendix G for a discussion of these weights).

[112]Again, local staff used a variety of measures to recommend students for the program. However, because performance standards for these measures may differ from those of the study-administered baseline test, 8 percent of students in Cohort 2 identified by local staff as in need of supplemental support and randomly assigned into either the enhanced or regular program group tested at or above the proficient level on the study-administered baseline test (SAT 10).

[113]In the regular after-school program, some staff members provided academic support to students, while other staff members were primarily involved in enrichment or recreational activities. The results presented in this section are based on staff in the former group only. Percentages are based on the number of staff who responded to the survey.

Figure 8.1
Academic Services Offered by Regular After-School Program Staff at Centers Implementing the Enhanced Reading Program



(continued)

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

NOTES: Percentages are calculated as the average of the regular after-school program staff in each year: 42 staff in Year 1 and 60 staff in Year 2.

[a]Across both years, of the 102 staff who filled out the survey, 15 staff (14.71 percent) did not respond to any of these questions.

[b]This question was only on Year 1's Evaluation of Academic Instruction in After-School Programs after-school staff survey; thus, values for Year 2 are not applicable (NA).

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

[d]Values have been suppressed to protect respondent confidentiality.

In the first year of implementation, 42 staff taught the regular after-school program, and, among them, 17 percent (seven instructors) reported providing some form of reading instruction beyond tutoring or homework help. Among these seven instructors, five reported that they formally assess student progress on a monthly basis and use student assessments to guide their instruction.

In the second year, 60 staff taught the regular after-school program, and, among them, 12 percent (seven instructors) reported providing some form of academic instruction in reading beyond tutoring or homework help. Among these seven instructors, five reported that they use student assessments to guide their instruction[114] and six reported that they provide instruction using a daily lesson plan and supporting materials.[115]

Responses to the after-school staff survey indicate that, when staff reported providing academic instruction in reading, they were providing at least one key element of the enhanced afterschool reading program — use of a structured after-school reading curriculum, frequent assessments to guide instruction, and/or use of a daily lesson plan. Hence, the reading instruction that the 17 percent of regular after-school staff in Year 1 and 12 percent in Year 2 indicated they provided was likely similar in nature to the enhanced program, thus dampening the service contrast in the study.[116]

Differences in Staff Providing Academic Support Services

Differences in the staffing strategy and support provided to staff for those offering academic support in the enhanced program group compared with those in the regular program group are also illustrated in the responses to the surveys of after-school program staff.[117]

Characteristics of Staff

Table 8.2 presents information on the characteristics of staff members in the enhanced and regular after-school programs. As shown in this table, staff members in the two types of

[114]In the second-year survey, staff were not asked whether they assessed student progress on a monthly basis.

[115]As part of the field research in the second year, two randomly selected regular program instructors in each after-school center were interviewed. These interviews were used to further explore the nature of the academic services provided by the regular program instructors. However, of the seven second-year instructors who reported providing reading-focused instruction, only one was part of the randomly selected staff to be interviewed. Therefore, these findings are not discussed.

[116]The difference between implementation years in the percentage of regular program staff who report providing academic instruction in reading is not statistically significant (p-value = 0.49)

[117]In the regular after-school program, some staff members provided academic support to students, while other staff members were primarily involved in enrichment or recreational activities. The results presented in this section are based on staff in the former group only (which includes 42 staff from the first year and 60 from the second year). Percentages are based on the number of staff who responded to the survey.

The Evaluation of Academic Instruction in After-School Programs

Table 8.2

Characteristics of After-School Staff at Centers Implementing the Enhanced Reading Program

Service Offering	Enhanced Program	Regular Program	Estimated Difference		P-Value for the Estimated Difference
First implementation year					
Certified in elementary education (%)	98.08	47.50	50.58	*	0.00
Years of elementary school teaching experience (%)					
0-2 years	11.54	45.00	-33.46		
3-4 years	11.54	7.50	4.04		
More than 4 years	76.92	47.50	29.42		
			chi-square	*	0.01
Staff-youth ratio (youth enrolled)	1:9	1:15	-5.52	*	0.00
Sample size (total = 98)	56	42			
Second implementation year					
Certified in elementary education (%)	100.00	46.67	53.33	*	0.00
Years of elementary school teaching experience (%)					
0-2 years	8.00	36.36	-28.36		
3-4 years	20.00	4.55	15.45		
More than 4 years	72.00	59.09	12.91		
			chi-square	*	0.02
Staff-youth ratio (youth enrolled)	1:9	1:12	-2.91	*	0.00
Sample size (total = 120)	60	60			

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

NOTES: This table reflects staff in the first and second year of the study in the 12 centers that implemented the program in both years. 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 any given characteristic varies by as much as 19 for the enhanced program group and 18 for the regular program group due to nonresponse on particular survey items. Staff for whom values are missing are not included in the calculations.

programs differ on several dimensions, and these differences were evidenced in both years of program implementation.

The top panel of Table 8.2 — which presents the characteristics of staff in the first implementation year — shows that staff members in the regular after-school program were less likely to be certified teachers. Forty-eight percent of regular program staff members were certified teachers, compared to 98 percent of enhanced program staff. This difference is statistically significant at the 5 percent level.

Regular program staff also had less teaching experience. Forty-eight percent of regular program staff had more than four years of elementary teaching experience (compared with 77 percent of enhanced program staff), while 45 percent had two years or fewer of elementary school teaching experience (compared with 12 percent of enhanced program staff). The overall difference in teaching experience between the two types of program is statistically significant.

The regular after-school program was also characterized by a higher staff-to-youth ratio. The staff-to-youth ratio was 1:15 on average in the regular after-school program, while the enhanced after-school program had an average staff-to-youth ratio of 1:9. This difference is also statistically significant.

The bottom panel of Table 8.2 shows that this pattern of differences in characteristics between staff in the enhanced and regular program is consistent across implementation years. And the difference in characteristics of staff members between the two years of implementation is not statistically significant.[118]

Support for Staff

The top panel of Table 8.3 — which describes the support provided to staff in the first implementation year — shows that staff in the regular after-school program were less likely than staff for the enhanced program to report having received high-quality training to carry out their work (51 percent and 100 percent, respectively, p-value = 0.00) or to report receiving ongoing support for how to teach children in their Adventure Island activity (56 percent and 96 percent, respectively, p-value = 0.00).

Regular program staff members were also less likely to report receiving paid daily preparation time. Sixty-two percent of regular program staff reported getting less than 30 minutes a day of paid preparation time, and 39 percent reported getting 30 minutes or more. In

[118]P-values for the test of the difference of service offering measures across implementation years are 0.32, 0.37, and 0.78, respectively, for certification, years of experience, and the staff-to-youth ratio.

The Evaluation of Academic Instruction in After-School Programs

Table 8.3

Support for After-School Staff at Centers Implementing the Enhanced Reading Program

Service Offering	Enhanced Program	Regular Program	Estimated Difference		P-Value for the Estimated Difference
First implementation year					
High-quality training to carry out activity[a] (%)	100.00	51.35	48.65	*	0.00
Ongoing support from district for how to teach children in activity[a] (%)	96.15	56.41	39.74	*	0.00
Amount of paid preparation time to carry out activity (%)					
No minutes to less than 30 minutes per day	25.49	61.54	-36.05		
30 or more minutes per day	74.51	38.46	36.05		
			chi-square	*	0.00
Sample size (total = 98)	56	42			
Second implementation year					
High-quality training to carry out activity[a] (%)	97.92	66.67	31.25	*	0.00
Ongoing support from district for how to teach children in activity[a] (%)	97.92	79.55	18.37	*	0.01
Amount of paid preparation time to carry out activity (%)					
No minutes to less than 30 minutes per day	34.69	64.44	-29.75		
30 or more minutes per day	65.31	35.56	29.75		
			chi-square	*	0.00
Sample size (total = 120)	60	60			

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

NOTES: This table reflects staff in the first and second year of the study for the 12 centers that implemented the program in both years. 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 15 for the enhanced program group and 19 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 presents percentages of after-school staff who responded "sort of true" or "very true" when surveyed.

comparison, 75 percent of enhanced program staff received 30 minutes or more of paid preparation time — a difference of 36 percentage points. The overall difference in paid preparation time between the two types of after-school program is statistically significant.

The bottom panel of Table 8.3 shows that this pattern of differences between staff in the enhanced and regular program in the first year is consistent with what occurred in the second year. And the differences between the two years of implementation — with respect to the support provided to staff members — are not statistically significant.[119]

Differences in Attendance and Hours of Academic Instruction in the After-School Program

Table 8.4 presents information on student attendance on the days that the enhanced program operated during the school year, as well as the yearly amount of after-school reading instruction received by students. In both years, nearly all students assigned to the enhanced program for one year participated in the enhanced services (fewer than five students attended zero days and received zero hours of instruction). The top panel presents yearly attendance and hours of instruction for the first implementation year (Cohort 1) while the bottom panel presents this information for the second implementation year (Cohort 2).

In the first year of implementation, students in the enhanced program group were offered the Adventure Island program for 104 days and attended 83 days, while students in the regular program group attend 79 days of the regular after-school program. Thus, students in the enhanced program group attended the after-school program for four days more on average than students in the regular program group (this difference is statistically significant, p-value = 0.02). However, in the second year of implementation, the difference between the two program groups' attendance is not statistically significant. The difference between the two years of implementation — with respect to attendance — is statistically significant (p-value = 0.02).

The amount of reading instruction received by students in the enhanced program group is statistically higher than that of students in the regular program group in each of the two implementation years (p-values = 0.00). In the first year of implementation (Cohort 1), students in the enhanced program group received 54 more hours of reading instruction than students in the regular program group, while in the second year of implementation (Cohort 2), this difference is 56 hours. The net difference in instructional reading hours between implementation years is not statistically significant (p-value = 0.63). This approximately 55 hours of extra

[119]P-values for the test of the difference in measures of support provided to staff across implementation years are 0.32, 0.59, and 0.77, respectively, for "received high quality training," "ongoing support," and "paid preparation time."

The Evaluation of Academic Instruction in After-School Programs

Table 8.4

Attendance of Students in the Reading Analysis Sample (One Year of Service)

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]					
Attendance in after-school program[b]					
Number of days attended	82.94	78.72	4.22 *	0.13	0.02
Total hours of reading instruction received[c]	63.49	9.02	54.47 *	2.58	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[d]					
Students receiving instruction (%)	34.19	23.96	10.24 *	0.21	0.00
Number of days per week[e]	1.05	0.57	0.48 *	0.31	0.00
Regular school day[f]					
Students receiving special support (%)	45.09	44.25	0.84	0.02	0.77
Minutes per week of individualized help	69.07	66.01	3.05	0.04	0.56
Sample size (total = 905)	504	401			
Cohort 2[g]					
Attendance in after-school program[b]					
Number of days attended	74.21	76.23	-2.02	-0.06	0.31
Total hours of reading instruction received[c]	63.52	8.01	55.51 *	2.63	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[d]					
Students receiving instruction (%)	42.23	29.29	12.94 *	0.27	0.00
Number of days per week[e]	1.23	0.82	0.41 *	0.26	0.00
Regular school day[f]					
Students receiving special support (%)	45.06	44.52	0.54	0.01	0.88
Minutes per week of individualized help	45.20	44.37	0.83	0.01	0.88
Sample size (total = 626)	352	274			

(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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

The estimated impacts are regression-adjusted using ordinary least squares, controlling for indicators of

Table 8.4 (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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

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

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

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 or 60, 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.

[d]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 four students who did not complete a survey.

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

[f]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.

[g]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

reading instruction between the enhanced and regular program group (54 for Cohort 1 and 56 for Cohort 2) was made up of approximately 73 sessions of 45 minutes each. It represents an estimated 23 percent increase in reading instruction over and above what is received during the regular school day.[120]

The students in the regular program group received fewer hours of instruction in reading instruction in both implementation years for two related reasons. First, while *all* staff in the enhanced program group provided reading instruction, 17 percent of regular program staff in the first year of implementation and 12 percent in the second year reported providing academic instruction in reading. Second, in the first year of implementation, students in the enhanced program group had higher attendance than students in the regular program group, as noted above.

Differences in Academic Support from Other Sources

This section examines whether students in the regular after-school program group sought out other supplemental reading programs outside of school — or whether they received additional help from their school-day teachers — in response to not having been selected to enroll in the enhanced after-school program. The second section in each panel of Table 8.4 presents findings for academic support from other nonschool sources and during the regular school day, based on student surveys, as well as surveys of regular-school-day teachers.

On the follow-up student survey, students were asked whether they attended a reading class or reading-related activity outside of the regular school day that was not part of the after-school program.[121] They were also asked how many days per week they attended this class or activity. Within each implementation year, students in the enhanced program group reported a statistically significantly greater amount of participation in a reading class or activity outside of school, compared to students in the regular program group. Specifically, in Cohort 1, 34 percent of students in the enhanced program group reported such participation compared to 24 percent of students in the regular program group (p-value = 0.00). And enhanced program group students participated in this type of activity 1.05 days per week, on average, while the

[120]This percentage increase is based on information about the number of minutes of school-day reading instruction. 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. Hence, the 55 additional hours of reading instruction received by students in the enhanced program group represents a 23 percent increase in instructional time in reading.

[121]These data are student self-reports of academic support received and are subject to bias inherent in such a method of data collection; however, there is no reason to believe that such bias would differ for enhanced program students compared to regular program students.

regular program group participated, on average, 0.57 day per week (p-value = 0.00). The difference between implementation years in participation in a reading class or activity outside of school, and days per week of participation, is not statistically significant (p-values are 0.55 and 0.63, respectively).

Additionally, surveys of regular-school-day teachers of students in the sample asked whether each of these students received "any special support in reading during the school day, such as pull-out tutoring, reading recovery, or a computer-assisted lab." Teachers were also asked to report the number of minutes of individualized instruction that they or an aide provided each student in the sample in reading during the prior week.

In both Cohort 1 and Cohort 2, there are no statistically significant differences in the amounts of individualized instruction received by students in the enhanced and regular program groups, nor is there a statistically significant difference in the percentage of students in each program group who received special in-school support. These findings do not differ by a statistically significant amount across the two years of implementation.[122]

Impacts on Student Achievement and Other Outcomes

This section examines whether one year of access to the enhanced after-school reading program improves student achievement and investigates whether this impact differs across the first and second years of program implementation. In addition to examining impacts on reading achievement, the effect of the enhanced program is also estimated for three academic behaviors: homework completion, attentiveness, and disruptiveness in class. When interpreting these impact findings, the key service contrast result to bear in mind is that students in the enhanced program group received 54 more hours of reading instruction than students in the regular program group in the first year and 56 more hours in the second year.

Impacts on Student Achievement

In the spring of each study year, the Stanford Achievement Test 10th Edition (SAT 10) abbreviated battery in reading was administered to all students in the sample.[123] Total scores on the reading test, as well as scores on three subtests — vocabulary, reading comprehension, and word study skills (for grades two through four) — are used to measure individual students' academic achievement in reading. In addition, the Dynamic Indicators of Basic Early Literacy

[122]P-values for the test of the difference in in-school support services across implementation years are 0.10 and 0.34, respectively, for the percentage of students receiving special support and for minutes of individualized help.

[123]Spring 2006 for Cohort 1 and Spring 2007 for Cohort 2.

Skills (DIBELS) oral reading fluency (ORF) and nonsense word fluency (NWF) tests were administered to students in the second and third grade.

The top panel of Table 8.5 presents the impact on SAT 10 reading scores and the DIBELS for students in the Cohort 1 sample. As seen in this table, one year of access to the enhanced reading program did not have a statistically significant effect on SAT 10 total reading scores. The first two bars in the top graph of Figure 8.2 places these impact estimates within the context of the actual and expected growth in total reading scores for students in the enhanced program group. The dark bar in the graph represents the actual growth of students in the enhanced program group, which for SAT 10 total scores was 21.83 points over the school year. The light bar in the graph represents the growth in test scores for the regular program group; this growth of 24.42 points provides the best indication of what the regular program group would have achieved on the SAT 10 had they not enrolled in the enhanced after-school reading program.[124] Thus, the estimated impact of the program is -2.59 scaled score points on SAT 10 total scores, which is not statistically significant.

To investigate whether specific *types* of reading knowledge are affected by the enhanced reading program, impacts in the first year were also examined for the three subtests embedded in the SAT 10 (vocabulary, reading comprehension, and word study skills) and for two fluency measures for students in grades two and three. As seen in Table 8.5 (and Figure 8.2 for SAT 10 measures), students in the enhanced program group had statistically significantly fewer gains on the reading comprehension SAT 10 subtest than students in the regular program group. Specifically, the reading comprehension score for students in the enhanced program group is 3.6 scaled score points lower than that of their counterparts in the regular program group (p-value = 0.04).[125] However, access to the enhanced program did not have a statistically significant effect on the other SAT 10 subtests or the two DIBELS fluency measures.

After-school teachers and centers potentially became more experienced with the delivery of the intervention in the second year. Thus, to determine whether the impact of offering students the opportunity to enroll in the enhanced after-school program differed from the first to

[124]The fall-to-spring growth in test scores for students in the sample is 24 scaled score points, based on the abbreviated SAT 10 test, whereas the fall-to-spring average growth for a nationally representative sample of students in grades two through five is 10 scaled score points, based on the full-length SAT 10 test. However, note that the average growth among the study sample may be partially attributable to regression to the mean. Regression to the mean is a statistical artifact that makes random variation in repeated data look like true growth. Specifically, even in the absence of true growth, students with below-average SAT 10 scores at baseline (such as the students in this sample) would score closer to the national mean on the follow-up test than they did on the baseline test, due to measurement error in the SAT 10 assessment.

[125]Given that the impact on the SAT 10 *total* score is not statistically significant, the statistical significance of the estimated impact on the reading comprehension subtest could be a Type I error and thus should be interpreted with caution.

The Evaluation of Academic Instruction in After-School Programs

Table 8.5

Impact of the Enhanced Reading Program on Student Achievement in the Reading Analysis Sample (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
Cohort 1[a]					
SAT 10 reading total scaled scores	588.66	591.25	-2.59	-0.08	0.06
Vocabulary	582.73	584.84	-2.12	-0.05	0.29
Reading comprehension	589.47	593.01	-3.55 *	-0.10	0.04
Word study skills (grades 2-4)[b]	589.44	590.04	-0.61	-0.01	0.81
DIBELS (grades 2-3)[c]					
Oral fluency score	73.61	71.93	1.68	0.05	0.44
Nonsense word fluency score	66.19	63.82	2.37	0.07	0.32
Sample size (total = 905)	504	401			
Cohort 2[d]					
SAT 10 reading total scaled scores	593.95	593.68	0.27	0.01	0.88
Vocabulary	587.45	585.92	1.53	0.03	0.56
Reading comprehension	595.75	596.99	-1.24	-0.03	0.56
Word study skills (grades 2-4)[b]	593.64	592.12	1.52	0.04	0.62
DIBELS (grades 2-3)[c]					
Oral fluency score	78.91	75.82	3.10	0.09	0.21
Nonsense word fluency score	75.52	70.59	4.94	0.14	0.11
Sample size (total = 626)	352	274			

(continued)

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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 374 to 787, 439 to 777, 412 to 739, and 410 to 740. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximum 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-

Table 8.5 (continued)

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 374 to 787, 439 to 777, 412 to 739, and 410 to 740. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximum 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98; nonsense = 36.13. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

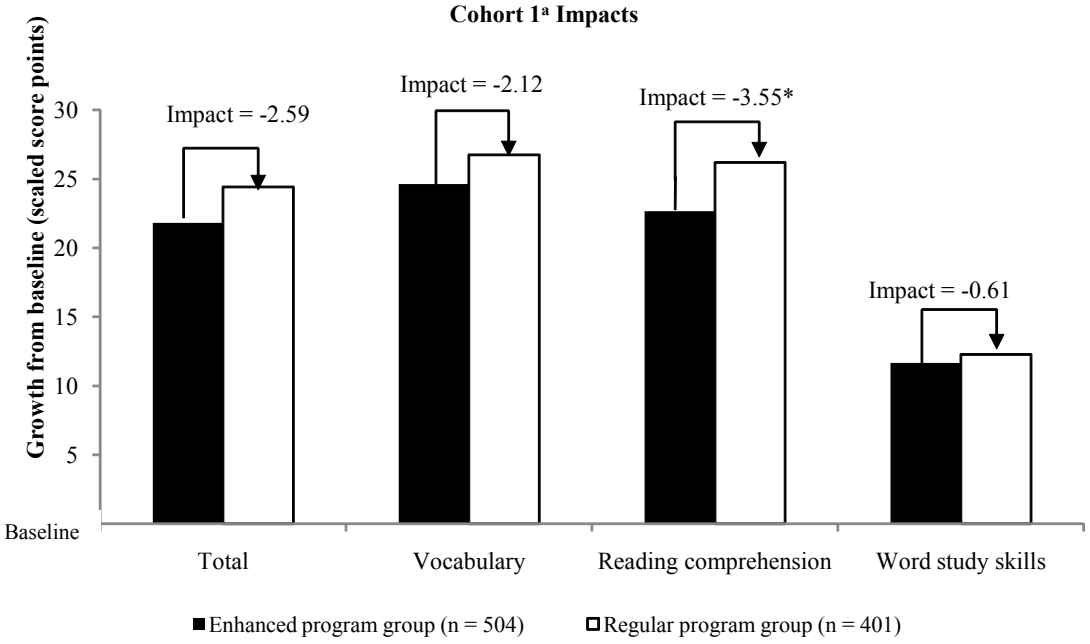
[c]The DIBELS sample includes only second- and third-grade students because the nonsense word fluency subtest and the oral fluency subtest were not administered to fourth- and fifth-grade students in both study years. Impacts on the SAT 10 for second- and third-grade students are of similar magnitude and direction as the SAT 10 impacts presented in this table for all grades combined. (SAT 10 impacts do not differ by a statistically significant amount for second- and third-grade students compared to fourth- and fifth-grade students.)

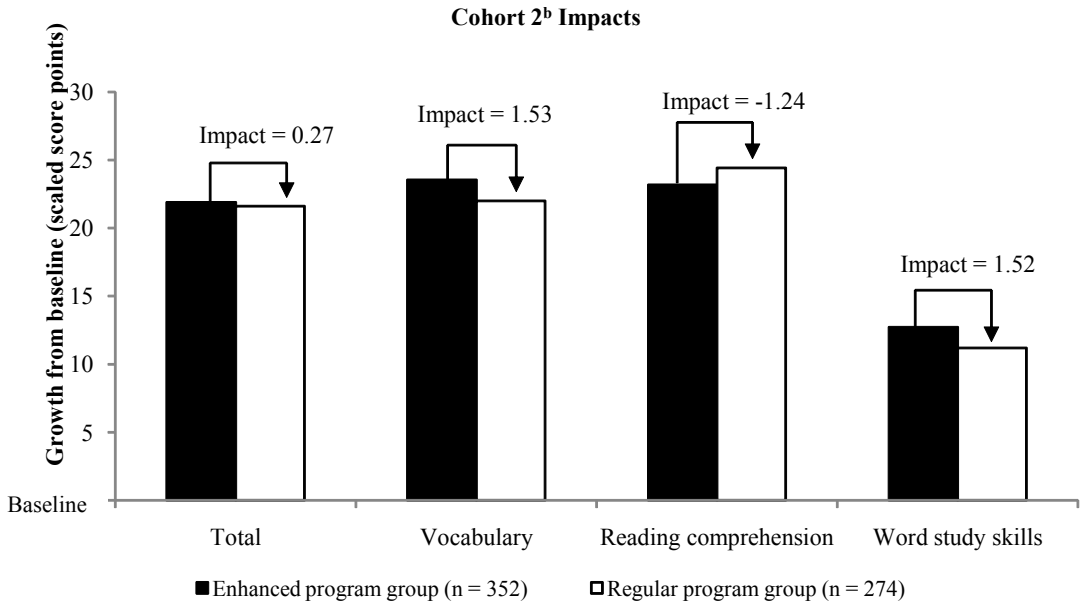
[d]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

The Evaluation of Academic Instruction in After-School Programs

Figure 8.2

SAT 10 Reading Test Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Reading Program (One Year of Service)





(continued)

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

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services

Figure 8.2 (continued)

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

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. For Cohort 1, these effect sizes are -0.08, -0.05, -0.10, and -0.01 for the reading total, vocabulary, reading comprehension, and word study skills, respectively. For Cohort 2, these effect sizes are 0.01, 0.03, -0.03, and 0.04 for the reading total, vocabulary, reading comprehension, and word study skills, respectively.

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 world 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.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

difference in impacts between implementation years (Cohort 1 and Cohort 2 samples) is not statistically significant for any measure of reading achievement.[126] Thus, it cannot be concluded that the enhanced after-school reading program was more or less effective in one implementation year than the other.

Another achievement measure of policy interest is the school district's locally administered standardized test, since scores on these tests are typically tied to local accountability provisions. For this reason, student scores on locally administered (state) tests were collected,

[126]The p-value for the difference in impacts between cohorts is 0.20, 0.26, 0.39, 0.59, 0.66, and 0.51 for the total, vocabulary, reading comprehension, word skills, oral fluency, and nonsense word fluency, respectively.

and impacts on these test scores are examined. Note, first, that because the locally administered tests were not available for second-grade students in some centers,[127] the impact analysis on locally administered tests is confined to students in grades three to five. Second, because the scale of the locally administered tests differs by site, all test scores were standardized within each study site by grade, and all estimated impacts on these tests are expressed in effect sizes. (See Appendix F for details on these outcomes measures.)

As was found for the SAT 10 total scores and DIBELS reading tests, the impact of the enhanced reading program on the locally administered reading test is not statistically significant for either of the cohort-specific samples. Nor is the difference in impacts across cohorts statistically significant (p-value = 0.67). These results can be found in Appendix Table G.4.

As noted earlier in this chapter, there are statistically significant differences in baseline characteristics between students in the enhanced and regular program groups. In order to address this problem, controls for various student characteristics were included in the impact model. Three sensitivity analyses were conducted to gauge whether these covariates adequately control for baseline differences between students in the two program groups. These three tests confirm that controlling for students' baseline characteristics — and particularly their pretest score — produces internally valid estimates of the impact of the enhanced program (see Appendix G for details on the nature and the results of these tests).[128]

Impacts on Academic Behaviors

The impact of the enhanced after-school reading program on student academic behaviors is uncertain in terms of its magnitude and direction. On the one hand, if students become better able to complete their school work, their classroom behavior may improve as a result of their enrollment in the enhanced reading program. On the other hand, the additional formal instruction that students receive in the after-school program may cause "fatigue" and, therefore,

[127]Grade two tests are not available in eight of the 12 centers in the first year and six of the 12 centers in the second year.

[128]The first two sensitivity tests examine whether the findings are robust to the specification of the impact model. In the first test, impacts are estimated for a model that does not include any background covariates. In the second test, impacts are estimated for a model that controls for prior achievement (pretest). These sensitivity analyses confirm the importance of controlling for prior achievement in the statistical model. For the third sensitivity test, impacts were estimated on a restricted sample that excludes the random assignment blocks with the largest baseline differences between the enhanced and regular program groups. Findings for this restricted sample are similar to those presented in this chapter.

Note that the robustness of the impact findings presented in this section was also tested by estimating program impacts based on the full sample (i.e., students who have SAT 10 total test scores, rather than students who have both SAT 10 scores *and* a regular-school-day teacher survey). These sensitivity tests yield similar results to those reported in this chapter (see Appendix G).

negatively affect their behavior during the regular school day. Furthermore, the enhanced program replaces time spent on homework help, which could adversely affect students' homework completion.

To assess whether the enhanced after-school program changed students' behavior in any way, impacts on three measures of academic behaviors — homework completion, attentiveness, and disruptiveness in class — were examined. These measures are drawn from the survey of regular-school-day teachers. All three measures 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 8.6 shows that one year of enrollment in the enhanced reading program did not interfere with or improve homework completion, nor did it have a statistically significant effect on the two classroom behavior measures in either of the two years of program implementation (Cohort 1 or Cohort 2 samples). Nor is the difference in impacts across implementation years (cohorts) statistically significant. However, these findings should be interpreted with caution because all three variables were measured with a single survey item, thus compromising the reliability of the measures.

The Evaluation of Academic Instruction in After-School Programs

Table 8.6

Impact of the Enhanced Reading Program on Student Academic Behavior in the Reading Analysis Sample (One Year of Service)

Student Academic Behavior Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]					
Student does not complete homework	2.35	2.34	0.01	0.01	0.87
Student is disruptive	2.20	2.15	0.04	0.04	0.52
Student is attentive	3.31	3.33	-0.02	-0.03	0.66
Sample size (total = 905)	504	401			
Cohort 2[b]					
Student does not complete homework	2.40	2.29	0.11	0.10	0.19
Student is disruptive	2.16	2.18	-0.01	-0.01	0.87
Student is attentive	3.41	3.38	0.04	0.04	0.53
Sample size (total = 626)	352	274			

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

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: homework = 1.12; disruptive = 1.10; attentive = 0.80.

The sample sizes reported represent the number of students from the analysis sample in each cohort. The sample size for each outcome varies by the number of regular-school-day teachers who responded to any given question.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

Chapter 9

Analysis of the Offer of Two School Years of Service in Reading: Sample Characteristics, Service Contrast, and Impacts

This chapter examines the third key research question in this report:

- What is the impact of offering students the opportunity to participate in the enhanced after-school reading program *for two consecutive school years*?

As explained in Chapter 2, the impact of offering students the opportunity to participate in the enhanced program for two consecutive years is estimated by comparing the outcomes of students who were randomly assigned to either the enhanced after-school program (enhanced program group) or the regular after-school program (regular program group) for two consecutive school years. As is common in experimental studies, not all students received the treatment to which they were randomly assigned. Thus, this analysis includes students assigned to two years of the enhanced program, whether or not they attended both years. In fact, 43 percent of the students assigned to the enhanced program in the fall of 2006 and then again in 2007 did not attend the afterschool program for a second year.[129] And 46 percent of the students assigned to the regular after-school program in the fall of 2006 and then again in 2007 did not attend the regular afterschool program for a second year. Hence, the impact findings presented later in this chapter are of a two-year offer of services (an intent-to-treat analysis), rather than the impact of two years of receipt of the enhanced program. This latter relationship is addressed nonexperimentally in Chapter 10.

Before presenting the impact findings, however, the chapter describes the sample of students included in the analysis and provides a comparison of the academic services offered to students in each of the two program groups across both years of implementation.

The Analysis Sample

The two-year sample used for the analysis includes 270 students; 169 (63 percent) were randomly assigned to the enhanced after-school program in both years of the study, and

[129]The most common reason for students not reenrolling in the enhanced program was that they no longer had physical access to the program, either because they had moved away or did not have a means of transportation to/from the program. This second-year nonparticipation rate across both program groups of 40 percent is lower than the student turnover seen in the prior national study of 21st Century Community Learning Centers programs (James-Burdumy et al., 2005), in which 60 percent of treatment group students did not return for the second year of the program.

101 (37 percent) were randomly assigned to remain in the regular after-school program in both years of the study. This sample is limited to students with follow-up data from both the evaluation-administered achievement test and the regular-school-day teacher survey.[130]

Table 9.1 presents the characteristics of these students in the two-year sample, for each of the two program groups. As seen in this table, the majority of students in the enhanced program group are either Hispanic (41 percent) or black (38 percent); half of students in the sample (53 percent) are male; 15 percent are overage for grade; 82 percent were eligible for free or reduced-price lunch; and 27 percent lived in a household with a single adult. Twenty-nine percent of students in the sample had a mother who did not finish high school. In addition, all students in the analysis sample were enrolled in grades two through four in the first year of the study, given the two-year nature of the treatment.[131] Finally, at the beginning of the first implementation year, 91 percent of the students in the two-year sample were performing at a level defined by the publisher of the achievement test used in this study as below proficient in reading.[132]

An overall F-test indicates that there is a systematic difference in the background characteristics of students in the enhanced and regular program groups. As in the previous chapter, this problem is addressed by including measures of student characteristics (including students' pretest scores in the fall of 2005) in the impact model in order to control for observed differences between the enhanced and the regular program group at baseline. (See Appendix H for a detailed description of the statistical model and sensitivity tests that were used to validate the sample and model.)

Finally, recall from Chapter 2 that given the size of the two-year sample (270 students), the study is equipped to detect a two-year impact of the enhanced program of 0.23 standard deviation or larger. This translates into an impact of 9.0 scaled score points on the SAT 10 total

[130]Among those in the two-year sample who did not apply to the second year of the study and did not receive the second year of program services, follow-up data were collected for 46 students in the enhanced after-school program group (E_1E_2) and 29 students in the regular after-school program group (R_1R_2).

[131]A student enrolled in grade five in the first year of the study typically could not have been offered the opportunity to enroll in the enhanced after-school program in the second year of the study, because the enhanced after-school program is only available to students in grades two through five. Seven students enrolled in grade five in the first year of the study were retained in the second year of the study; however, these students were excluded from the analysis, because assuming that the enhanced program has an impact on grade promotion, retained students in the regular program group may no longer have a counterpart in the enhanced program group.

[132]As mentioned in Chapter 2, local staff used a variety of measures to recommend students for the program. However, because performance standards for these measures may differ from those of the study-administered baseline test, not all students identified by local staff as in need of supplemental support tested below the proficient level on the study-administered baseline test (SAT 10).

The Evaluation of Academic Instruction in After-School Programs

Table 9.1

Baseline Characteristics of Students in the Reading Analysis Sample (Offer of Two Years of Service)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Enrollment						
2nd grade	100	64	36			
3rd grade	87	52	35			
4th grade	83	53	30			
Total	270	169	101			
Race/ethnicity (%)						
Hispanic		41.23	43.31	-2.09	-0.04	0.69
Black, non-Hispanic		37.68	33.53	4.16	0.08	0.28
White, non-Hispanic		12.49	17.19	-4.70	-0.13	0.32
Other		9.82	6.11	3.71	0.18	0.14
Gender (%)						
Male		52.95	46.51	6.44	0.12	0.33
Average age (years)		8.04	7.96	0.08	0.13	0.25
Overage for grade[a] (%)		14.90	9.72	5.17	0.14	0.31
Free/reduced-price lunch (%)						
Eligible (among information providers)		81.88	83.05	-1.18	-0.03	0.76
No information provided		3.09	4.74	-1.66	-0.08	0.50
Average household size		2.12	2.20	-0.08	-0.06	0.60
Single-adult household (%)		26.97	23.42	3.56	0.08	0.55
Mother's education level (%)						
Did not finish high school		28.87	21.99	6.87	0.15	0.21
High school diploma or GED certificate		32.29	21.02	11.27 *	0.23	0.05
Some postsecondary study		35.86	44.54	-8.68	-0.16	0.19
No information provided		6.86	10.62	-3.75	-0.14	0.43
SAT 10 baseline reading total scaled scores		550.65	558.66	-8.01 *	-0.24	0.04
Vocabulary/word reading[b]		541.88	552.10	-10.22	-0.23	0.08
Reading comprehension		549.90	559.96	-10.06 *	-0.26	0.02
Word study skills[c]		564.11	565.48	-1.36	-0.03	0.76
Sample size (total = 270)		169	101			

(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 enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Table 9.1 (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 enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 effect size for each characteristic is calculated as a proportion of the standard deviation for students in the two-year sample regular program group.

An F-test was calculated 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. The F-value (F = 1.67) is significant at the 5 percent level.

[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.

reading test, which is equivalent to 32 percent of the expected growth in test scores for a nationally representative sample of students in grades two through four.[133, 134, 135]

[133]The growth from "fall one year" to "spring the next school year" in average SAT 10 total reading scores for a nationally representative sample of students (based on normed averages for each grade from the test developers) with the same grade composition as the two-year sample is 28.4 scaled score points. Specifically, a weighted average of fall scores of nationally representative second-, third-, and fourth-graders is calculated where the weights are the proportion in the two-year sample that were in the second, third, and fourth grade at baseline. This weighted average is subtracted from the weighted average of spring scores of nationally representative third-, fourth-, and fifth-graders (the weights are the same as before) and derives the 28.4 point difference. Therefore, a 9.0 scaled score point impact is equivalent to 32 percent of the expected two-year improvement of nationally representative students in the same grade levels.

[134]Note that the minimum detectable effect size (MDES) for the test of the *difference* between the impact on students of their *first* year of enrollment vs. the impact on students of being offered the opportunity to enroll for two years is 0.24 standard deviation.

[135]The actual precision of estimated impacts may differ somewhat from those calculated as part of the statistical power analyses presented here and in Appendix B. These differences are due to such factors as variation across after-school centers in samples sizes, random assignment ratios, pretest scores, and outcome levels.

The Academic Service Contrast Between the Enhanced and Regular After-School Programs

This section describes the extent to which the academic support services received by students in the enhanced program group during both years of implementation differ from the "business as usual" services received by students in the regular program group. This cumulative two-year service contrast is what produces the impact of offering the enhanced after-school reading program to students in both years of the study.

As seen in Chapter 8, the services received by the enhanced and regular program group differed as intended with respect to instructional offerings and the qualifications and experience of staff, in both years of the study (i.e., for both cohorts of students). However, for the purposes of understanding the impact of offering a student the opportunity to enroll in two years of enhanced services, the other aspects of the service contrast discussed in Chapter 8 — i.e., student attendance in the after-school program, hours of after-school reading instruction, and student academic support from other sources — are less useful because they reflect the service contact over the course of only one year of enrollment. Hence, the remainder of this section examines the cumulative difference between students assigned to the enhanced and regular program groups (across both years of program implementation), for these three aspects of the service contrast, drawing on data from surveys of after-school program staff, attendance records, and surveys of students and regular-school-day teachers.

Differences in Attendance and Hours of Academic Instruction in the After-School Program

Table 9.2 presents information on student attendance on the days that the enhanced program was operating, as well as the amount of after-school reading instruction received by students in each program group. The top panel presents average attendance and instructional hours across both years of the study, while the bottom two panels present this information separately for each year of the study.

Cumulatively across both study years, students assigned to the enhanced program were offered the Adventure Island program for 194 days and attended, on average, 130 days (for an average of 103 hours of academic instruction in reading), whereas students in the regular program attended for 125 days (for an average of five hours of academic instruction in reading) over the two-year span. For days attended, the difference of five days is not statistically significant, nor is there a statistically significant difference in either study year.[136] However, for hours of reading instruction, the difference of 98 hours is statistically significant at the 5 percent level and

[136]Attendance for the regular program group is only counted for the days during which the enhanced reading program was operating.

The Evaluation of Academic Instruction in After-School Programs

Table 9.2

Attendance of Students in the Reading Analysis Sample (Offer of Two Years of Service)

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cumulative across both study years					
Attendance in after-school program[a]					
Number of days attended	130.31	125.02	5.28	0.16	0.47
Total hours of reading instruction received[b]	102.73	5.23	97.50 *	5.78	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	43.60	28.90	14.70 *	0.31	0.01
Number of days per week[d]	0.92	0.53	0.38 *	0.25	0.00
Regular school day[e]					
Students receiving special support (%)	67.22	59.56	7.65	13.99	0.21
Minutes per week of individualized help	68.75	63.86	4.88	0.06	0.52
Sample size (total = 270)	169	101			
Study year					
First year (2005-2006 school year)					
Attendance in after-school program[a]					
Number of days attended	85.90	82.36	3.55	0.11	0.24
Total hours of reading instruction received[b]	65.49	4.94	60.55 *	3.59	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	34.06	17.13	16.94 *	0.35	0.00
Number of days per week[d]	1.09	0.44	0.65 *	0.42	0.00
Regular school day[e]					
Students receiving special support (%)	47.21	41.25	5.96	10.88	0.30
Minutes per week of individualized help	87.24	90.07	-2.82	-0.04	0.81
Sample size (total = 270)	169	101			

(continued)

Table 9.2 (continued)

Attendance Measure	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Second year (2006-2007 school year)					
Attendance in after-school program[a]					
Number of days attended	44.40	42.67	1.74	0.05	0.77
Total hours of reading instruction received[b]	37.24	0.29	36.95 *	2.19	0.00
Reading support from other sources					
Out-of-school reading class or tutoring[c]					
Students receiving instruction (%)	27.59	20.42	7.17	0.15	0.17
Number of days per week[d]	0.74	0.62	0.12	0.08	0.49
Regular school day[e]					
Students receiving special support (%)	52.19	50.02	2.16	3.95	0.73
Minutes per week of individualized help	50.25	37.66	12.59	0.16	0.15
Sample size (total = 270)	169	101			

(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 enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample 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 on the days they were present, or 60 minutes in centers that met only three days a week (one center in the first year and five centers in the second year). Total hours is calculated for these students by multiplying each student's total days of attendance by 45 (or 60).

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 or 60, 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.

[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?"

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

Table 9.2 (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?"

[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.

ıe

7

Most of the service contrast as measured in hours occurred during the first year of the program (see Table 9.2, first- and second-year findings). This statistically significant decrease between implementation years in hours of instruction (p-value = 0.01) is not surprising given that 43 percent of students assigned to the enhanced program for two consecutive years did not actually reapply to the study and participate in a second year of enhanced services (and therefore received zero hours of academic after-school instruction during that year.)[138]

Differences in Academic Support from Other Sources

The second section of the first panel in Table 9.2 presents findings on the supplementary academic support services received by each program group over both years of the study, whether from non-school sources or during the regular school day. On average, across both years of the study, students in the regular program group received statistically significantly *less* out-of-school reading support (classes or tutoring) than students in the enhanced program group. Specifically, 44 percent of students in the enhanced program group reported participating in a reading class or activity outside of school at some point during the two study years, compared to 29 percent of students in the regular program group (p-value = 0.01). And enhanced program group students participated in this type of activity 0.9 day per week on average, while the regular program group participated 0.5 day per week on average (p-value = 0.00).[139]

[137]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, or 486 hours across two school years. Therefore, the 98 hours of after-school reading instruction received by students in the enhanced program group represents a 20 percent increase in instructional time over the two-year period.

[138]The exploratory analysis in Chapter 10 examines the association between *receiving* two years of enhanced services and the amount of instruction received, for students who actually participated in the enhanced program in the second year of the study.

[139]These findings are based on the follow-up student survey, administered in the spring of each school year. The survey asked students whether they attended a reading class or activity outside the regular school (continued)

Table 9.2 also shows that across both years of the study, there are no statistically significant differences in the percentage of students in the enhanced and regular program groups receiving special support during the regular school day or in the amount of individualized help received.[140]

Impacts on Student Achievement and Other Outcomes

This section examines whether being offered the opportunity to participate in the enhanced after-school reading program for two consecutive years improves student achievement. Specifically, this intent-to-treat analysis indicates what the impact may be when a school offers a program to students for two consecutive years, although approximately 43 percent of the students do not return to the program after the first year. The effect of the enhanced program is also estimated for three academic behaviors: homework completion, attentiveness, and disruptiveness in class.

Impacts on Student Achievement

In the spring of each study year, the Stanford Achievement Test 10th Edition (SAT 10) abbreviated battery in reading was administered to students in the sample.[141] Total scores on the reading test, as well as scores on three subtests — vocabulary, reading comprehension, and word study skills (for grades two through four) — are used to measure individual students' academic achievement in reading. In addition, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency (ORF) test was also administered.

Table 9.3 shows that the average total reading score of students in the enhanced program group is 5.6 points lower than that of their counterparts in the regular program group, which is statistically significant and translates into an effect size of -0.17 standard deviation (p-value = 0.04). The estimated impact of the enhanced program on both the vocabulary and reading comprehension subtests is -7.6 scale score points, both of which are also statistically

day that was not part of the after-school program. (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.

[140]These results are based on the survey of the school-day teachers of students in the sample, administered at the end of each school year. Each teacher was asked whether each student in the sample received "any special support in reading during the school day, such as pull-out tutoring, a computer lab, or a special class." Teachers were also asked to report the number of minutes of individualized instruction that they or an aide provided each sample member in reading or reading during the prior week.

[141]Spring 2006 for Year1 1 and Spring 2007 for Year 2.

The Evaluation of Academic Instruction in After-School Programs

Table 9.3

Impact of the Enhanced Reading Program on Student Achievement in the Reading Analysis Sample (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
SAT 10 reading total scaled scores	595.99	601.61	-5.63 *	-0.17	0.04
Vocabulary	590.26	597.84	-7.58 *	-0.17	0.05
Reading comprehension	596.83	604.38	-7.55 *	-0.21	0.02
Word study skills (grades 2-4)[a]	594.16	595.47	-1.31	-0.03	0.80
DIBELS					
Oral fluency score	87.89	88.03	-0.15	0.00	0.96
Sample size (total = 270)	169	101			

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: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 416 to 787, 464 to 777, 455 to 739, and 450 to 740. The DIBELS oral reading fluency scores have a minimum score of zero, but no set maximum 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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[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.

significant. However, the estimated impact on the DIBELS oral fluency measure is not statistically significant. And, when accounting for multiple test corrections using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), the estimated impact on the SAT 10 is no longer statistically significant. Therefore, this result may be due to chance.[142]

To place the impacts on the SAT 10 into context, the impact of students' first year in the enhanced program was also estimated. However, the estimated impact of students' first year of enhanced services does not statistically differ from the estimated impact on these students of assigning them to two years of enhanced services (p-value = 0.46). Thus, while it can be said that being assigned to two years of enhanced services produces significantly fewer gains on test scores than experienced by the regular program group, it cannot be concluded that assigning students to enroll in the enhanced program for two years has a different impact on their reading achievement than assigning them to enroll in one year of the enhanced program.

Figure 9.1 places these impact estimates in the context of the actual and expected two-year achievement growth of students in the enhanced program group. The figure plots the two-year growth in SAT 10 total reading scores for students in the enhanced program group, as well as the expected growth that these students would have achieved had they not been assigned to the enhanced program for two consecutive years (as represented by the growth of students in the regular program group). As another frame of reference, the figure also plots the test score growth for a nationally representative sample of students with the same grade composition in each period as the two-year sample. As shown in this figure, the SAT 10 total scores of students in the enhanced program group grew by 42.8 points across both years of the study (25.1 points in the first year and another 17.7 points in the second year). However, the test scores of students in the regular program group also grew during this period — by 48.4 points across both years (28.7 points in the first year and another 19.7 points in the second year). The higher growth rate of students in the regular program group produces the estimated impacts mentioned above, a difference of -5.6 points after two years (in favor of the regular program group). Note that the average test score growth exhibited by students in both program groups may represent a closing of the achievement gap, but it could also be partially attributable to regression to the mean.[143]

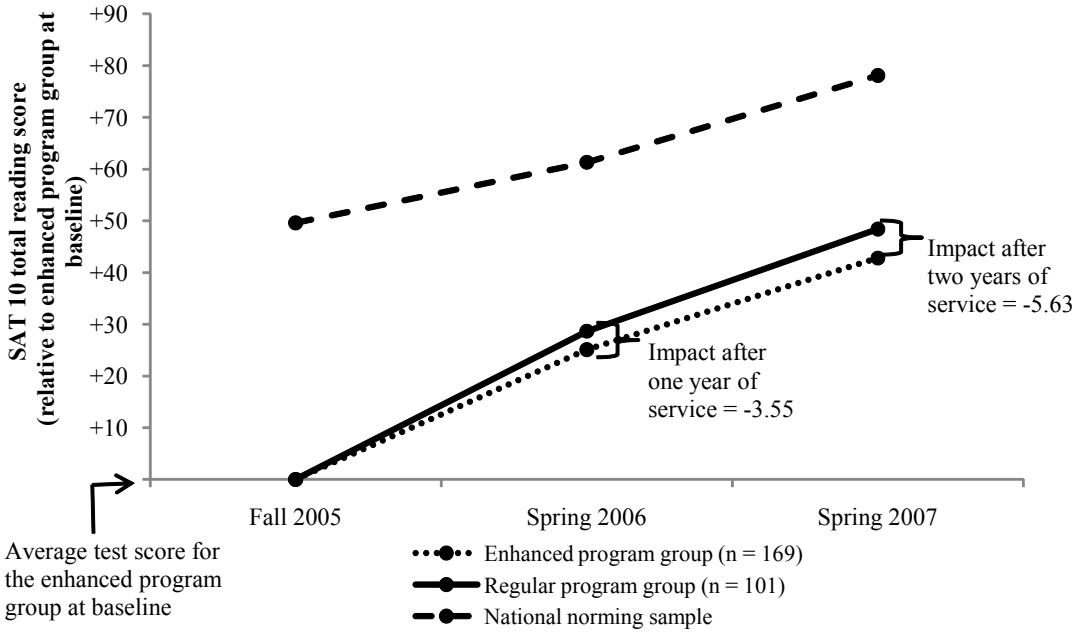
[142]Because impacts on reading achievement for all students are assessed using two measures, the SAT 10 and the DIBELS oral reading fluency test, a multiple comparison adjustment is applied.

[143]Regression to the mean is a statistical artifact that makes random variation in longitudinal data look like true growth. Specifically, even in the absence of true growth, students with below-average SAT 10 scores at baseline (such as the students in this sample) would score closer to the national mean on the follow-up test than they did on the baseline test, due to measurement error in the SAT 10 assessment.

The Evaluation of Academic Instruction in After-School Programs

Figure 9.1

SAT 10 Total Reading Scores from Baseline to Follow-Up and the Associated Impact of the Enhanced Reading Program After One Year and Two Years of Service



SOURCES: MDRC calculations are from baseline and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery. National norming sample calculations are from the SAT 10 (2002 norming sample): Stanford Achievement Test Series: Tenth Edition: Technical Data Report (Harcourt Assessment, 2004, pp. 312-338).

NOTES: The growth line for the enhanced program group is based on the observed mean baseline and follow-up test scores of students assigned to the enhanced after-school program for two consecutive years (baseline is Fall 2005; follow-ups are Spring 2006 and Spring 2007). The growth line for the regular program group represents the test scores that students in the enhanced program group would have obtained had they not been assigned to the enhanced program (calculated as the mean test score for the enhanced program group minus the estimated impact at a given time point). The growth line for the national norming sample is based on the average SAT 10 total reading scores for a nationally representative sample of students with the same grade composition in each period as the two-year sample. Specifically, at each point in time (the fall baseline, the first spring, and the second spring), the SAT 10 national norm scores for second-, third-, and fourth-graders are averaged weighting each grade average score according to their proportion in the two-year study sample at baseline. This creates an expected two-year improvement of nationally representative students at the same grade levels as this study's sample. The baseline for the national norming sample is set relative to the average baseline score of the enhanced program group.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data was not collected. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

Impacts on locally administered (state) tests were also examined, given the policy-relevance of these test scores.[144] Though not statistically significant, the estimated impact on locally administered standardized test scores of offering students in grades three through five the opportunity to participate in the enhanced program for two consecutive years is 0.06 standard deviation (p-value = 0.60).[145] Appendix Table H.4 presents the results of this analysis.

As noted earlier, the estimated impact of assigning students to the enhanced program for two consecutive years must be interpreted in light of the fact that 43 percent of students in the enhanced program group did not actually attend the program for a second year. This means that the results presented in Table 9.3 are a weighted average of the impact for students who attended both years of the enhanced program and the impact for students who did not reapply and therefore attended the enhanced program in the first year only. Thus, the results discussed in this section represent the impact of offering the enhanced program to the same students in two consecutive years, rather than the impact of receiving two years of enhanced after-school services. Because estimating the association between receiving two years of enhanced services and student outcomes cannot be estimated within the experimental framework of the study design, this question will be examined in the next chapter, which presents findings from some nonexperimental exploratory analyses.

Finally, as noted earlier in this chapter and in the previous chapter, there are statistically significant differences in baseline characteristics between students in the enhanced and the regular program groups, most notably with respect to reading pretest scores. In order to address this problem, controls for various student characteristics were included in the impact model and three sensitivity analyses were conducted to gauge whether these covariates adequately control for baseline differences between students in the two program groups. These three tests again confirm that controlling for students' baseline characteristics — and particularly their pretest scores — produces internally valid estimates of the impact of the enhanced program (see Appendix H for details on the nature and the results of these tests).[146]

---

[144]Because the scale of the locally administered tests differs by site, all test scores were standardized within each study site by grade, and all estimated impacts on these tests are expressed in effect sizes. (See Appendix F for details on these outcomes measures.)

[145]Locally administered tests are not available for students in grade two.

[146]The first two sensitivity tests examine whether the findings are robust to the specification of the impact model. In the first test, impacts are estimated for a model that does not include any background covariates. In the second test, impacts are estimated for a model that controls for prior achievement (pretest). These sensitivity analyses confirm the importance of controlling for prior achievement in the statistical model. For the third sensitivity test, impacts were estimated on a restricted sample that excludes the random assignment blocks with the largest baseline differences between the enhanced and regular program groups. Findings for this restricted sample are similar to those presented in this chapter.

Note that the robustness of the impact findings presented in this section was also tested by estimating program impacts based on the full sample instead of the analysis sample (i.e., students who have SAT 10

(continued)

Impacts on Academic Behaviors

As noted in previous chapters of this report, the expected impact of after-school academic instruction on students' behavior during the school day is of uncertain magnitude and direction. Hence, a secondary analysis was conducted to estimate the impact of being offered the opportunity to enroll in the enhanced reading program in two consecutive years on three measures of student academic behavior: homework completion, attentiveness in class, and disruptiveness in class.[147] All three measures in this domain are based on a scale that ranges from 1 to 4, with "1" indicating that the specific behavior *never* occurred and "4" indicating that it occurred *often.*

Table 9.4 shows that being assigned to the enhanced after-school program in two consecutive years did not interfere with or improve homework completion and had no statistically significant impacts on the two classroom behavior measures. However, as previously mentioned, these findings should be interpreted with caution because all three variables were measured with a single survey item, thus compromising the reliability of the measures.

total test scores rather than students who have both SAT 10 scores *and* a regular-school-day teacher survey). These sensitivity tests yield similar results to those reported in this chapter (see Appendix H).

[147]These measures are drawn from the survey of students' regular-school-day teachers.

The Evaluation of Academic Instruction in After-School Programs

Table 9.4

Impact of the Enhanced Reading Program on Student Academic Behavior in the Reading Analysis Sample (Offer of Two Years of Service)

Student Academic Behavior Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Student does not complete homework	2.57	2.39	0.17	0.15	0.23
Student is disruptive	2.30	2.32	-0.02	-0.02	0.89
Student is attentive	3.19	3.32	-0.12	-0.16	0.21
Sample size (total = 270)	169	101			

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

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: homework = 1.12; disruptive = 1.10; attentive = 0.80.

The sample sizes reported represent the number of students from the analysis sample. The sample size for each outcome varies by the number of regular-school-day teachers who responded to any given question.

Chapter 10

Exploratory Analyses of the Impact of the Enhanced After-School Reading Program

This chapter reports on a set of exploratory analyses whose purpose is to provide information that may inform the design and implementation of the enhanced reading program. However, because these analyses are nonexperimental, they should be viewed as hypothesis-generating since they may not reflect true causal relationships.

As noted in the previous chapter, not all students assigned to the enhanced program group for two consecutive years actually participated in the enhanced program in the second year. In order to provide information about the treatment for those who actually received it in both years, and to examine whether longer exposure to the program is associated with improved student outcomes, the first exploratory analysis examines the relationship between achievement and program participation for those students who participated in both years of the enhanced after-school services.

Additionally, the enhanced program was offered in a variety of different settings. Understanding how variation in the local school context, as well as variation in program implementation (across centers and the two implementation years), is associated with impacts on achievement can help one interpret the generalizability of the overall findings, as well as generate possible avenues for program improvement. Thus, the second exploratory analysis examines whether the impact of one year of enhanced services is associated with the characteristics of program implementation in the after-school center and/or with the characteristics of the local school context in which the program was implemented.

The Association Between Receiving Two Years of Enhanced After-School Reading Instruction and Student Achievement

This section examines the association between receiving enhanced after-school services for two consecutive years and reading achievement by focusing on the students in the enhanced program group who were randomly assigned to — and participated in — the enhanced after-school reading program in both years of the study.

As discussed in the exploratory analysis chapter for the math centers (Chapter 6), estimating the two-year impact for these students is challenging because students who received two years of enhanced after-school services chose to attend a second year of the enhanced program, perhaps based on factors related to their experience in the enhanced program during the first

year of the study. Because these students' decision processes are not known, it is not possible to identify students in the regular program group who would have made the same choice had they been given the option to participate. In other words, it is not clear which students in the regular program group provide the appropriate counterfactual for students in the enhanced program group who applied to and who received two years of enhanced services.

Thus, the association between receiving two years of enhanced services and reading achievement is estimated from nonexperimental methods, using an instrumental variables analysis. This technique identifies who among the regular program group are most like those in the two-year enhanced program group and essentially compares outcomes of like individuals.[148]

Table 10.1 shows that there is a statistically significantly negative association between students receiving two years of enhanced after-school services and reading achievement (-7.5 scaled score points for SAT 10 total reading scores, p-value = 0.04). However, this nonexperimental estimate of receiving two years of enhanced services does not statistically differ from the estimated impact of receiving one year of enhanced services (p-value = 0.47).

Taken together, the experimental findings from the previous chapter and the above nonexperimental findings suggest that for this population of struggling readers, two years of enhanced after-school services — whether offered or received — produces significantly fewer gains in reading achievement than experienced by the regular program group.[149] Yet, it cannot be concluded that two years in the enhanced program has a different impact on students' reading achievement than one year in the enhanced program.

---

[148]Specifically, estimated comparisons are based on students who were randomly assigned to one of three conditions: two years of enhanced services, two years of regular services, or enhanced services in the first year of the study but not the second. Based on this sample of students, impact estimates were obtained from an instrumental variable analysis in which the two treatment conditions (that is, two years of enhanced services and enhanced services in the first year but not the second) are used as instrumental variables for the number of years of enhanced services that were actually received (one year or two years). This model was fitted using two-stage least squares. Estimated associations 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. Appendix I further describes the conceptual underpinnings of the analysis and the statistical model in greater detail, as well as the sample of students included in the analysis.

[149]In order to interpret the two-year associations in Table 10.1, it is important to understand the extent to which the services received by students in the enhanced program group who applied in the second year differ from the services received by their counterparts in the regular program group who also applied in the second year. For this reason, the association between receiving two years of enhanced services and the *hours of reading instruction received* by students was estimated (see Appendix I for details). As seen in the service contrast section in the previous chapter, *offering* students the opportunity to participate in enhanced services for two years increases the amount of reading instruction that they receive by 98 hours across both years of the study. Based on an instrumental variables analysis (see Appendix I), *receiving* two years of enhanced services increases the amount of instruction by 113 hours (p-value = 0.00).

The Evaluation of Academic Instruction in After-School Programs

Table 10.1

Association Between Receiving Two Years of the Enhanced Reading Program and Student Achievement

Student Achievement Outcome	Students Who Received Two Years of Services	Estimated Counterfactual	Estimated Comparison	Standardized Estimated Comparison[a]	P-Value for the Estimated Comparison
SAT 10 reading total scaled scores	600.03	607.55	-7.51 *	-0.23	0.04
Vocabulary	596.60	605.23	-8.62 *	-0.19	0.03
Reading comprehension	600.45	609.61	-9.15	-0.25	0.09
Sample size (total = 408)[b]	NA	NA			

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, and vocabulary scaled scores, respectively, have the following possible ranges: 416 to 787, 464 to 777, 455 to 739.

Estimated comparisons are based on students who were randomly assigned to one of three conditions: two years of enhanced services, two years of regular services, or enhanced services in the first year of the study but not the second. Based on this sample of students, impact estimates were obtained from an instrumental variable analysis in which the two treatment conditions (that is, two years of enhanced services; enhanced services in the first year but not the second) are used as instrumental variables for the number of years of enhanced services that were actually received (one year or two years). This model was fitted using two-stage least squares. Estimated associations 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 "Students Who Received Two Years of Services") are the observed means for students who were assigned to and received two years of enhanced services. The values in column 2 (labeled "Estimated Counterfactual") are the estimated outcomes that these students would have obtained had they not received two consecutive years of enhanced services. Rounding may cause slight discrepancies in calculating sums and differences.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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.

[a]The standardized estimated comparison for each outcome is calculated as a proportion of the standard deviation for students in the regular program group. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[b]Group-specific sample sizes are not presented because the analysis is not based on a direct comparison of students who received two years of enhanced services to students who did not receive two years of enhanced services.

Linking the Impact of One Year of Enhanced Services on Reading Achievement with School and Program Characteristics

As discussed in Chapter 8, the estimated impact of one year of enhanced program enrollment on SAT 10 total reading scores is -2.59 scaled score points (or -0.08 standard deviation) for Cohort 1, but 0.27 scaled score points (or 0.01 standard deviation) for Cohort 2. Additionally, each year, not all centers in the study sample experienced this exact cohort-specific impact.[150] Understanding how variation in the local school context, as well as variation in program implementation, is linked to impacts on achievement may suggest settings or implementation features that may be associated with different impacts. Thus, this section explores whether the impact of one year of enhanced services on SAT 10 total reading scores in an after-school center is associated with (1) the characteristics of the school that housed the after-school center and (2) the characteristics of a center's implementation of the enhanced program. Using both study years allows these characteristics to vary both within centers over time and across centers within a given implementation year.[151]

The analysis is conducted by using a linear interaction model to estimate the association between these center characteristics and program impacts on SAT 10 total scores in the participating after-school centers in both study years (i.e., the 24 center-level impacts).[152] Because students were not randomly assigned to programs with different school characteristics, this analysis is exploratory rather than experimental; as such, these results should be viewed as hypothesis-generating rather than as establishing causal inferences.

Three measures of *program implementation* are included in the analysis: the number of days over the course of the school year that the enhanced reading program was offered (included as a measure of program dosage), whether one or more teachers teaching the enhanced program left during the school year (included as a measure of disruption in instruction), and the difference between the total hours of after-school academic instruction received by students in the enhanced program group relative to students in the regular program group (a measure of service contrast). The analysis also includes four measures of the *local school context* that capture the characteristics of the regular school day, as well as the characteristics of the school's

[150]Center-by-year impact estimates on SAT 10 total reading scores range from -11.1 scaled score points to 7.2 scaled score points. An F-test indicates that the overall variation in impacts across centers and implementation years is not statistically significant at the 5 percent level (p-value = 0.55). Nonetheless, statistically significant associations between school-level predictors and impacts may still be found, thus providing information that can be used to improve the design and implementation of the program. See Appendix J for a discussion of variation in impacts across centers and implementation years.

[151]Variation in each of the program implementation and local school context measures across centers and years is statistically significant (p-value for variation of each measure is 0.00).

[152]Twelve centers * two implementation years = 24 center-level impacts.

student body. These measures are: whether the school met its Adequate Yearly Progress (AYP) goals, whether the in-school student-to-teacher ratio is greater than in the enhanced after-school program (13:1),[153] the amount of reading instruction that students received during the regular school day,[154] and the proportion of students receiving free or reduced-price lunch. Details on these measures are provided in Appendix J.

Table 10.2 presents the estimated association between the program impacts on SAT 10 total reading scores and these school-level characteristics. None of the individual school context or program implementation characteristics are correlated with impacts on total reading scores. Therefore, associations with individual variables do not highlight aspects of program implementation or school context that are likely to improve impacts.[155]

[153]As noted in Chapter 2, the planned student-teacher ratio was 10:1; however, up to 13 students were randomly assigned to each class, in order to account for the possibility that some students might not attend on a given day.

[154]School administrators were asked how many minutes teachers spend per day teaching 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. Specifically, across both cohorts, 8 percent of schools offer fewer than 90 minutes, 42 percent offer 90 minutes, 25 percent offer 90 to 120 minutes, and the remaining 25 percent offer 120 minutes or more (rounding may cause slight discrepancies in calculating sums and differences). Thus, the natural split for this subgroup is between schools offering 90 minutes or less of school-day reading instruction and schools offering more than 60 minutes.

[155]Two additional school-level measures were available for the second year of program implementation in the reading centers. The first is the average yearly achievement gain of students in the school, which serves as a proxy for the level and quality of instruction and leadership at the school.

The second measure is the percentage of enhanced program teachers in the second year of the study who also taught during the first year (i.e., "returning" teachers). This measure is intended to gauge program implementation strength, since one would expect returning teachers to be better able to deliver the enhanced curriculum than new teachers.

Given the availability of these additional measures, a separate analysis was conducted focusing on the second year of the study only (i.e., 12 center-level impacts) and using all available school-level characteristics in the second year of the study. None of the individual school context or implementation characteristics were associated with program impacts by a statistically significant amount.

The Evaluation of Academic Instruction in After-School Programs

Table 10.2

Associations Between School and Program Characteristics and the Impact of the Enhanced Reading Program on Student Achievement After One Year of Service

Interaction Characteristic	Estimated Coefficient	P-Value for the Estimated Coefficient
School context		
More than 90 minutes of reading instruction	1.20	0.64
Student-to-teacher ratio greater than that in the enhanced program[a]	-2.89	0.33
Did not meet Adequate Yearly Progress (AYP) goals	-0.98	0.73
Percentage of student body that is low-income	0.01	0.67
Program implementation		
Total days enhanced program was offered	-0.05	0.58
Service contrast between enhanced and regular program group[b]	0.08	0.11
Enhanced teacher left the program during the school year	-1.29	0.64
F-test of all characteristics		0.45
F-test of school context characteristics		0.65
F-test of program implementation characteristics		0.30
Size of student sample (total = 1,531)		
Size of school sample (total = 12 schools times 2 years = 24)		

(continued)

SOURCES: MDRC calculations 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 (CCD) Web site, http://nces.ed.gov/ccd/. Program implementation characteristics are from the Evaluation of Academic Instruction in After-School Programs attendance data and from Bloom Associates. These data reflect the 2005-2006 and 2006-2007 school years.

NOTES: The estimated coefficients represent how the impact of the reading program on SAT 10 reading total scaled scores varies with each school characteristic. These estimates were obtained by fitting an impact model that includes an indicator of treatment status, as well as a set of interaction terms between the treatment indicator and each of the school characteristics listed above; the findings reported in the table are the coefficients of the interaction between treatment status and the school characteristics. The model also controls for random assignment strata, students' baseline reading total scaled score, race/ethnicity, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education. The F-test tested whether the coefficients of the school characteristic variables are jointly equal to zero. Within each center, the analysis sample includes, on average, 64 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]Schools are classified as having a high student-to-teacher ratio if the ratio is greater than 13:1.

[b]Service contrast is measured as the difference between the total hours of after-school academic instruction received by students in the enhanced program group relative to students in the regular program group. This difference is obtained from a regression model that estimates the impact of the enhanced program on the number of hours of after-school academic instruction received by students, controlling for random assignment strata and student characteristics. This regression model is estimated for each center in each year of the study.

Table 10.2 (continued)

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

[b]Service contrast is measured as the difference between the total hours of after-school academic instruction received by students in the enhanced program group relative to students in the regular program group. This difference is obtained from a regression model that estimates the impact of the enhanced program on the number of hours of after-school academic instruction received by students, controlling for random assignment strata and student characteristics. This regression model is estimated for each center in each year of the study.

Appendix A

Findings After the First Implementation Year and Differences Between Centers that Participated in Both Years of the Study and Centers that Participated Only in the First Year

During the first year of program operations (school year 2005-2006), the enhanced instruction was implemented in 50 after-school centers — 25 to test the reading program and 25 to test the math program. The study was then extended to include a second year of operations (2006-2007) for the purpose of examining questions that can only be answered after two years of program implementation. Continuation in the study was voluntary, and 27 of the original 50 after-school centers were able to participate in the study for another year. This appendix summarizes findings from the first report and provides a comparison of first-year impacts and implementation in the 27 after-school centers that participated in both years of the demonstration and the 23 centers that participated in the first year only.

Findings After the First Year of Implementation

The first report for this study examined the impacts of the enhanced after-school programs in their first year of implementation in 50 after school-centers (Black and others, 2008). Impacts on student achievement were presented, as well as impacts on three academic behaviors (student attentiveness in class, student disruptiveness in class, and homework completion). The report also examined whether impacts differed across grades and for subgroups of students with different levels of prior academic achievement. Findings from the first report are summarized below.

Early Findings for Math

In the first year of the demonstration, the enhanced math program was implemented in 25 after-school centers. The program was implemented as intended, in terms of staff characteristics, training, and usage of instructional materials. Students in the enhanced program were offered an average of 179 minutes of math instruction per week, and 84 percent of the instructors reported that maintaining the intended pace of the daily lesson in the allotted time was not consistently a problem for them. The following key findings were reported:

- Students in the enhanced program group received 49 more hours of academic instruction in math during the school year than students in the regular program group (p-value = 0.00). This represents an estimated 30 percent increase in the hours of math instruction that students received over the school year.

- The enhanced program produced a positive and statistically significant impact of 2.8 scaled score points on the SAT 10 total math test (p-value = 0.01). This

impact translates into an effect size of a 0.06 standard deviation,[1] and represents an 8.5 percent improvement in students' test score growth, over and above what they would have experienced had they not enrolled in the enhanced program.

- The math program did not produce statistically significant impacts (either positive or negative) on any of the three school-day academic behavior measures: homework completion, attentiveness, and disruptiveness in class.

Impacts were also estimated for subgroups of students defined by grade (grades four and five vs. grades two and three) and for subgroups of students with different levels of math achievement at baseline (below basic, basic, and proficient). While impacts on SAT 10 scores were found to be positive and statistically significant for some student subgroups (i.e., students in grades four and five and students with basic proficiency), the difference in impacts across subgroups is not statistically significant. Thus, it could not be concluded that the enhanced program was more effective for some subgroups of students than others.

Early Findings for Reading

In the first year of the demonstration, the enhanced reading program was implemented in 25 after-school centers. The reading program was implemented as intended, in terms of staff characteristics, training, and usage of instructional materials. Students were offered an average of 176 minutes of reading instruction per week, though nearly half of the instructors (42 percent) reported that it was consistently difficult for them to include all aspects of the reading program and maintain the intended pace of the daily lesson plan. The following key findings were reported:

- Students in the enhanced program group received 48 more hours of academic instruction in reading during the school year than students in the regular program group (p-value = 0.00). This represents an estimated 20 percent increase in the hours of reading instruction that students received over the school year.

- The enhanced reading program did not have a statistically significant impact on reading achievement as measured by the SAT 10 reading test.[2]

[1]Effect sizes are 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).

[2]Among students in grades two to three, the enhanced program did have a positive and statistically significant impact on one of the two DIBELS fluency subtests (0.12 standard deviation, p = 0.03). However, this sub-

(continued)

- 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.

Additional analyses indicated that impacts on SAT 10 scores and academic behaviors were not statistically significant for any of the student subgroups defined by grade and reading achievement at baseline.

Differences Between After-School Centers that Participated in Both Years of the Study and Centers that Participated Only in the First Year

Near the end of the first year of implementation, the evaluation was extended to include an additional year of program operations and data collection. Continuation in the study was voluntary, and 27 of the original 50 after-school centers were able to participate in the study for another year (15 math centers and 12 reading centers). Although sites were recruited for a second year of participation in the study after they had experienced the program for the majority of the first implementation year, the first report for this study was published in the middle of the second implementation year.[3] Therefore, after-school centers' decisions about whether or not to continue in the second year could not have been based on knowledge about the effectiveness of the enhanced program in their center.

This section examines differences between after-school centers that participated in both years of the study (and which are the focus of this report) and those that did not continue for a second year. In particular, this section examines whether the continuing after-school centers had different impacts in the first year of the study than the non-continuing centers, in order to determine whether the continuing centers differed from the original set of centers in terms of their impacts. Where differences in impacts between the two subgroups of centers are found, differences in centers' first-year implementation characteristics are also examined.

Centers that Implemented the Enhanced Math Program

Appendix Table A.1 shows that the first-year impact of the enhanced math program in the 15 math centers that continued into the second year of the study did not differ significantly from program impacts in the 10 centers that did not. Estimated impacts on the total math score,

---

test is one of six reading measures used to estimate impacts for second- and third-grade students. When accounting for multiple hypothesis testing, this estimate is no longer statistically significant (see Benjamini and Hochberg, 1995).

[3]Recruitment of sites for a second year of the study began in the spring of 2006.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table A.1

Impact of the Enhanced Math Program on Student Achievement in the First Year of the Study, by Whether or Not a Center Participated in the Second Year of the Study

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Centers that continued in the second year of the study (n = 15)[a]					
SAT 10 math total scaled scores	607.01	603.52	3.49 *	0.09	0.02
Problem solving	607.88	605.38	2.50	0.06	0.11
Procedures	607.63	601.80	5.82 *	0.11	0.00
Sample size (total = 1,144)	634	510			
Centers that did not continue in the second year of the study (n = 10)					
SAT 10 math total scaled scores	602.38	600.02	2.36	0.06	0.15
Problem solving	603.71	601.17	2.54	0.06	0.14
Procedures	602.01	599.10	2.91	0.06	0.22
Sample size (total = 817)	447	370			

Difference in impacts between centers that continued and centers that did not continue in the second year	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
Centers that continued minus centers that did not continue			
SAT 10 math total scaled scores	1.13	0.03	0.60
Problem solving	-0.04	0.00	0.99
Procedures	2.91	0.06	0.35

(continued)

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

NOTES: The 25 after-school centers that participated in the first year of the study are divided into two groups: those that continued in the second year and those that did not continue.

Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: 389 to 796, 414 to 776, and 413 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

Appendix Table A.1 (continued)

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 effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

[a]Students enrolled in these centers in the first year of the study are the Cohort 1 sample.

for example, were 3.49 scale score points in the continuing centers and 2.36 scale score points in the centers that did not continue and the difference is not statistically significant. Estimated impacts on subscales also did not vary by a statistically significant amount between the two groups of math centers.[4] Thus, the 15 math centers that continued into the second year of the study do not systematically differ (in terms of their impacts) from the 10 centers that did not continue after the first year.

Centers that Implemented the Enhanced Reading Program

Appendix Table A.2 shows that the first-year impacts of the enhanced reading program in the 12 reading centers that continued into the second year of the study and in the 13 reading centers that did not continue. These results show that the impacts of the enhanced program on SAT 10 total reading scores and reading comprehension scores in the continuing centers were lower by a statistically significant amount than those in the non-continuing centers. Estimated impacts on the total reading score were -2.59 scale score points in the continuing centers and 1.48 scale score points in centers that did not continue, and the p-value of the difference was 0.03. Estimated impacts on the reading comprehension score were -3.55 scale score points in

[4]Estimated impacts for the problem-solving subscale were 2.50 scale score points in centers that continued and 2.54 in centers that did not, and the p-value of the difference was 0.99. Estimated impacts for the procedures subscale were 5.82 scale score points in the centers that continued and 2.91 in the centers that did not, and the p-value was 0.35.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table A.2

Impact of the Enhanced Reading Program on Student Achievement in the First Year of the Study, by Whether or Not a Center Participated in the Second Year of the Study

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Centers that continued in the second year of the study (n = 12)[a]					
SAT 10 reading total scaled scores	588.66	591.25	-2.59	-0.08	0.06
Vocabulary	582.73	584.84	-2.12	-0.05	0.29
Reading comprehension	589.47	593.01	-3.55 *	-0.10	0.04
Word study skills (grades 2-4)[b]	589.44	590.04	-0.61	-0.01	0.81
DIBELS (grades 2-3)[c]					
Oral fluency score	73.61	71.93	1.68	0.05	0.44
Nonsense word fluency score	66.19	63.82	2.37	0.07	0.32
Sample size (total = 905)	504	401			
Centers that did not continue in the second year of the study (n = 13)					
SAT 10 reading total scaled scores	586.28	584.80	1.48	0.04	0.24
Vocabulary	579.29	576.43	2.85	0.06	0.12
Reading comprehension	588.02	585.64	2.38	0.07	0.15
Word study skills (grades 2-4)[b]	583.53	586.60	-3.07	-0.07	0.16
DIBELS (grades 2-3)[c]					
Oral fluency score	68.15	65.38	2.77	0.08	0.17
Nonsense word fluency score	62.54	58.06	4.48	0.12	0.10
Sample size (total = 923)	544	379			

Difference in impacts between centers that continued and centers that did not continue in the second year	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
Centers that continued minus centers that did not continue			
SAT 10 reading total scaled scores	-4.06 *	-0.12	0.03
Vocabulary	-4.97	-0.11	0.07
Reading comprehension	-5.92 *	-0.16	0.01
Word study skills (grades 2-4)[b]	2.46	0.06	0.46

(continued)

Appendix Table A.2 (continued)

Difference in impacts between centers that continued and centers that did not continue in the second year	Difference in Impacts	Difference in Impact Effect Sizes	P-Value for Difference
Centers that continued minus centers that did not continue			
DIBELS (grades 2-3)[c]			
Oral fluency score	-1.09	-0.03	0.71
Nonsense word fluency score	-2.11	-0.06	0.56

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: The 25 after-school centers that participated in the first year of the study are divided into two groups: those that continued in the second year and those that did not continue.

Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 374 to 787, 439 to 777, 412 to 739, and 410 to 740. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximum 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 effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98; nonsense = 36.13. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[a]Students enrolled in these centers in the first year of the study are the Cohort 1 sample.

[b]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.

[c]The DIBELS sample includes only second- and third-grade students because the nonsense word fluency subtest and the oral fluency subtest were not administered to fourth- and fifth-grade students in both study years.

the continuing centers and 2.38 scale score points in centers that did not continue, and the p-value of the difference was 0.01. Estimated impacts on the other three subscales did not vary by a statistically significant amount.[5]

Given that differences in impacts exist between the continuing and non-continuing centers, the first-year implementation characteristics were also examined. This analysis found that continuing and non-continuing centers differed by a statistically significant amount with respect to three features of their first-year operations. Relative to the centers that did not continue in the second year, continuing centers:

- Employed a lower percent of certified teachers (76 percent versus 93 percent, p-value = 0.01),
- Had a higher student-to-staff ratio on average (12 versus 10, p-value = 0.01), and
- Had staff who were less likely to report having been given more paid time to prepare (p-value = 0.01) than their regular program counterparts.

[5]Estimated impacts for the vocabulary subscale were -2.12 scale score points in centers that continued and 2.85 in centers that did not, and the p-value of the difference was 0.07. Estimated impacts for the word study skills were -.61 scale score points in the centers that continued and -3.07 in the centers that did not, and the p-value was.46. Estimated impacts for the oral fluency subscale were 1.68 points in centers that continued and 2.77 in centers that did not, and the p-value of the difference was 0.71. Estimated impacts for the nonsense word fluency score were 2.37 points in the centers that continued and 4.48 in the centers that did not, and the p-value was 0.56.

Appendix B

Statistical Precision and Minimum Detectable Effect Size

This appendix presents the minimum detectable effect size (MDES) for the analyses in this report. 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.

Appendix Table B.1 presents the MDES for each of the impact analyses in this report (calculations are for impacts on SAT 10 achievement test scores). The table also shows the size of the corresponding sample used for analysis, which is a key factor in the determination of the MDES (the larger the sample size, the smaller the MDES). Note that these MDES calculations are based on the actual parameter values related to the standard error of the impact estimates. Details on the MDES calculations are presented at the end of the appendix.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table B.1

Sample Sizes and Minimum Detectable Effect Sizes for Math and Reading Analysis Samples

	Cohort 1 Sample	Cohort 2 Sample	Two-Year Sample
Math			
Sample size	1,144	792	367
Minimum detectable effect size	0.10	0.15	0.21
Reading			
Sample size	905	626	270
Minimum detectable effect size	0.11	0.14	0.23

NOTE: Calculations are based on the formula discussed in Appendix B and the parameters values in Appendix Table B.2.

[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.

MDES for the Impact of Assigning Students to the Enhanced Program for One Year

Table B.1 presents the impact of assigning students to enroll in the enhanced program *for one school year*. As shown, the smallest program impact that can be estimated with confidence (given random sampling and estimation error in the sample) for the Cohort 1 and Cohort 2 samples is 0.10 and 0.15 standard deviation, respectively, for the math analysis, and 0.11 and 0.14 standard deviation for the reading analysis. Notice that the MDES for the Cohort 1 sample is smaller because it includes a larger number of students (see Chapter 2 for a more detailed explanation of why the two cohort-specific samples differ in size).

MDES for the Impact of Assigning Students to the Enhanced Program for Two Years

The impact of assigning students to enroll in the enhanced program for two school years is based on students who were randomly assigned to enroll in either the enhanced or regular program for two consecutive years. The smallest program impact that can be estimated with confidence (given random sampling and estimation error in the sample) is 0.21 standard deviation for the math analysis and 0.23 standard deviation for the reading analysis.

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 includes covariates for each block and 11 additional covariates.

σ_y^2 = The within-block variance of the outcome in question[2]

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.

P = The proportion of students randomly assigned to the enhanced program group

N = The number of independent observations (students) in the sample

J = The number of random assignment blocks in the study[3]

τ_y^2 = The cross-block variance in the mean value of the outcome measure *y*.

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.[4]

The values of these parameters were estimated based on the analysis samples, using SAT 10 total scores as the outcome. These values are presented in Appendix Table B.2 and were used in the MDES calculations.[5]

[2]All between-block variation is explained by the block fixed-effects included in the impacts model (see Appendix G and H), so within-block variation is the only unexplained variation in the analysis.

[3]For the impact of being assigned to the enhanced program for *one school year*, random assignment blocks are defined by grade *j* and center *c* in fall 2005 for the Cohort 1 sample, and by first-year status (regular program, or new to the study) and grade *j* and center *c* in fall 2006 for the Cohort 2 sample.

For the impact of being assigned to the enhanced program for *two school years*, random assignment blocks are defined by grade *j* and center *c* (in fall of 2005).

[4]This assumption is justified by the fact that the sites for the study were selected purposefully. Therefore, the results are not generalizable statistically to a larger universe of after-school programs other than the centers included in this particular study.

[5] The second component in the MDES formula (square root portion) represents the standard error of the impact estimate. This standard error is known, so the MDES in Table B.1 could also have been calculated directly as follows: $MDES = M_{N-J-12} * s.e.(impact\_estimate)$.

Appendix Table B.2

Parameter Values Used to Calculate the Minimum Detectable Effect Size for Math and Reading Analysis Samples

	Cohort 1 Sample	Cohort 2 Sample	Two-Year Sample
Math			
R^2	0.43	0.39	0.36
$\sigma^2/(\tau^2+\sigma^2)$	0.66	0.68	0.72
P	0.55	0.58	0.63
J	59	103	45
N (independent observations)	1,144	792	367
Reading			
R^2	0.51	0.47	0.47
$\sigma^2/(\tau^2+\sigma^2)$	0.72	0.69	0.78
P	0.56	0.56	0.61
J	47	83	33
N (independent observations)	905	626	270

NOTE: Parameters are estimated based on the relevant sample, using the SAT 10 total score in math or reading as the outcome of interest.

Appendix C

Creation of the Analysis Sample (Math Centers)

This appendix describes the creation of the analysis samples used in the math impact analyses presented in Chapters 4 and 5. The appendix is divided into two sections — the first section describes the creation of the Cohort 1 and Cohort 2 samples (for estimating the impact of offering students one year of the enhanced after-school program *in the first or second year of implementation*), while the second section describes the creation of the two-year sample (for estimating the impact of offering students the opportunity to enroll in the enhanced math program *for two consecutive school years*).

Each section begins by providing information on the characteristics of students in the full study sample (i.e., all students who were randomized to answer a given research question). The purpose of this exercise is to determine whether random assignment resulted in two statistically equivalent groups of students at baseline (enhanced vs. regular program group).

Each section then presents response rates in the full study sample for each follow-up data source used in the analysis and describes how the analysis sample was constructed based on available follow-up data

Each section then ends with an examination of the characteristics of students in the resulting analysis sample. The key questions underlying this part of the response analysis are (1) whether students in the analysis sample are representative of students in the full study sample (which affects the *generalizability* of the findings to the full study sample), and (2) whether the analysis sample preserves the random assignment design (which affects whether or not the impact estimates are *unbiased*). Both of the issues are examined in the appendix.

One-Year Sample (Offer of One Year of Service)

As explained in Chapter 2, two cohorts of students were randomly assigned to enroll in the enhanced after-school math program for one school year (enhanced program group) or to remain in the regular after-school program during that time (regular program group). Students who were randomly assigned in the first year of implementation are referred to as the Cohort 1 sample. Students who were randomly assigned in the second year of implementation — and who did not participate in the enhanced program in Year 1 — are referred to as the Cohort 2 sample (see Figure 2.2). The analyses presented in this appendix are based on data from both of these samples.

Characteristics of Students in the Full Study Sample

The Cohort 1 full study sample includes 1,218 identified low-performing students who applied to the study and were randomly assigned to either the enhanced after-school math program or the regular program group. The Cohort 2 full study sample includes 833 newly

identified low-performing students applicants and students from Cohort 1 who applied to the second year of the study, all of whom were randomly assigned to either the enhanced after-school math program or the regular program group.

Appendix Table C.1 presents the baseline characteristics of students in these full study samples for each research group (enhanced program group and regular program group). An overall F-test indicates that there is no systematic difference in the baseline characteristics of students in the enhanced and regular program groups in the full Cohort 1 or Cohort 2 study samples. This indicates that random assignment was successful in creating two equivalent research groups at baseline.

Response Rates

Appendix Table C.2 presents response rates for each data source on follow-up student outcomes, both overall and by program group.[1] Response rates are presented for each of the two cohort-specific samples. The first two rows of each panel show the response rates for the key data source used in the impact analysis — the follow-up SAT 10 total score. The last three rows in each panel report the response rates for the other data sources used in the analysis — the regular-school-day teacher questionnaire (used to measure student academic behavior), the student survey (used to measure the service contrast), and the follow-up state test score (used as a supplementary measure of students' academic performance).[2] As seen in this table, with the exception of the state assessment in Cohort 1, response rates for all data sources are above 93 percent.[3] Moreover, the response rates of students in the enhanced and regular program group do not differ by a statistically significant amount on any measure.

Constructing the Analysis Sample

To keep the sample of students consistent across the outcome measures, the analysis sample is limited to students with data on both the follow-up SAT 10 assessment and the regular-school-day teacher survey.[4] The consort chart in Figure C.1 describes the construction of the samples used for analysis. As shown, in Cohort 1, 74 students are excluded from the math analysis and, in Cohort 2, 41 students are excluded. Thus, the samples used for analysis consist

[1]Spring 2006 for Cohort 1 and Spring 2007 for Cohort 2.

[2]Second-grade students are not included in the impact analysis for state tests because a subset of the 15 after-school centers do not administer a local assessment to their second-grade students (nine centers in the first year and seven centers in the second year). Response rates for school records are therefore based on students in grades three through five in the full study sample.

[3]The response rates for Cohort 1 state tests are all above 82 percent.

[4]The state test data are also used for an outcome measure; however, these data are not used when creating the analysis sample because state test data are not available for second-grade students.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table C.1

Baseline Characteristics of Students in the Math Full Study Sample (One Year of Service)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Cohort 1[a]						
Enrollment						
2nd grade	302	167	135			
3rd grade	314	170	144			
4th grade	308	170	138			
5th grade	294	167	127			
Total	1,218	674	544			
Race/ethnicity (%)						
Hispanic		29.27	24.32	4.94 *	0.10	0.03
Black, non-Hispanic		37.72	42.26	-4.54 *	-0.08	0.04
White, non-Hispanic		25.71	25.94	-0.24	0.00	0.91
Asian		1.49	2.16	-0.67	-0.05	0.39
Other		5.79	5.26	0.53	0.02	0.69
Gender (%)						
Male		47.63	43.70	3.92	0.07	0.17
Average age (years)		8.64	8.65	-0.01	-0.01	0.82
Overage for grade[b] (%)		17.66	17.18	0.48	0.01	0.82
Free/reduced-price lunch (%)						
Eligible (among information providers)		77.66	75.99	1.67	0.04	0.40
No information provided		3.12	1.73	1.39	0.08	0.09
Average household size		1.95	1.88	0.08	0.08	0.16
Single-adult household (%)		33.15	34.87	-1.72	-0.03	0.52
Mother's education level (%)						
Did not finish high school		16.91	15.92	0.99	0.02	0.65
High school diploma or GED certificate		31.31	31.66	-0.35	-0.01	0.90
Some postsecondary study		44.07	45.36	-1.29	-0.02	0.64
No information provided		7.72	7.06	0.66	0.03	0.64
SAT 10 baseline math total scaled scores		567.15	565.41	1.75	0.05	0.31
Problem solving		573.03	570.72	2.31	0.06	0.21
Procedures		560.03	558.78	1.25	0.02	0.57
Sample size (total = 1,218)		674	544			

(continued)

Appendix Table C.1 (continued)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Cohort 2[c]						
Enrollment						
2nd grade	272	161	111			
3rd grade	193	111	82			
4th grade	184	105	79			
5th grade	184	108	76			
Total	833	485	348			
Race/ethnicity (%)						
Hispanic		29.20	26.24	2.96	0.06	0.31
Black, non-Hispanic		37.85	39.28	-1.43	-0.03	0.58
White, non-Hispanic		24.39	28.05	-3.66	-0.08	0.17
Asian		1.55	1.35	0.20	0.01	0.83
Other		6.89	5.02	1.88	0.08	0.25
Gender (%)						
Male		41.69	46.50	-4.81	-0.09	0.18
Average age (years)		8.66	8.66	0.00	0.01	0.93
Overage for grade[b] (%)		14.24	16.19	-1.96	-0.05	0.45
Free/reduced-price lunch (%)						
Eligible (among information providers)		76.92	75.78	1.15	0.03	0.67
No information provided		3.23	3.05	0.18	0.01	0.88
Average household size		1.97	1.88	0.09	0.09	0.24
Single-adult household (%)		29.91	35.55	-5.63	-0.11	0.08
Mother's education level (%)						
Did not finish high school		18.27	16.08	2.19	0.05	0.42
High school diploma or GED certificate		31.64	29.70	1.94	0.04	0.56
Some postsecondary study		45.29	49.39	-4.10	-0.07	0.25
No information provided		4.80	4.83	-0.03	0.00	0.98
SAT 10 baseline math total scaled scores		570.96	571.05	-0.09	0.00	0.97
Problem solving		577.62	577.41	0.21	0.01	0.93
Procedures		562.91	562.43	0.48	0.01	0.87
Sample size (total = 833)		485	348			

(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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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 average

Appendix Table C.1 (continued)

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 average observed mean for members randomly assigned to the enhanced program group. The regular program group values in the next column are the average 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined.

F-tests were calculated for these full study samples 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.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

[c]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

of 1,144 students in Cohort 1 (which represents 94 percent of the full study sample) and 792 students in Cohort 2 (which represents 95 percent of the full study sample).[5]

Characteristics of Students in the Analysis Sample

This section examines whether (1) the analysis sample differs from the full study sample (which affects the generalizability of the findings to the full study sample), and whether (2) the statistical equivalence of the two research groups is preserved in the analysis sample (which affects whether the impact estimates are unbiased).

To examine the first issue, statistical tests were conducted to determine whether students in the analysis sample are different from students in the full study sample who were *excluded* from the analysis due to missing follow-up data. An overall F-test indicates that, for Cohort 1, these two groups of students are systematically different in terms of their background characteristics (F = 1.88, p-value = 0.01). Students were more likely to be included in the

[5]In Cohort 1, 94 percent of students in both the enhanced and regular program groups had follow-up data on both the SAT 10 and the teacher survey. The difference in response rates between the two program groups is not statistically significant (p-value = 0.81). In Cohort 2, 95 percent of students in both groups had follow-up data on both of these data sources, and the difference in response rates between the two program groups is not statistically significant (p-value = 0.97).

The Evaluation of Academic Instruction in After-School Programs

Appendix Table C.2

Response Rates to Tests and Surveys for Students in the Math Study Sample (One Year of Service)

Data Source	Full Sample	Enhanced Program	Regular Program
Cohort 1[a]			
Key outcome measure			
Follow-up SAT 10[b] (%)	94.01	94.21	93.75
Additional outcome measures			
Regular-school-day teacher survey (%)	99.51	99.26	99.82
Student survey (%)	98.19	98.37	97.98
Follow-up state test score[c] (%)	85.04	86.98	82.64
Full study sample size (total = 1,218)		674	544
Cohort 2[d]			
Key outcome measure			
Follow-up SAT 10[b] (%)	95.68	95.46	95.98
Additional outcome measures			
Regular-school-day teacher survey (%)	99.40	99.59	99.14
Student survey (%)	97.60	97.73	97.41
Follow-up state test score[c] (%)	93.94	93.83	94.09
Full study sample size (total = 833)		485	348

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

NOTES: Response rates are calculated from the full study sample for all students in the study and separately for students in each program group. The difference between the enhanced and regular program group response rates is not significant for any of the measures.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

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

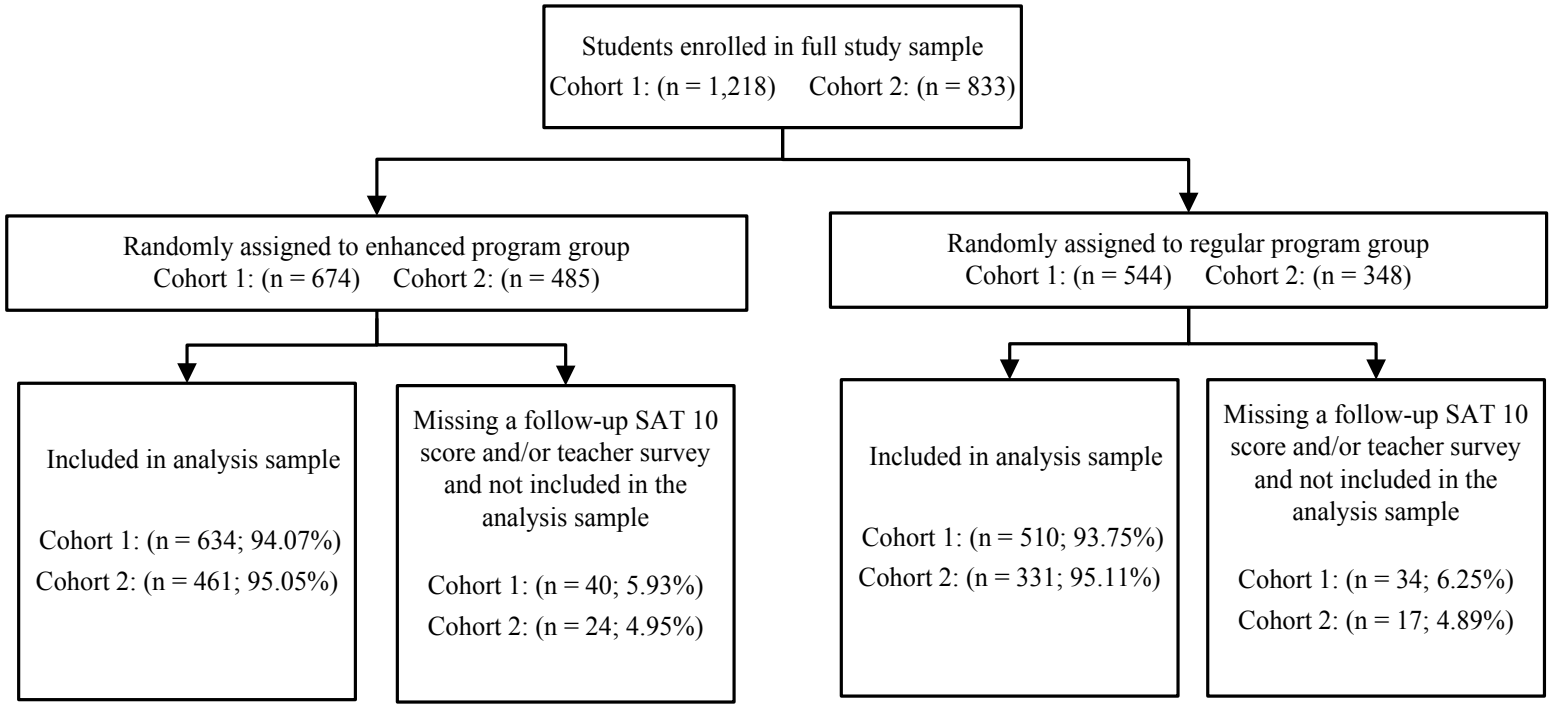
[c]This calculation is based on students in grades three to five only. Second-grade students are excluded from the analysis of state test data because most sites do not test their students until third grade.

[d]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

The Evaluation of Academic Instruction in After-School Programs

Appendix Figure C.1

Flow of Students from Enrollment to Analysis in the Math Sample (One Year of Service)



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 full study sample. All percentages are based on the number of students randomly assigned to either the enhanced or the regular program group.

analysis sample if their families had not moved in the two years prior to the start of the study (p-value = 0.00) or if there were more adults in their household (p-value = 0.00). For Cohort 2, the overall F-test indicates that these two groups of students are not systematically different in terms of their background characteristics (F = 0.75, p-value = 0.77). Because the analysis samples include almost all students (94 percent in Cohort 1 and 95 percent in Cohort 2) in the full study samples, and because the samples used for analysis still reflect the general characteristics of the full study samples, it is reasonable to assume that the impact findings presented in this report can be generalized to all students in the full study sample.

To examine whether randomization is preserved in the analysis sample, the characteristics of students in the enhanced and regular program were also compared. The characteristics of students in the samples used for analysis are presented in Table 4.1. As discussed in Chapter 4, an overall F-test indicates that there is no systematic difference in the background characteristics of students in the enhanced and regular program groups, in either of the two cohort-specific samples. This indicates that the statistical equivalence of the two research groups is preserved in the analysis sample.

Two-Year Sample (Offer of Two Years of Service)

As explained in Chapter 2, the impact of offering students the opportunity to enroll in the enhanced program for two consecutive years is estimated by comparing the outcomes of students who were randomly assigned to either the enhanced after-school program (enhanced program group) or the regular after-school program (regular program group) for two consecutive school years. These students are referred to as the two-year sample (see Figure 2.3). Also recall that this sample includes students assigned to two years of the enhanced program, whether or not they attended both years (i.e., the intent-to-treat sample).

Characteristics of Students in the Full Study Sample

The full two-year study sample includes 470 students, 62 percent of whom (293 students) were randomly assigned to the enhanced after-school program in both years of the study and 38 percent of whom (177 students) were assigned to the regular after-school program group in both years.

Appendix Table C.3 presents the baseline characteristics of students in the full two-year study sample for each research group (enhanced program group and regular program group). An overall F-test indicates that there is no systematic difference in the characteristics of students in the enhanced and regular program groups at baseline in the full two-year study sample. This indicates that random assignment was successful in creating two equivalent research groups at baseline.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table C.3

Baseline Characteristics of Students in the Math Full Study Sample (Offer of Two Years of Service)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Enrollment						
2nd grade	157	99	58			
3rd grade	166	98	68			
4th grade	147	96	51			
Total	470	293	177			
Race/ethnicity (%)						
Hispanic		28.72	31.19	-2.48	-0.05	0.60
Black, non-Hispanic		36.85	41.17	-4.33	-0.08	0.33
White, non-Hispanic		27.06	21.11	5.94	0.12	0.14
Other		4.79	7.33	-2.55	-0.10	0.37
Gender (%)						
Male		49.67	45.59	4.08	0.07	0.47
Average age (years)		8.06	8.11	-0.05	-0.09	0.39
Overage for grade[a] (%)		15.81	16.93	-1.12	-0.03	0.78
Free/reduced-price lunch (%)						
Eligible (among information providers)		75.04	79.41	-4.37	-0.10	0.25
No information provided		4.57	2.01	2.56	0.15	0.19
Average household size		1.95	1.88	0.07	0.07	0.50
Single-adult household (%)		33.76	41.87	-8.11	-0.15	0.13
Mother's education level (%)						
Did not finish high school		19.15	16.89	2.25	0.06	0.62
High school diploma or GED certificate		26.19	33.17	-6.97	-0.14	0.17
Some postsecondary study		46.20	40.17	6.03	0.11	0.26
No information provided		8.47	9.77	-1.30	-0.05	0.59
SAT 10 baseline math total scaled scores		553.03	547.77	5.25	0.14	0.10
Problem solving		560.54	553.24	7.30 *	0.18	0.03
Procedures		543.66	539.67	3.99	0.08	0.33
Sample size (total = 470)		293	177			

(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.

Appendix Table C.3 (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 enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 effect size for each characteristic is calculated as a proportion of the standard deviation for students in the two-year sample regular program group.

An F-test was calculated 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-value is not significant.

[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.

Response Rates

Appendix Table C.4 presents response rates for each data source on follow-up student outcomes, both overall and by program group, in the full two-year study sample.[6] The first row presents response rates for the key source of data used in the impact analysis — the follow-up SAT 10 total score. The last three rows report the response rates for the other data sources used in the analysis — the regular-school-day teacher questionnaire (used to measure student academic behavior), the student survey (used to measure the service contrast) and the follow-up state test score (used as a supplementary measure of students' academic performance).

As seen in this table, response rates for all data sources are above 78 percent, and the response rates of students in the enhanced and regular program group do not differ by a statistically significant amount on any measure. More specifically, 78 percent of students in the enhanced program group had follow-up data on both the SAT 10 and the teacher survey, while 79 percent of students in the regular program group had data on both of these sources. The difference in

[6]Outcomes are measured in Spring 2007.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table C.4

Response Rates to Tests and Surveys for Students in the Math Study Sample (Offer of Two Years of Service)

Data Source	Full Sample	Enhanced Program	Regular Program
Key outcome measures			
Follow-up SAT 10[a] (%)	78.51	77.82	79.66
Additional outcome measures			
Regular-school-day teacher survey (%)	81.28	81.57	80.79
Student survey (%)	80.21	80.55	79.66
Follow-up state test score[b] (%)	78.30	78.16	78.53
Full study sample size (total = 470)		293	177

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

NOTES: Response rates are calculated from the full study sample for all students in the study and separately for students in each program group. The difference between the enhanced and regular program group response rates is not significant for any of the measures.

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

[b]This calculation is based on students in grades three to five only. Second-grade students are excluded from the analysis of state test data because most sites do not test their students until third grade.

response rates between the two program groups is not statistically significant (p-value = 0.68). Recall that all eligible students from the fall of 2005 were randomly assigned in the second year, whether or not they reapplied for the second year of the study. And for those who did not reapply, the study team tried to collect follow-up data. Response rates in the full two-year study sample are thus driven downwards by the fact that 43 percent (84 students) of students who did not reapply to the study the second year also did not consent to follow-up data collection.[7] That said, among students who *did* consent to follow-up data collection (all students who participated in the second year and 57 percent of students who did not reapply to the study the second year), response rates are above 95 percent on all four of the data sources in Table C.4.

[7]Forty-two percent of students in the enhanced program group who did not reapply also did not consent to follow-up data collection, while 46 percent of students in the regular program group who did not reapply did not consent. The difference in non-consent rates between program groups does not differ by a statistically significant amount (p-value = 0.56).

Constructing the Analysis Sample

As noted earlier, the analysis sample is limited to students with data on both the follow-up SAT 10 assessment and the regular-school-day teacher survey. The consort chart in Figure C.2 describes the construction of the two-year sample used for analysis. As shown, 103 students are excluded from the math analysis sample (84 of these students are students who did not apply in the second year and did not provide consent for follow-up data collection). Thus, the two-year sample used for analysis consists of 367 students (which represents 78 percent of the full two-year study sample).

Characteristics of Students in the Analysis Sample

Statistical tests were conducted to determine whether students in the two-year sample used for analysis are different at baseline from students in the full two-year study sample who were *excluded* from the analysis due to missing follow-up data. An overall F-test indicates that these two groups of students are systematically different in terms of their background characteristics ($F = 2.9$, p-value = 0.00). Students were more likely to be included in the analysis sample if their families had not moved in the two years prior to the start of the study (p-value = 0.00) or if they were Hispanic (p-value = 0.00).[8]

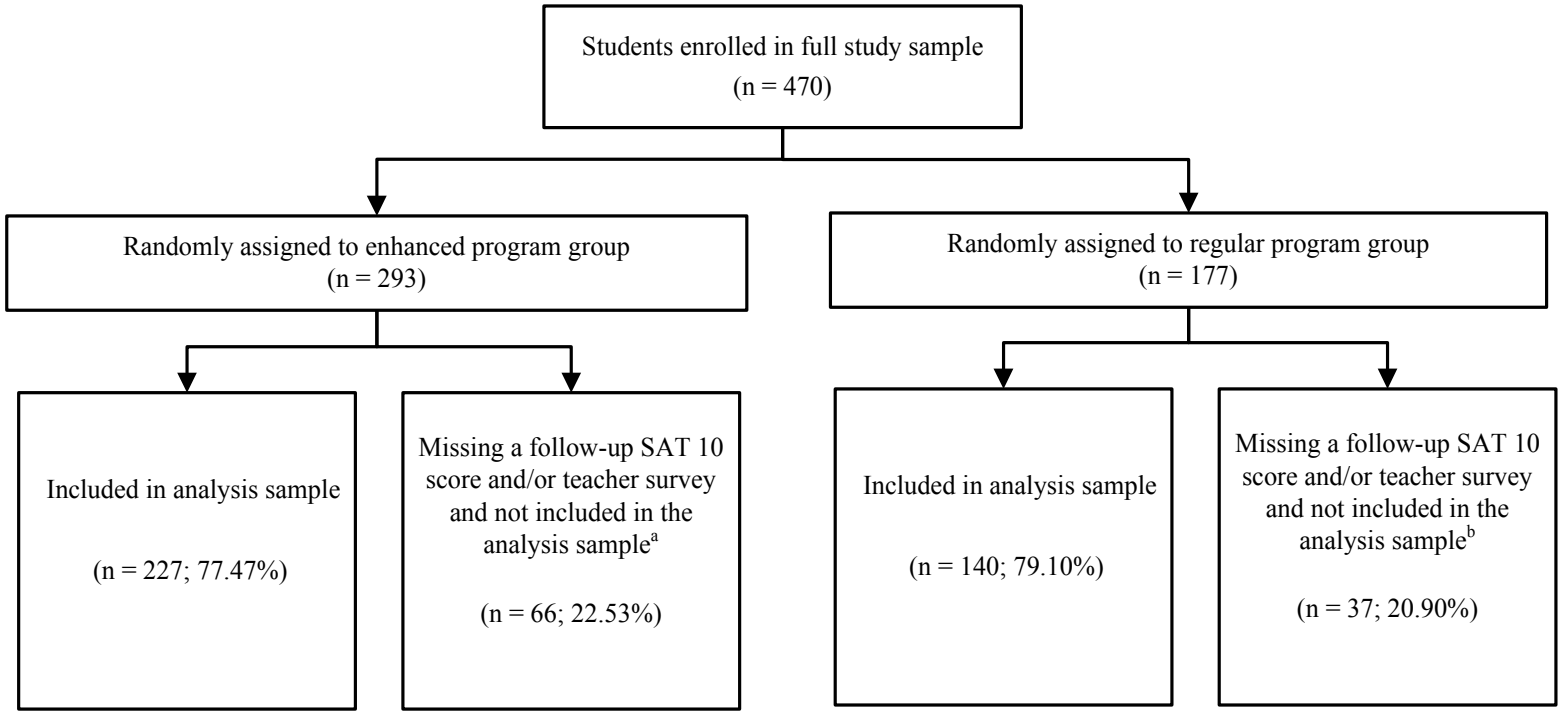
To examine whether the randomization is preserved in the analysis sample, the characteristics of students in the enhanced and regular program groups were also compared (see Table 5.1). As discussed in Chapter 5, an overall F-test indicates that there is no systematic difference in the background characteristics of students in the enhanced and regular program groups. This indicates that the statistical equivalence of the two research groups is preserved in the analysis sample. Thus, while the analysis sample may not be representative of the full study sample, there is no bias between the enhanced and regular program groups.

[8]As noted earlier in this appendix, students excluded from the two-year analysis sample are primarily students who did not apply in the second year (nonapplicants) and did not consent to a second year of data collection. As will be explained in Appendix H, nonapplicants who *did* consent to follow-up data collection are weighted to account for nonapplicants students who did *not* consent to follow-up data collection.

The Evaluation of Academic Instruction in After-School Programs

Appendix Figure C.2

Flow of Students from Enrollment to Analysis in the Math Sample (Offer of Two Years of Service)



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

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

[a]Among these, 51 are students who did not reapply and did not provide consent for follow-up data collection.

[b]Among these, 33 are students who did not reapply and did not provide consent for follow-up data collection.

Appendix D

Creation of the Analysis Sample (Reading Centers)

This appendix describes the creation of the analysis samples used in the reading impact analyses. The appendix is divided into two sections — the first section describes the creation of the Cohort 1 and Cohort 2 samples (for estimating the impact of offering students one year of the enhanced after-school program *in the first or second year of implementation*), while the second section describes the creation of the two-year sample (for estimating the impact of offering students the opportunity to enroll in the enhanced reading program *for two consecutive school years*).

Each section begins by providing information on the characteristics of students in the full study sample (i.e., all students who were randomized to answer a given research question). The purpose of this exercise is to determine whether random assignment resulted in two statistically equivalent groups of students at baseline (enhanced vs. regular program group).

Each section then presents response rates in the full study sample for each follow-up data source used in the analysis and describes how the analysis sample was constructed based on available follow-up data.

Each section ends with an examination of the characteristics of students in the resulting analysis sample. The key questions underlying this part of the response analysis are (1) whether students in the analysis sample are representative of students in the full study sample (which affects the *generalizability* of the findings to the full study sample), and (2) whether the analysis sample preserves the random assignment design (which affects whether or not the impact estimates are *unbiased*). Both of the issues are examined in the appendix.

One-Year Sample (Offer of One Year of Service)

As explained in Chapter 2, two cohorts of students were randomly assigned to enroll in the enhanced after-school reading program for one school year (enhanced program group) or to remain in the regular after-school program during that time (regular program group). Students who were randomly assigned in the first year of the study comprise the Cohort 1 sample. Students who were randomly assigned in the second year of the study — and who did not participate in the enhanced program in the first year of the study — comprise the Cohort 2 sample (see Figure 2.2). The analyses presented in this appendix are based on data from both of these samples.

Characteristics of Students in the Full Study Sample

The full Cohort 1 study sample includes 989 identified low-performing students who applied and were randomly assigned to either the enhanced after-school reading program or the regular program group. The Cohort 2 full study sample includes 668 newly identified low-

performing student applicants and students from Cohort 1 who applied to the second year of the study, all of whom were randomly assigned to either the enhanced after-school reading program or the regular program group.

Appendix Table D.1 presents the baseline characteristics of students in these full study samples for each research group (enhanced program group and regular program group). An overall F-test indicates that there is a systematic difference in the background characteristics of students in the enhanced and regular program groups in both of the full study samples (Cohort 1: F-value = 2.94, p-value = 0.00; Cohort 2: F-value = 2.22, p-value = 0.00).[1] This means that, taken together, individual differences between the enhanced and regular program groups are greater than what would be predicted by chance. In the Cohort 1 sample, for example, there is a statistically significant difference between the enhanced and regular program groups in terms of baseline reading test scores (students in the enhanced group have lower baseline scores on average), while, in the Cohort 2 sample, there is a statistically significant difference in household composition (students in the enhanced program group are more likely to come from a single-adult household).[2] The former difference in baseline test scores is especially important because reading achievement is also a key outcome measure in this evaluation. (See Appendix G for a discussion of the analysis model that was used to control for observed differences in baseline characteristics between the enhanced and regular program groups, as well as the tests that were used to test the sensitivity of the impact findings to model and sample specifications).

Response Rates

Appendix Table D.2 presents response rates for each data source on follow-up student outcomes, both overall and by program group.[3] Response rates are presented for each of the two cohort-specific samples. The first three rows of each panel show the response rates for the three key sources of data used in the impact analysis — the follow-up SAT 10 total score and the DIBELS Oral Reading Fluency (ORF) and Nonsense Word Fluency (NWF) scores (administered to second- and third-graders in the sample). The last three rows in each panel report the response rates for the other three data sources used in analysis — the regular-school-day teacher questionnaire (used to measure student academic behavior), the student survey (used to measure the service contrast) and the follow-up state test score (used as a supplementary measure of

[1]Note that baseline differences between the enhanced and regular program group were also found in the first report for the 25 after-school reading centers that participated in the first year of the study (Black et al., 2008.). The Cohort 1 sample represents a subset of the students included in the sample for the first-year report.

[2]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.

[3]Spring 2006 for Cohort 1 and Spring 2007 for Cohort 2.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table D.1

Baseline Characteristics of Students in the Reading Full Study Sample (One Year of Service)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Cohort 1[a]						
Enrollment						
2nd grade	253	140	113			
3rd grade	241	137	104			
4th grade	252	139	113			
5th grade	243	134	109			
Total	989	550	439			
Race/ethnicity (%)						
Hispanic		37.40	40.18	-2.78	-0.05	0.21
Black, non-Hispanic		41.15	37.78	3.36	0.06	0.07
White, non-Hispanic		14.23	14.62	-0.39	-0.01	0.83
Asian		2.01	2.75	-0.74	-0.04	0.44
Other		5.11	4.56	0.55	0.03	0.67
Gender (%)						
Male		48.18	45.70	2.48	0.05	0.44
Average age (years)		8.60	8.56	0.04	0.07	0.24
Overage for grade[b] (%)		17.09	14.36	2.73	0.07	0.24
Free/reduced-price lunch (%)						
Eligible (among information providers)		83.58	82.44	1.14	0.03	0.56
No information provided		4.55	4.65	-0.11	0.00	0.94
Average household size		2.09	2.00	0.09	0.07	0.23
Single-adult household (%)		29.51	29.48	0.03	0.00	0.99
Mother's education level (%)						
Did not finish high school		25.45	19.00	6.45 *	0.14	0.02
High school diploma or GED certificate		32.55	26.53	6.02 *	0.12	0.04
Some postsecondary study		37.09	47.57	-10.48 *	-0.19	0.00
No information provided		4.91	6.90	-1.99	-0.07	0.20
SAT 10 baseline reading total scaled scores		565.88	570.55	-4.67 *	-0.14	0.01
Vocabulary/word reading[c]		556.98	563.65	-6.67 *	-0.15	0.01
Reading comprehension		566.76	572.69	-5.94 *	-0.16	0.01
Word study skills[d]		575.55	577.07	-1.52	-0.04	0.50
Sample size (total = 989)		550	439			

(continued)

Appendix Table D.1 (continued)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Cohort 2[e]						
Enrollment						
2nd grade	209	125	84			
3rd grade	160	85	75			
4th grade	144	81	63			
5th grade	155	88	67			
Total	668	379	289			
Race/ethnicity (%)						
Hispanic		38.27	40.15	-1.88	-0.03	0.50
Black, non-Hispanic		37.56	39.94	-2.39	-0.05	0.30
White, non-Hispanic		17.05	14.05	3.00	0.08	0.21
Asian		2.27	2.87	-0.59	-0.03	0.63
Other		4.94	3.09	1.85	0.09	0.22
Gender (%)						
Male		55.83	47.72	8.11 *	0.15	0.04
Average age (years)		8.56	8.56	0.00	0.00	0.98
Overage for grade[b] (%)		14.68	15.52	-0.84	-0.02	0.77
Free/reduced-price lunch (%)						
Eligible (among information providers)		82.24	83.94	-1.70	-0.04	0.50
No information provided		4.56	3.01	1.55	0.07	0.33
Average household size		1.97	2.19	-0.22 *	-0.19	0.01
Single-adult household (%)		34.01	23.71	10.29 *	0.22	0.00
Mother's education level (%)						
Did not finish high school		21.99	26.67	-4.67	-0.10	0.18
High school diploma or GED certificate		26.74	30.45	-3.70	-0.08	0.31
Some postsecondary study		45.83	37.92	7.91 *	0.15	0.05
No information provided		5.43	4.96	0.47	0.02	0.80
SAT 10 baseline reading total scaled scores		570.32	572.07	-1.75	-0.05	0.43
Vocabulary/word reading[c]		561.51	561.86	-0.36	-0.01	0.91
Reading comprehension		570.94	573.10	-2.16	-0.06	0.41
Word study skills[d]		579.27	580.73	-1.46	-0.04	0.59
Sample size (total = 668)		379	289			

(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.

Appendix Table D.1 (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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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 average observed mean for members randomly assigned to the enhanced program group. The regular program group values in the next column are the average 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined.

F-tests were calculated for these full study samples 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. The F-value for the Cohort 1 sample (F = 2.94) and the Cohort 2 sample (F = 2.22) are signficant at the 5 percent level.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

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

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

[e]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

students' academic performance).[4] As seen in this table, with the exception of the state assessment, response rates for all data sources are above 90 percent.[5] In addition, the response rates of students in the enhanced and regular program group do not differ by a statistically significant amount on any measure.

[4]Second-grade students are not included in the impact analysis for state tests because six of the 12 after-school centers do not administer a local assessment to their second-grade students. Response rates for school records are therefore based on students in grades three through five in the full study sample.

[5]The response rates for the state tests are all above 80 percent.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table D.2

Response Rates to Tests and Surveys for Students in the Reading Study Sample (One Year of Service)

Data Source	Full Sample	Enhanced Program	Regular Program
Cohort 1[a]			
Key outcome measures			
Follow-up SAT 10[b] (%)	93.23	93.27	93.17
DIBELS oral reading fluency (%)	91.09	90.61	91.71
DIBELS nonsense word fluency (%)	91.09	90.61	91.71
Additional outcome measures			
Regular-school-day teacher survey (%)	97.17	97.09	97.27
Student survey (%)	97.37	98.18	96.36
Follow-up state test score[c] (%)	83.42	85.37	80.98
Full study sample size (total = 989)		550	439
Cohort 2[d]			
Key outcome measures			
Follow-up SAT 10[b] (%)	93.86	93.14	94.81
DIBELS oral reading fluency (%)	94.31	93.33	95.60
DIBELS nonsense word fluency (%)	94.04	93.33	94.97
Additional outcome measures			
Regular-school-day teacher survey (%)	99.70	99.74	99.65
Student survey (%)	96.11	95.25	97.23
Follow-up state test score[c] (%)	84.10	82.68	85.85
Full study sample size (total = 668)		379	289

(continued)

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: Response rates are calculated from the full study sample for all students in the study and separately for students in each program group. The difference between the enhanced and regular program group response rates is not significant for any of the measures.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

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

[c]This calculation is based on students in grades three to five only. Second-grade students are excluded from the analysis of state test data because most sites do not test their students until third grade.

[d]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to

Appendix Table D.2 (continued)

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

[c]This calculation is based on students in grades three to five only. Second-grade students are excluded from the analysis of state test data because most sites do not test their students until third grade.

[d]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

y-
ıe
ɔf
ıg

analysis and in Cohort 2, 42 are excluded. Thus, the analysis samples consist of 905 students in Cohort 1 (which represents 92 percent of the full study sample) and 626 students in Cohort 2 (which represents 94 percent of the full study sample).[6]

Characteristics of Students in the Analysis Sample

This section examines (1) whether the analysis sample differs from the full study sample (which affects the generalizability of the findings to the full study sample), and (2) whether the statistical equivalence of the two research groups is preserved in the analysis sample (which affects whether the impact estimates are unbiased).

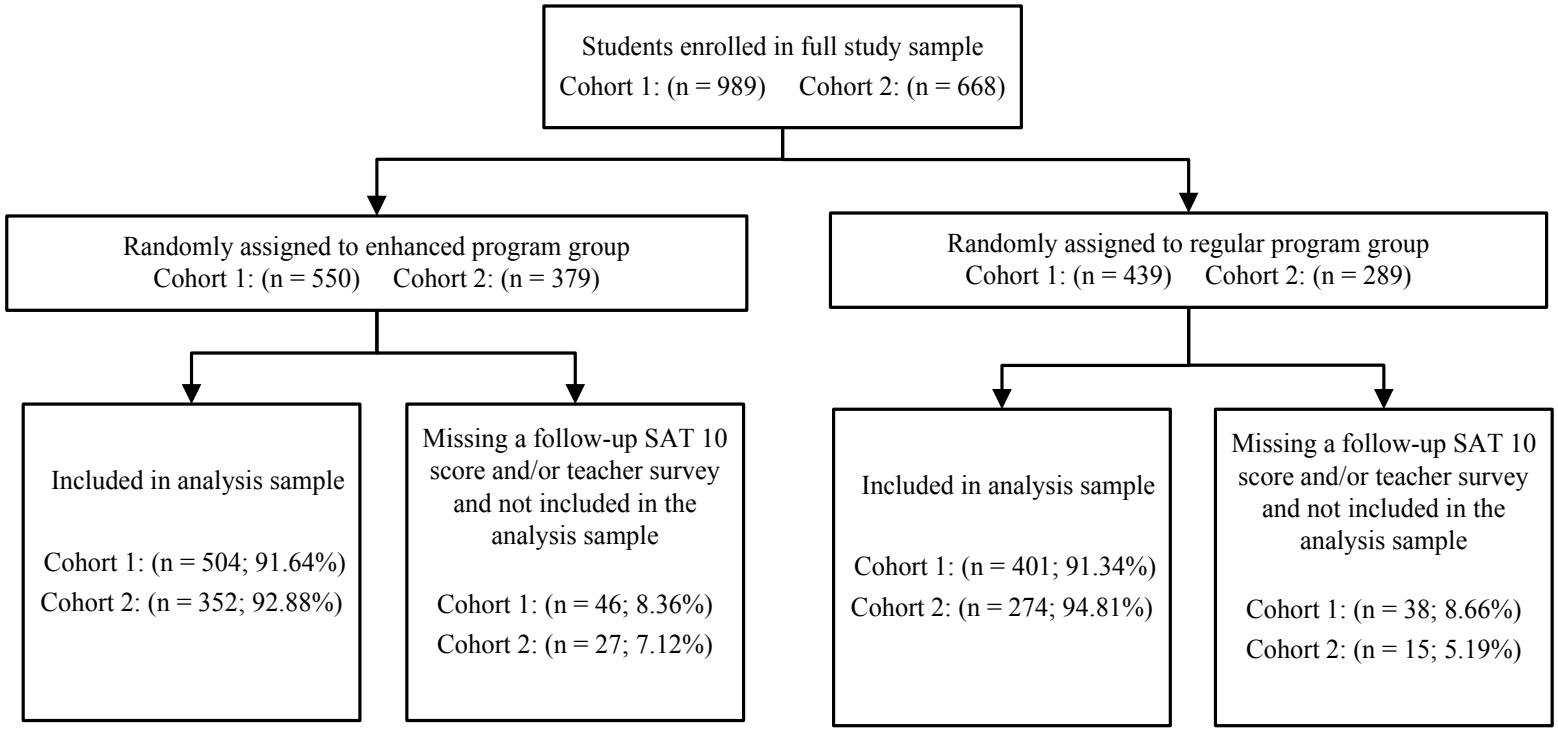
To examine the first issue, statistical tests were conducted to determine whether students in the analysis sample are different from students in the full study sample who were *excluded* from the analysis due to missing follow-up data. An overall F-test indicates that for both cohorts, these two groups of students are *not* systematically different in terms of their background characteristics (Cohort 1: F-value = 1.27, p-value = 0.19; Cohort 2: F-value = 1.19, p-value = 0.26). Thus, it is reasonable to assume that the impact findings presented in this report (based on the analysis samples) can be generalized to all students in the full study samples.

[6]In Cohort 1, 92 percent of students in the enhanced program groups had follow-up data on both the SAT 10 and the teacher survey, and 91 percent of students in the regular program groups had follow-up data on both these measures. The difference in response rates between the two program groups is not statistically significant (p-value = 0.87). In Cohort 2, 93 percent of students in the enhanced program group had follow-up data on both of these data sources compared with 95 percent students in the regular program group, and the difference in response rates between the two program groups is not statistically significant (p-value = 0.30).

The Evaluation of Academic Instruction in After-School Programs

Appendix Figure D.1

Flow of Students from Enrollment to Analysis in the Reading Sample (One Year of Service)



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 full study sample. All percentages are based on the number of students randomly assigned to either the enhanced or the regular program group.

In order to examine the second issue related to randomization, the characteristics of students in the enhanced and regular program were also compared. The characteristics of students in the analysis samples are presented in Table 8.1. As in the full study sample, an overall F-test indicates that there is a systematic difference in the background characteristics of students in the enhanced and regular program groups, in the two cohort-specific samples.

Two-Year Sample (Offer of Two Years of Service)

As explained in Chapter 2, the impact of offering students the opportunity to enroll in the enhanced program for two consecutive years is estimated by comparing the outcomes of students who were randomly assigned to either the enhanced after-school program (enhanced program group) or the regular after-school program (regular program group) for two consecutive school years. These students are referred to as the two-year sample (see Figure 2.3). Also recall that in order to preserve the experimental design of the study this sample includes students assigned to two years of the enhanced program, whether or not they attended both years (i.e., the intent-to-treat sample).

Characteristics of Students in the Full Study Sample

The full two-year study sample includes 370 students, 61 percent of whom (227 students) were randomly assigned to the enhanced after-school program in both years of the study and 39 percent of whom (143 students) were assigned to the regular after-school program group in both years.

Appendix Table D.3 presents the baseline characteristics of students in the full two-year study sample for each research group (enhanced program group and regular program group). An overall F-test indicates that there is a systematic difference in the background characteristics of students in the enhanced and regular program groups in the full two-year study sample ($F = 1.9$, p-value = 0.02).[7] This means that random assignment may not have been successful in creating two statistically equivalent research groups at baseline. (See Appendix H for a discussion of the analysis model that was used to control for observed differences in baseline characteristics between the enhanced and regular program groups, as well as the tests that were used to test the sensitivity of the impact findings to model and sample specifications).

[7]This occurs because students in the two-year study sample are a subset of the students in enrolled in grades two to four in the first-year study sample.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table D.3

Baseline Characteristics of Students in the Reading Full Study Sample (Offer of Two Years of Service)

Characteristic	Full Sample	Enhanced Program	Regular Program	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Enrollment						
2nd grade	133	80	53			
3rd grade	123	77	46			
4th grade	114	70	44			
Total	370	227	143			
Race/ethnicity (%)						
Hispanic		40.54	43.18	-2.64	-0.05	0.61
Black, non-Hispanic		39.38	33.51	5.87	0.11	0.12
White, non-Hispanic		12.07	17.45	-5.38	-0.15	0.23
Other		9.13	6.08	3.05	0.15	0.21
Gender (%)						
Male		53.73	44.83	8.90	0.16	0.17
Average age (years)		8.07	7.97	0.09	0.16	0.16
Overage for grade[a] (%)		15.68	10.25	5.43	0.14	0.27
Free/reduced-price lunch (%)						
Eligible (among information providers)		82.44	83.23	-0.79	-0.02	0.83
No information provided		2.87	4.41	-1.54	-0.07	0.51
Average household size		2.09	2.15	-0.06	-0.05	0.66
Single-adult household (%)		27.00	25.46	1.54	0.03	0.79
Mother's education level (%)						
Did not finish high school		27.46	22.28	5.18	0.11	0.34
High school diploma or GED certificate		29.51	21.88	7.63	0.16	0.16
Some postsecondary study		36.65	45.90	-9.25	-0.17	0.15
No information provided		6.39	9.95	-3.56	-0.13	0.42
SAT 10 baseline reading total scaled scores		550.97	559.47	-8.51 *	-0.26	0.02
Vocabulary/word reading[b]		541.65	552.78	-11.13 *	-0.25	0.05
Reading comprehension		550.46	561.24	-10.78 *	-0.28	0.01
Word study skills[c]		564.16	565.52	-1.35	-0.03	0.76
Sample size (total = 370)		227	143			

(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.

Appendix Table D.3 (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 enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 effect size for each characteristic is calculated as a proportion of the standard deviation for students in the two-year sample regular program group.

An F-test was calculated 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-value (F = 1.89) is signficant at the 5 percent level.

[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.

Response Rates

Appendix Table D.4 presents response rates for each data source on follow-up student outcomes, both overall and by program group, in the full two-year study sample.[8] The first two rows present response rates for the two key sources of data used in the impact analysis — the follow-up SAT 10 total score and the DIBELS Oral Reading Fluency (ORF) score.[9] The last three rows in each panel report the response rates for the other data sources used in analysis — the regular-school-day teacher questionnaire (used to measure student academic behavior), the student survey (used to measure the service contrast), and the follow-up state test score (used as a supplementary measure of students' academic performance).

[8]Outcomes are measured in Spring 2007.

[9]Response rates for the DIBELS Nonsense Word Fluency (NWF) assessment are not presented because impacts on this measure are not examined (due to the fact that data on this measure are only available for third-grade students at follow-up).

The Evaluation of Academic Instruction in After-School Programs

Appendix Table D.4

Response Rates to Tests and Surveys for Students in the Reading Study Sample (Offer of Two Years of Service)

Data Source	Full Sample	Enhanced Program	Regular Program
Key outcome measures			
Follow-up SAT 10[a] (%)	73.24	74.89	70.63
DIBELS oral reading fluency (%)	73.24	74.89	70.63
Additional outcome measures			
Regular-school-day teacher survey (%)	77.30	78.41	75.52
Student survey (%)	75.41	76.21	74.13
Follow-up state test score[b] (%)	64.32	67.84	58.74
Full study sample size (total = 370)		227	143

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: Response rates are calculated from the full study sample for all students in the study and separately for students in each program group. The difference between the enhanced and regular program group response rates is not significant for any of the measures.

[a]This calculation is based on responses to the total reading scaled score.

[b]This calculation is based on students in grades three to five only. Second-grade students are excluded from the analysis of state test data because most sites do not test their students until third grade.

As seen in this table, response rates for all data sources are above 64 percent, and the response rates of students in the enhanced and regular program group do not differ by a statistically significant amount on any measure. Recall that all eligible students from the fall of 2005 were randomly assigned in the second year, whether or not they reapplied for the second year of the study. And for those who did not reapply, the study team tried to collect follow-up data. Thus, response rates in the full two-year study sample are driven downwards by the fact 49 percent (81 students) of students who did not reapply to the study the second year also did not consent to follow-up data collection[10] That said, among students who *did* consent to

[10]Forty-seven percent of students in the enhanced program group who did not reapply also did not consent to follow-up data collection, while 53 percent of students in the regular program group who did not reapply did (continued)

follow-up data collection (all students who participated in the second year and 51 percent of students who did not reapply to the study the second year), response rates are above 82 percent on all data sources in Table D.4.

Constructing the Analysis Sample

As noted earlier, the analysis sample is limited to students with data on both the follow-up SAT 10 assessment and the regular-school-day teacher survey. The consort chart in Figure D.2 describes the construction of the two-year analysis sample. As shown, 100 students are excluded from the reading analysis sample (81 of these students are students who did not apply in the second year and did not provide consent for follow-up data collection). Thus, the two-year analysis sample consists of 270 students (which represents 73 percent of the full two-year study sample).[11]

Characteristics of Students in the Analysis Sample

Statistical tests were conducted to determine whether students in the two-year analysis sample are different at baseline from students in the full two-year study sample who were *excluded* from the analysis due to missing follow-up data. An overall F-test indicates that these two groups of students are systematically different in terms of their background characteristics ($F = 2.6$, p-value = 0.00). Students were more likely to be included in the two-year analysis sample if they were Hispanic (p-value = 0.04) or if they were not overage for grade at baseline (p-value = 0.01).[12]

The characteristics of students in the enhanced and regular program groups were also compared (see Table 9.1). As in the full two-year study sample, an overall F-test indicates that there is a systematic difference in the background characteristics of these two groups of students in the two-year analysis sample ($F = 1.5$, p-value = 0.05).

---

not consent. The difference in non-consent rates between program groups does not differ by a statistically significant amount (p-value = 0.45).

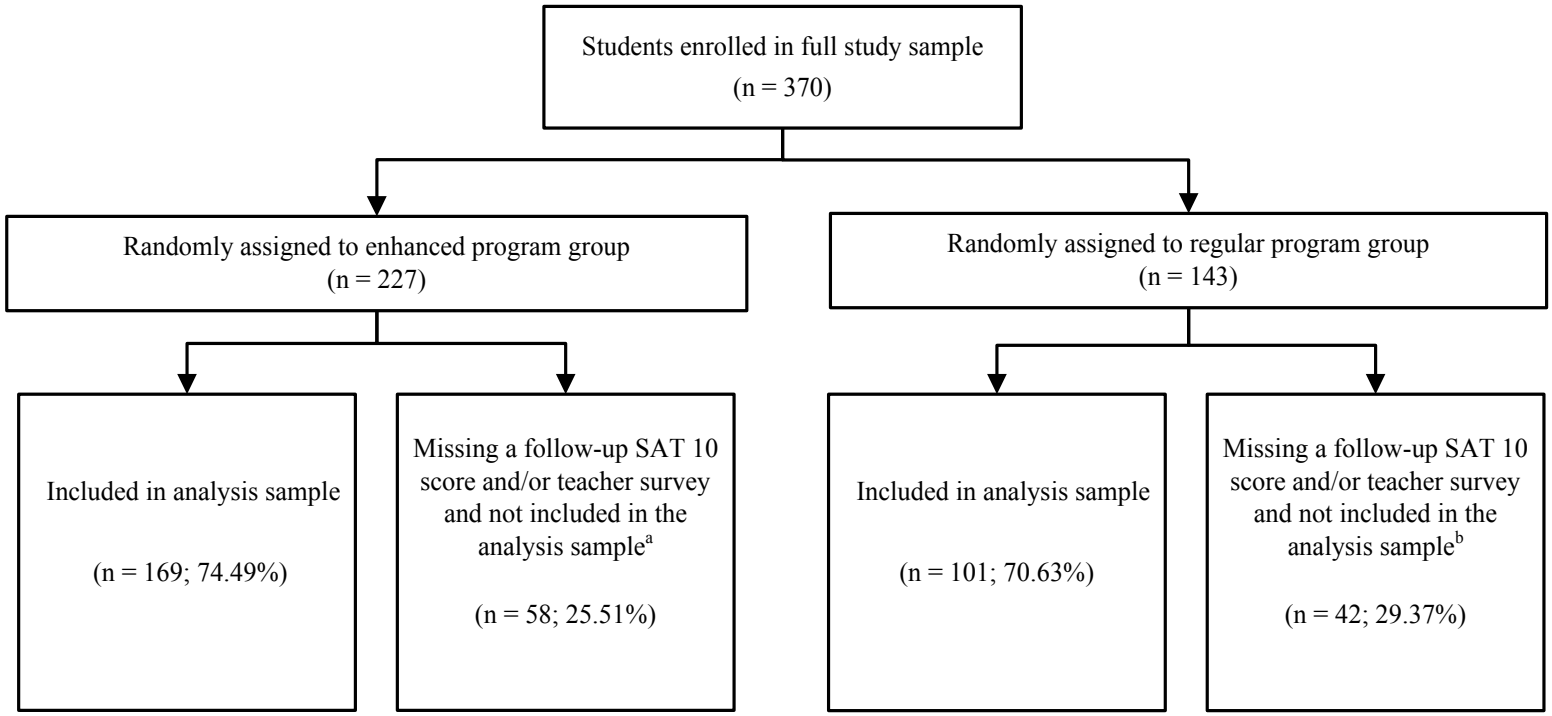
[11] Seventy-four (74) percent of students in the enhanced program group had follow-up data on both the SAT 10 and the teacher survey, while 71 percent of students in the regular program group had data on both of these sources. The difference in response rates between the two program groups is not statistically significant (p-value = 0.43)

[12] As noted earlier in this appendix, students excluded from the two-year analysis sample are primarily students who did not apply in the second year (nonapplicants) and did not consent to a second year of data collection. As will be explained in Appendix H, nonapplicants who *did* consent to follow-up data collection are weighted to account for nonapplicants who did *not* consent to follow-up data collection.

The Evaluation of Academic Instruction in After-School Programs

Appendix Figure D.2

Flow of Students from Enrollment to Analysis in the Reading Sample (Offer of Two Years of Service)



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

NOTES: This figure explains how the two-year reading analysis sample was created from the full two-year study sample. All percentages are based on the number of students randomly assigned to either the enhanced or the regular program group.

[a]Among these, 46 are students who did not reapply and did not provide consent for follow-up data collection.

[b]Among these, 35 are students who did not reapply and did not provide consent for follow-up data collection.

Appendix E

Implementation Measures from Structured Protocol Observations and Class Record Forms

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. However, no formal measure of reliability was computed for these data. 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 E.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 six points per observation by using all the instructional elements (shown in Appendix Box E.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 E.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 were 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.

Uses curriculum materials. 1 point was 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 pages 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 was 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.[a] ½ point was 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 E.1 (continued)

Direct/differentiated instruction (to individuals and small groups in rotations). 1 point was 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 was 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 was awarded if:

- Observer checked box indicating each component section (warm-ups, workout session and cool-downs) is completed in the allotted timeframe

NOTE: [a]For warm-ups and cool-downs, ½ point was awarded by district coordinators for each in the first implementation year and 1 point was awarded by district coordinators in the second implementation year. In the second year, researchers re-scaled the 1 point to ½ point to make it consistent and comparable to the first year.

Each class was observed, on average, three times during the year. For each class, observation scores were averaged together.[1] In other words, if a class that was observed three times received 5 of 6 possible points during two of the observations and 6 of 6 possible points during a third observation, then the observation rating for that class is 5.3. In the first year, the average observation rating was 5.98 (standard deviation = 0.13), and, in the second year, the average observation rating was 6 (standard deviation = 0.00).

Observations of Adventure Island

Appendix Box E.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

[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 Box E.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.

Uses curriculum. 1 point was 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 was 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 E.2 (continued)

Completes in allotted time. 1 point was 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.[a] 1/2 point was 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.[a] 1/2 point was 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.[a] 1/2 point was 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.[a] 1/2 point was 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 E.2 (continued)

Teaches phonics in Alphie's Lagoon and Captain's Cove. 1 point was 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

NOTE: [a]For uses cooperative learning strategies, awards points for cooperative learning, models fluency, and awards points for fluency, ½ point was awarded by district coordinators for each in the first implementation year and 1 point was awarded by district coordinators in the second implementation year. In the second year, researchers re-scaled the 1 point to ½ point to make it consistent and comparable to the first year.

slightly different components for the higher and lower reading levels, with a maximum score of five points per observation for Discovery Bay and Treasure Harbor classes and six points per observation for Alphie's Lagoon and Captain's Cove classes.[2] The instructional elements (shown in Appendix Box E.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] In other words, if a Discovery Bay class that was observed three times received 4 of 5 possible points during two of the observations and 5 of 5 possible points during a third observation, then the observation rating for that class is 4.3. In the first year, the average observation rating for Alphie's Lagoon and Captain's Cove classes was 5.14 (standard deviation = 0.56) and ,in the second year, 5.09 (standard deviation = 0.61). For

[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.

Discovery Bay and Treasure Harbor classes, the average observation rating in the first year was 4.20 (standard deviation = 0.41) and, in the second year, 3.90 (standard deviation = 0.30).

Harcourt School Publishers' Class Record Forms

As a way for teachers to keep track of student progress through the Mathletics' skills, Harcourt School Publishers created Class Record Forms and trained the teachers on how to fill out the form for classroom management purposes. As part of these forms, teachers enter the student's pretest score, then check off the skills for which they provide direct instruction (as a result of not mastering these skills on the pretest). The form also captures, for each skill, which elements are assigned to students (e.g. computer instruction, board games, etc.), as well as the skill-by-skill posttest.

The average number of days spent on a skill is calculated for each student using the total number of skills for which the teacher indicated providing direct instruction and the students' overall number of days attended (as captured by the attendance data).[4]

[4]Attendance data were collected from students in the enhanced and regular program groups for the days on which the enhanced program met.

Appendix F

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 F.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 two through five 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.

The Stanford Achievement Test, Tenth Edition (SAT 10), abbreviated battery was selected and administered by local data collection staff, who were part of the research team, at both baseline and follow-up.[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

[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.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table F.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 F.2 and F.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.

> 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 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 2002. 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 two through five 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 two through five 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. Word study skills measures structural and phonetic analysis, such as identifying and decoding compound words and contractions and recognizing sounds of consonants and vowels. 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 by local data collection staff, who were part of the research team, at follow-up to second- and third-graders in the reading centers during the first implementation year and to all students in the reading centers during the second implementation year, in addition to the SAT 10.[2]

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). In this study, students were tested on measures of fluency — oral reading fluency (ORF) and nonsense word fluency (NWF).[3]

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

[2]A professional trainer of DIBELS from Sopris West was hired to go to each reading center and train the data collection staff, who then administered the DIBELS to students at their center. This trainer was then available for questions and advice.

[3]The nonsense word fluency measure is not recommended for older grades; thus fourth- and fifth-graders were not administered this part of the test.

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 one 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 Records 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 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 F.2 and F.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:

$$Z_{ijg} = \frac{(Y_{ijg} - \bar{Y}_{jg})}{s.d._{jg}(Y_{ijg})}$$

where:

Z_{ijg} = the standardized score for student *i* in grade *g* from site *j*.

Y_{ijg} = the raw score for student *i* in grade *g* from site *j* on the locally administered test.

$\bar{Y}_{jg}$ = the average raw score for students in grade *g* in site *j* on the locally administered test.

$s.d._{jg}(Y_{ijg})$ = the standard deviation of the raw test scores for students in grade *g* in site *j*.

This transformed measure was then used as an outcome for student achievement. The z-score represents a student's deviation from the average level of achievement among students in their grade, as a proportion of the variation in achievement among students in their grade (i.e., standard deviation or effect size units).

Appendix Table F.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 F.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
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.

Appendix Table F.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
New Mexico Standards Based Assessment (NMSBA)	Criterion-referenced	Reading and Listening for Comprehension; Writing and Speaking for Expression; Literature and Media
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

(continued)

Appendix Table F.3 (continued)

Test	Criterion- or Norm-Referenced	Test Content
Pennsylvania System of School Assessment (PSSA)	Criterion-referenced	Comprehension and Reading Skills; Interpretation and Analysis of Fiction and Non-Fiction Text
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.

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."

However, it should be noted that no additional instructions were given about the survey questions, definition of terms, or the rating scale. Teachers were only instructed in the logistics of distribution and collection of surveys. And all three variables were measured with a single survey item, thus compromising the reliability of these measures.

Appendix G

Statistical Model and Sensitivity Analyses (Impact of Offering One Year of Service)

This appendix describes the statistical model used to estimate the impact of offering students one year of the enhanced after-school program *either in the first or second year of implementation* and presents findings for additional impact analyses that were conducted to test the sensitivity of the results to sample and model specifications.

The first additional analysis examines impacts on locally administered standardized (state) tests. This outcome has policy relevance, given that scores on these tests are typically tied to rewards and sanctions in the local accountability system. An important issue to note here is that locally administered test data were not always available for second-graders in some study sites, since testing usually begins in the third grade. As a result, the impacts on state tests presented in this appendix are based on students in grades three to five only; impacts on the SAT 10 for this same subgroup of students are also presented for comparative purposes.

The second additional analysis presented in this appendix examines impacts for the SAT 10 respondent sample. The impact findings presented in the main body of the report are based on an analysis sample restricted to students with spring follow-up data on both the SAT 10 assessment *and* the regular-school-day teacher survey. The latter restriction was imposed because measures of students' academic behavior are created from the teacher survey. As discussed in the report, however, the enhanced program did not affect students' academic behaviors. Hence, the second criteria for sample inclusion was dropped, and impacts on the SAT 10 were re-estimated based on *all* students that completed the SAT 10 assessment (whether or not they had teacher survey data).

The third additional analysis presented in this appendix is a sensitivity test of the impact findings to the chosen specification for the impact model. Specifically, the impact model includes student baseline covariates in order to explain random differences in the outcomes of students (and therefore improve the precision of the impact estimates). Strictly speaking, these covariates need not be included in the analysis because randomization creates the expectation that students assigned to the enhanced and regular program are similar on observed and unobserved characteristics prior to the intervention, and any subsequent differences between the outcomes of students in the these two groups can be fairly attributed to the effects of the enhanced program. Rather, such covariates are typically included to increase the *precision* of the estimates. Hence, this appendix presents impacts from models that do not include these baseline covariates. (As will be explained in the reading section, this sensitivity analysis differs somewhat for the reading sample because randomization did not produce two statistically equivalent groups at baseline.)

Analysis of Program Impacts

Impacts on student outcomes are estimated for each of the two academic programs separately (the math or the reading program) by comparing the outcomes of students assigned to the enhanced program for one school year (enhanced program group) and the outcomes of students assigned to the regular after-school program for one school year (regular program group). As explained in Chapter 2, this analysis is based on students in the Cohort 1 sample and the Cohort 2 sample (see Figure 2.2).

The Model

The impact of enrolling in the enhanced program for one school year is estimated for each outcome using the following statistical model:

$$Y_{ik} = \gamma_0 Y_{-1,ik} + \beta_0 T_{ik} + \sum_k \gamma_{1k} B_{ik} + \sum_S \gamma_{2s} X_{sik} + \varepsilon_{ik} \quad (1)$$

where:

T_{ik} = Indicator of program group membership (treatment status). This indicator is equal to 1 if student i from random assignment block k was assigned to the enhanced program and zero otherwise

Y_{-1ik} = The pretest score for student i from random assignment block k before random assignment[1]

B_{ik} = Random assignment block indicators; equal to 1 if student i is in random assignment block k and zero otherwise[2]

X_{sik} = The set of s other student-level covariates for student i in random assignment block k

ε_{ik} = A student-level random error term assumed to be independently and identically distributed.

[1]Pretest scores are scaled scores from the SAT 10 tests (SAT 9 for a couple of centers) in reading and math administered in the fall of 2005 (for Cohort 1) and either the fall of 2006 (for new students in Cohort 2) or the spring of 2006 (for returning students in Cohort 2), before the start of the after-school program. Total scores for math are used in the math analysis, and total scores for reading are used in the reading analysis.

[2]In the Cohort 1 sample, random assignment block is defined by grade j and center c in fall 2005 (60 blocks for math and 48 for reading). In the Cohort 2 sample, random assignment block is defined by first-year treatment status (regular program, or new to study) by grade j and center c in fall 2006 (104 blocks for math and 84 blocks for reading).

The coefficient, β_0, represents the overall impact of being randomized to one year of the enhanced program 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. There are several features to note about this model:

- β_0 is a "fixed-effect" estimate that addresses the question: What is the effect of the enhanced program 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 squared (OLS) regression is used to estimate Equation (1).
- Indicators for random assignment blocks (B_{ik}) are included in the model to reflect the design feature (i.e., differential rates of treatment assignment by block) and to control for variation in mean outcome levels across blocks (which can be due to different characteristics of centers, school settings, etc).
- The model controls for the student's pretest achievement score. This information can increase the precision of impact estimates, especially for fixed-effect models, because pretests substantially reduce within-block random error in the outcome measure, 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 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.

Other Analytical Issues

Missing Covariates

For the baseline achievement test, there are 13 missing cases (two for math and 11 for reading). For other covariates, there are 7 percent or fewer missing cases.[3] To keep the sample

[3]Across both cohort samples for the math analysis, four students are missing a race/ethnicity indicator, 55 are missing a free lunch status indicator, 35 are missing information about single-adult household, and 122 are missing information about mother's education. Across both cohort samples for the reading analysis, six are missing a race/ethnicity indicator, 67 are missing a free lunch status indicator, 15 are missing information about (continued)

as complete as possible, the missing values were imputed with the mean value of the random assignment block and program group[4] to which the student belongs.[5] 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.

Weighting of Grades in the Cohort 2 Sample

As shown in the tables of student baseline characteristics in Appendix C (math) and D (reading), the Cohort 1 and Cohort 2 samples are characterized by different grade distributions. While students in the Cohort 1 sample are approximately equally distributed across grades, the Cohort 2 sample includes a proportionately larger percentage of students in grade two than other grades.

This occurs because of the way in which the Cohort 2 sample is defined. Recall from Chapter 2 that the Cohort 2 sample excludes students *who were in the enhanced program group* in the first year of the study (the "EE" and "ER" students in Figure 2.1). Because students enrolled in grade two in the second year of the study could *not* have been part of the study in its first year unless they were retained, the Cohort 2 sample includes a proportionately larger percentage of students in second grade (32 percent) than other grades.[6]

In order to ensure that second-grade students do not have a disproportionate weight in the findings, all analyses that include the Cohort 2 sample are weighted to reflect the distribution of students across grades in the full second-year randomization sample (i.e., the sample prior to the exclusion of the "EE" and "ER" students).[7]

---

single-adult household, and 91 are missing information about mother's education. (No students are missing gender or age.)

[4]In other words, for Cohort 1, the mean value for students in program group *p* (enhanced or regular) in grade *j* in center *c* in fall 2005. For Cohort 2, the mean value for students in program group *p* (enhanced or regular) by first-year treatment status, in grade *j* in center *c* in fall 2006.

[5]Rather than imputing the missing reading or math SAT 10 total scaled score, the mean raw score for the missing subtest was imputed and then the subtest raw scores were added to obtain an imputed total raw score. The student was then assigned the scaled score associated with their imputed total raw score. This was done so that — if there is an actual score for one or more of the subtests — the imputed total score incorporates that information.

[6]In other words, most second-grade students (97 percent) in the second year of the study are "new" to the study ("NE" and "NR" in Figure 2.2) and therefore not among the students excluded from this analysis.

[7]Specifically, in each after-school center, students in the Cohort 2 sample in each grade are weighted up to account for returning students in their grade who were randomized in the second year of the study but who are not part of the Cohort 2 sample (i.e., returning students in the "EE" and "ER" group, see Figure 2.3). These weights were then normalized to sum to the actual Cohort 2 sample.

Additional Analyses for the Math Sample

This section presents supplementary impact findings for the enhanced math program. The section begins with a discussion of impacts on locally administered math assessments. This is followed by a presentation of impacts on the SAT 10 respondent sample. The section concludes by examining impacts based on an alternate specification of the statistical model.

Impact on State Assessments

Table G.1 presents estimated program impacts on students' performance on locally administered math tests (grades three to five). Because these test scores were standardized within each study site, all estimated impacts are in effect size units.[8] Also, because local assessment data are only available for students in grades three to five, the table also shows program impacts on the study-administered SAT 10 tests for this specific sample of students, for comparative purposes.

As shown in this table, the impact of the enhanced math program on the locally administered math test for students in Cohort 1 is 0.05 standard deviation and not statistically significant (p-value = 0.35). For students in Cohort 2, there is a statistically significant difference in the impact on locally administered tests of 0.18 standard deviation (p-value = 0.01). However, the difference in impacts across cohorts is not statistically significant (p-value = 0.67); thus, it cannot be concluded that the impact of the enhanced program on locally administered tests differs between implementation years.

Impact on the SAT 10 Respondent Sample

Impacts on student achievement were re-estimated for the sample of all SAT 10 respondents to make sure that no imbalance was created when the full sample was limited to the analysis sample. This change in the sample added six observations. Table G.2 presents impacts on SAT 10 math test scores for the SAT 10 respondent sample. As seen in the table, the magnitude of the estimates changes very little relative to what was presented in Chapter 4 of the report, and the patterns of statistical significance are the same.

[8] Appendix F describes the standardization of the test score variable.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table G.1

Impact of the Enhanced Math Program on Student Achievement in the Math Analysis Sample for Grades 3 to 5 (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]				
State test scaled scores	0.06	0.01	0.05	0.35
SAT 10 math total scaled scores	622.26	619.39	0.07	0.10
Sample size (total = 767)	434	333		
Cohort 2[b]				
State test scaled scores	0.01	-0.16	0.18 *	0.01
SAT 10 math total scaled scores	619.62	615.68	0.10	0.09
Sample size (total = 516)	297	219		

(continued)

SOURCES: MDRC calculations are from results on state tests administered in the 2006-2007 school year and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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 analysis is restricted to students for whom a state test score was obtained. The resulting state test analysis sample represents 92 percent of the third- through fifth-graders in the 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 F for details.

Based on the SAT 10 national norming sample, math total scaled scores range from 428 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

Appendix Table G.1 (continued)

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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: SAT 10 = 38.90; state test = 1.13. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

Model Specification Tests

All impacts were re-estimated with a model that has no covariates other than the random assignment block indicators and the treatment status indicator (i.e., without student pretests and background characteristics):

$$Y_{ik} = \beta_0 T_{ik} + \sum_k \gamma_{1k} B_{ik} + \varepsilon_{ik} \qquad (2)$$

Because this study is based on a randomized experiment, both sets of impact estimates — those that are and are not adjusted for student characteristics — should provide similar estimates of the treatment effect. The precision of the estimated impact, however, should be higher for the adjusted estimates.

As can be seen in Table G.3, dropping the student characteristics from the statistical model and only controlling for the randomization strata does not substantially affect the magnitude of the impact findings, as expected. Also, the patterns of statistical significance are the same as those presented in Chapter 4.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table G.2

Impact of the Enhanced Math Program on Student Achievement for the SAT 10 Respondent Sample (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]					
SAT 10 math total scaled scores	607.12	603.62	3.50 *	0.09	0.01
Problem solving	608.06	605.51	2.55	0.06	0.11
Procedures	607.69	601.90	5.79 *	0.11	0.00
Sample size (total = 1,145)	635	510			
Cohort 2[b]					
SAT 10 math total scaled scores	606.77	603.37	3.40	0.09	0.07
Problem solving	608.85	606.24	2.61	0.07	0.17
Procedures	605.26	600.73	4.52	0.09	0.09
Sample size (total = 797)	463	334			

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

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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: 389 to 796, 414 to 776, and 413 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table G.3

Impact of the Enhanced Math Program on Student Achievement for the Analysis Sample, with Random Assignment Indicators as the Only Model Covariates (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]					
SAT 10 math total scaled scores	607.01	602.49	4.52 *	0.12	0.02
Problem solving	607.88	604.09	3.78	0.09	0.06
Procedures	607.63	600.98	6.65 *	0.13	0.01
Sample size (total = 1,144)	634	510			
Cohort 2[b]					
SAT 10 math total scaled scores	606.72	603.08	3.64	0.09	0.13
Problem solving	608.80	606.21	2.60	0.06	0.28
Procedures	605.20	600.10	5.09	0.10	0.11
Sample size (total = 792)	461	331			

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

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: 389 to 796, 414 to 776, and 413 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 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

Additional Analyses for the Reading Sample

This section presents additional impact findings for the enhanced reading program. The section begins with a discussion of impacts on locally administered reading assessments. This is followed by a presentation of impacts on the SAT 10 respondent sample. The section concludes by examining impacts based on alternate specifications of the statistical model.

Impact on State Assessments

Table G.4 presents estimated program impacts on students' performance on locally administered reading tests (grades three to five). Because these test scores were standardized within each study site, all estimated impacts are in effect size units.[9] Also, because local assessment data is only available for students in grades three to five, the table also shows program impacts on the study-administered SAT 10 tests for this specific sample of students, for comparative purposes.

As seen in this table, the impact of the enhanced reading program on the locally administered reading test for this particular sample of students is not statistically significant for either of the two cohort-specific samples. (Impacts on SAT 10 total reading scores for these students are also not statistically significant.)

Impacts for the SAT 10 Respondent Sample

Impacts on student achievement were re-estimated 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 18 observations. Table G.5 presents impacts on SAT 10 reading test scores for this SAT 10 respondent sample. As seen in the table, the magnitude of the estimates changes very little relative to what was presented in Chapter 8 of the report, and the patterns of statistical significance are the same, with the exception of the impact on reading comprehension in the Cohort 1 sample (which is no longer statistically significant).

Model Specification Tests and Other Sensitivity Tests

As noted in the introduction, randomization creates the expectation that students assigned to the enhanced and program group are similar on average at baseline within random assignment block. Hence, the purpose of including student covariates in the impact model is simply to improve the precision of the impact estimates (reduce the standard error).

[9] Appendix F describes the standardization of the test score variable.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table G.4

Impact of the Enhanced Reading Program on Student Achievement in the Reading Analysis Sample for Grades 3 to 5 (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact Effect Size	P-Value for the Estimated Impact
Cohort 1[a]				
State test scaled scores	-0.08	-0.05	-0.03	0.62
SAT 10 reading total scaled scores	600.85	603.53	-0.08	0.10
Sample size (total = 589)	337	252		
Cohort 2[b]				
State test scaled scores	0.07	0.06	0.01	0.90
SAT 10 reading total scaled scores	605.42	605.29	0.00	0.95
Sample size (total = 380)	208	172		

(continued)

SOURCES: MDRC calculations are from results on state tests administered in the 2006-2007 school year and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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 analysis is restricted to students 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 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 F for details.

Based on the SAT 10 national norming sample, reading total scaled scores range from 416 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: SAT 10 = 33.19; state test = 1.16. The standard deviation in the

Appendix Table G.4 (continued)

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: SAT 10 = 33.19; state test = 1.16. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

However, in the reading sample, randomization did not produce two statistically equivalent groups at baseline (see Chapter 8). Most notably, students in the enhanced program group had lower pretest scores on average than students in the regular program group. Hence, in this situation, it is important to control for student background characteristics in the impact model, especially student pretests. Otherwise, the analysis may produce biased estimates of the program's impact. The three sensitivity analyses presented in this section confirm that including student pretests and background characteristics in the model effectively controls for baseline differences between the enhanced and regular program groups.

No Covariates Other Than Block

As a first step, all impacts were re-estimated with a model that has no covariates other than the "block" (random assignment unit) indicators and the treatment status indicator (see equation 2). Had random assignment resulted in two statistically equivalent research groups, this model would produce impact estimates that are similar in magnitude to the adjusted estimates presented in Chapter 8.

However, as seen in Table G.6, when all student covariates are dropped from the impact model, the estimated impacts become smaller. As a result, some estimates become more negative to the extent that they become statistically significant. This happens because the enhanced reading group was lower-achieving on average before the start of the program (most notably in the Cohort 1 sample), and the impact model no longer controls for this difference in prior achievement. Hence, the impact estimates in Table G.6 are biased, in that they do not represent the true causal effect of the program on student achievement. These results confirm the importance of controlling for student background characteristics in the model.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table G.5

Impact of the Enhanced Reading Program on Student Achievement for the SAT 10 Respondent Sample (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
Cohort 1[a]					
SAT 10 reading total scaled scores	589.38	591.47	-2.09	-0.06	0.13
Vocabulary	583.18	585.42	-2.25	-0.05	0.26
Reading comprehension	590.25	593.02	-2.78	-0.08	0.11
Word study skills (grades 2-4)[b]	589.91	590.28	-0.37	-0.01	0.88
DIBELS (grades 2-3)[c]					
Oral fluency score	73.94	72.33	1.61	0.05	0.46
Nonsense word fluency score	66.87	64.37	2.50	0.07	0.29
Sample size (total = 922)	513	409			
Cohort 2[d]					
SAT 10 reading total scaled scores	593.82	593.64	0.19	0.01	0.91
Vocabulary	587.28	585.84	1.44	0.03	0.58
Reading comprehension	595.63	596.98	-1.36	-0.04	0.52
Word study skills (grades 2-4)[b]	593.56	592.08	1.48	0.04	0.63
DIBELS (grades 2-3)[c]					
Oral fluency score	78.81	75.80	3.01	0.09	0.22
Nonsense word fluency score	75.44	70.56	4.88	0.14	0.11
Sample size (total = 627)	353	274			

(continued)

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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

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: 374 to 787, 439 to 777, 412 to 739, and 410 to 740. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximum 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

Appendix Table G.5 (continued)

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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98; nonsense = 36.13. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

[c]The DIBELS sample includes only second- and third-grade students because the nonsense word fluency subtest and the oral fluency subtest were not administered to fourth- and fifth-grade students in both study years.

[d]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

No Covariates Other Than Block and Pretest

Impacts were also re-estimated based on a model that includes *prior achievement* as a student-level covariate (the variable on which the two research groups differed the most at baseline) but that does not include the set of student demographic characteristics:

$$Y_{ik} = \gamma_0 Y_{-1,ik} + \beta_0 T_{ik} + \sum_k \gamma_{1k} B_{ik} + \varepsilon_{ik} \qquad (3)$$

As can be seen from Table G.7, the magnitudes of the estimates produced by this model are *not* substantially different from those presented in Chapter 8. This suggests that controlling for students' pretest scores effectively adjusts for observed differences between the enhanced and regular program groups at baseline.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table G.6

Impact of the Enhanced Reading Program on Student Achievement for the Analysis Sample, with Random Assignment Indicators as the Only Model Covariates (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
Cohort 1[a]					
SAT 10 reading total scaled scores	588.66	595.17	-6.51 *	-0.20	0.00
Vocabulary	582.73	589.71	-6.98 *	-0.16	0.01
Reading comprehension	589.47	596.86	-7.39 *	-0.20	0.00
Word study skills (grades 2-4)[b]	589.44	594.96	-5.52	-0.13	0.06
DIBELS (grades 2-3)[c]					
Oral fluency score	73.61	78.68	-5.07	-0.15	0.09
Nonsense word fluency score	66.19	68.46	-2.27	-0.06	0.41
Sample size (total = 905)	504	401			
Cohort 2[d]					
SAT 10 reading total scaled scores	593.95	594.83	-0.89	-0.03	0.71
Vocabulary	587.45	586.54	0.91	0.02	0.77
Reading comprehension	595.75	598.67	-2.92	-0.08	0.26
Word study skills (grades 2-4)[b]	593.64	594.34	-0.70	-0.02	0.84
DIBELS (grades 2-3)[c]					
Oral fluency score	78.91	78.76	0.16	0.00	0.96
Nonsense word fluency score	75.52	72.49	3.03	0.08	0.37
Sample size (total = 626)	352	274			

(continued)

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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 374 to 787, 439 to 777, 412 to 739, and 410 to 740. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximum 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 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 mean covariate values for the

Appendix Table G.6 (continued)

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 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98; nonsense = 36.13. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

[c]The DIBELS sample includes only second- and third-grade students because the nonsense word fluency subtest and the oral fluency subtest were not administered to fourth- and fifth-grade students in both study years.

[d]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

Exclusion of Blocks with Baseline Differences

While controlling for students' pretest scores appears to adjust for *observed* baseline differences between the two research groups, it may not control for remaining *unobserved* differences between the two groups, in which case the impact findings would be biased.

An additional sensitivity test was conducted to explore this possibility. For each cohort-specific sample, the center-by-grade blocks with the largest differences in baseline characteristics between students in the enhanced and regular program groups were dropped from the analysis (14 center-by grade blocks were excluded in total; six blocks in Cohort 1

The Evaluation of Academic Instruction in After-School Programs

Appendix Table G.7

Impact of the Enhanced Reading Program on Student Achievement for the Analysis Sample, Without Demographic Characteristics as Model Covariates (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
Cohort 1[a]					
SAT 10 reading total scaled scores	588.66	591.34	-2.68 *	-0.08	0.05
Vocabulary	582.73	584.90	-2.17	-0.05	0.27
Reading comprehension	589.47	593.22	-3.75 *	-0.10	0.03
Word study skills (grades 2-4)[b]	589.44	590.22	-0.78	-0.02	0.76
DIBELS (grades 2-3)[c]					
Oral fluency score	73.61	73.10	0.51	0.02	0.81
Nonsense word fluency score	66.19	64.82	1.37	0.04	0.56
Sample size (total = 905)	504	401			
Cohort 2[d]					
SAT 10 reading total scaled scores	593.95	593.44	0.51	0.02	0.77
Vocabulary	587.45	584.95	2.50	0.06	0.32
Reading comprehension	595.75	597.31	-1.56	-0.04	0.46
Word study skills (grades 2-4)[b]	593.64	591.96	1.69	0.04	0.57
DIBELS (grades 2-3)[c]					
Oral fluency score	78.91	75.65	3.26	0.10	0.19
Nonsense word fluency score	75.52	70.39	5.13	0.14	0.09
Sample size (total = 626)	352	274			

(continued)

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: Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 374 to 787, 439 to 777, 412 to 739, and 410 to 740. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximum 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 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

Appendix Table G.7 (continued)

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 p-value is less than or equal to 5 percent.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98; nonsense = 36.13. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

[c]The DIBELS sample includes only second- and third-grade students because the nonsense word fluency subtest and the oral fluency subtest were not administered to fourth- and fifth-grade students in both study years.

[d]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and who thus were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

and eight blocks in Cohort 2).[10] In the remaining sample of students, it was found that there was no longer a systematic difference between students in the enhanced and regular program

[10]Center-by-grade blocks were excluded in two stages. In the first stage, the baseline characteristic with the most statistically significant difference between students in the enhanced and regular program group was identified, and the blocks with the largest between-group differences on this characteristic were excluded (top 10 percent excluded). If there still remained a systematic difference in the background characteristics of students in this restricted sample (based on an overall F-test), this exercise was repeated again based on the center-by-grade blocks in the restricted sample.

For Cohort 1, the most notable difference between the two program groups was in terms of their reading pretest score (which was lower on average in the enhanced program group). Thus, the difference in reading pretest scores between students in the enhanced and regular program group was calculated for each grade *j* within center *c*, and the 10 percent of blocks with the largest negative differences were dropped from the analysis (i.e., below the 10^{th} percentile). However, after dropping these blocks, there still remained a systematic difference in baseline characteristics between students in the enhanced and regular program group in the Cohort 1 sample, with the most notable difference now being in terms of the percentage of students with missing data on race/ethnicity (this percentage was larger in the enhanced program group). Hence, of the remaining grade-by-center blocks, the 10 percent of blocks with the largest between-group differences on this variable were excluded from the analysis (i.e., above the 90^{th} percentile).

For Cohort 2, the most notable difference at baseline between the two program groups was in terms of the percentage of students living in single-adult households (which was higher on average in the enhanced program group). Thus, the difference in the percentage of students in single-adult households between the enhanced and regular program group was calculated for each grade *j* within center *c*, and the 10 percent of blocks with the largest differences were dropped from the analysis (i.e., above the 90^{th} percentile). However, after dropping these blocks, there still remained a systematic difference in baseline characteristics between students in the enhanced and regular program group in the Cohort 2 sample, with the most notable difference

(continued)

group at baseline.[11] All impacts were therefore re-estimated using this restricted sample. As seen in Table G.8, the cohort-specific impact estimates based on the restricted sample are similar in magnitude to those presented in Chapter 8 (though the impact on total SAT 10 reading scores in the Cohort 1 sample is now statistically significant).[12] This similarity in the magnitude of impact estimates suggests that including the baseline characteristics of students in the impact model effectively controls for observed and unobserved differences between the two program groups at baseline.

now being in terms of the percentage of students with missing data on whether they receive free or reduced-price lunch (this percentage was larger in the enhanced program group). Hence, of the remaining grade-by-center blocks, the 10 percent of blocks with the largest between-group differences on this variable were excluded from the analysis (i.e., above the 90th percentile).

[11]Cohort 1 restricted sample: $F = 1.56$, $p = 0.07$; Cohort 2 restricted sample: $F = 1.33$, $p = 0.16$.

[12]Differences in impacts between cohorts are not statistically significant.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table G.8

Impact of the Enhanced Reading Program on Student Achievement Based on a Reading Analysis Sample That Excludes the Random Assignment Blocks with the Largest Between-Group Differences in Baseline Characteristics (One Year of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
Cohort 1[a]					
SAT 10 reading total scaled scores	588.71	592.66	-3.95 *	-0.12	0.01
Vocabulary	582.60	586.75	-4.15	-0.09	0.05
Reading comprehension	589.50	594.36	-4.86 *	-0.13	0.01
Word study skills (grades 2-4)[b]	589.64	591.29	-1.65	-0.04	0.55
DIBELS (grades 2-3)[c]					
Oral fluency score	75.17	72.67	2.50	0.08	0.30
Nonsense word fluency score	67.77	65.45	2.32	0.06	0.36
Sample size (total = 787)	441	346			
Cohort 2[d]					
SAT 10 reading total scaled scores	593.48	593.50	-0.02	0.00	0.99
Vocabulary	586.91	585.91	0.99	0.02	0.72
Reading comprehension	594.50	595.91	-1.42	-0.04	0.51
Word study skills (grades 2-4)[b]	593.29	591.83	1.46	0.04	0.67
DIBELS (grades 2-3)[c]					
Oral fluency score	78.91	76.28	2.63	0.08	0.33
Nonsense word fluency score	77.29	70.20	7.08 *	0.20	0.03
Sample size (total = 546)	306	240			

(continued)

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: The restricted sample excludes blocks with the largest differences in baseline characteristics between students in the enhanced and regular program groups. In Cohort 1, 6 center-by-grade blocks are excluded on the basis of differences in reading pretest scores and the percentage of students with missing data on race/ethnicity. In Cohort 2, 8 center-by-grade blocks are excluded on the basis of between-group differences in the percentage of students living in a single-adult household and missing data on free/reduced-price lunch status.

Students in the enhanced program group were assigned to one year of enhanced after-school services, while students in the regular program group were assigned to one year of the regular after-school program.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 374 to 787, 439 to 777, 412 to 739, and 410 to 740. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximum score; the maximum score is determined

Appendix Table G.8 (continued)

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 374 to 787, 439 to 777, 412 to 739, and 410 to 740. The DIBELS oral reading fluency and nonsense word fluency scores have a minimum score of zero, but no set maximum 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.

For both samples, the estimated effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98; nonsense = 36.13. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[a]Cohort 1 includes the students who were randomly assigned in the fall of the first year of the study.

[b]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.

[c]The DIBELS sample includes only second- and third-grade students because the nonsense word fluency subtest and the oral fluency subtest were not administered to fourth- and fifth-grade students in both study years.

[d]Cohort 2 includes the students who were randomly assigned in the fall of the second year of the study and were not offered the enhanced services in the first year of the study. Cohort 2 estimates are weighted to reflect the distribution of students across grades for all students who applied to the second year of the study and were randomly assigned in the fall of 2006.

Appendix H

Statistical Model and Sensitivity Analyses (Impact of Offering Two Years of Service)

This appendix describes the statistical model used to estimate the impact of being offered the opportunity to enroll in the enhanced after-school program *for two consecutive school years* and presents findings for additional impact analyses that were conducted to test the sensitivity of the results to sample and model specifications. As explained in Chapter 2, the two-year sample used for analysis includes both students who voluntarily applied to the second year of the study (applicants) and students from the first-year study sample who did not apply to the second year of the study (nonapplicants).

The first supplementary analysis examines impacts on locally administered standardized (state) tests. This outcome has policy relevance, given that scores on these tests are typically tied to rewards and sanctions in the local accountability system. An important issue to note here is that locally administered test data were not always available for second-graders in some study sites, since testing usually begins in the third grade. As a result, the impacts on state tests presented in this appendix are based on students in grades three to five only; impacts on the SAT 10 for this same subgroup of students are also presented for comparative purposes.

The second additional analysis presented in this appendix examines impacts for the SAT 10 respondent sample. The impact findings presented in the main body of the report are based on an analysis sampled restricted to students with spring follow-up data on both the SAT 10 assessment *and* the regular-school-day teacher survey. The latter restriction was imposed because measures of students' academic behavior are created from the teacher survey. As discussed in the report, however, the enhanced program did not affect students' academic behaviors. Hence, the second criteria for sample inclusion was dropped, and impacts on the SAT 10 were re-estimated based on *all* students that completed the SAT 10 assessment (whether or not they had teacher survey data).

The third additional analysis presented in this appendix is a sensitivity test of the impact findings to the chosen specification for the impact model. Specifically, the impact model includes student baseline covariates in order to explain random differences in the outcomes of students (and therefore improve the precision of the impact estimates). Strictly speaking, these covariates need not be included in the analysis because randomization creates the expectation that students in the enhanced and regular program groups are similar on average at baseline and that the difference in their outcomes can be attributed to the effects of the enhanced program. Rather, such covariates are typically included to increase the *precision* of the estimates. Hence, this appendix presents impacts from models that do not include these baseline covariates. (As will be explained in the reading section, this sensitivity analysis differs somewhat for the reading sample because randomization did not produce two statistically equivalent groups at baseline.)

Analysis of Program Impacts

Impacts on student outcomes are estimated for each of the two academic programs separately (the math or the reading program) by comparing the outcomes of students who were randomly assigned to the enhanced program in both years of the study (enhanced program group) and the outcomes of students assigned to the regular after-school program in both years (regular program group).

The Model

The impact of being assigned to (or offered the opportunity to participate in) the enhanced program for two consecutive school years is estimated for each outcome using the following statistical model:

$$Y_{ik} = \gamma_0 Y_{-1,ik} + \beta_0 T_{ik} + \sum_k \gamma_{1k} B_{ik} + \sum_S \gamma_{2s} X_{sik} + \varepsilon_{ik} \qquad (1)$$

where:

T_{ik} = Indicator of program group membership (treatment status). This indicator is equal to 1 if student i from random assignment block k was assigned to the enhanced program in both years of the study and zero otherwise

Y_{-1ik} = The pretest score for student i from random assignment block k before random assignment[1]

B_{ik} = Random assignment block indicators; equal to 1 if student i is in random assignment block k and zero otherwise[2]

X_{sik} = The set of s other student-level covariates for student i in random assignment block k

ε_{ik} = A student-level random error term assumed to be independently and identically distributed.

[1]Pretest scores are scaled scores from the SAT 10 tests in reading and math administered in the fall of 2005, before the start of the first year of the after-school program. Total scores for math are used in the math analysis, and total scores for reading are using in the reading analysis.

[2]Random assignment block is defined by students' grade j and center c at the start of the study (fall 2005). There are 46 random assignment blocks in the two-year sample for math and 34 blocks in the two-year sample for reading.

The coefficient, β_0, represents the overall impact of being randomized to the enhanced program for two consecutive years 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.

There are several features to note about this model:

- β_0 is a "fixed-effect" estimate that addresses the question: What is the effect of being assigned to the enhanced program for two consecutive years 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 squared (OLS) regression is used to estimate Equation (1).
- Indicators for random assignment blocks (B_{ik}) are included in the model to reflect the design feature (i.e., differential rates of treatment assignment by block) and to control for variation in mean outcome levels across blocks (which can be due to different characteristics of centers, school settings, etc.).
- The model controls for the student's pretest achievement score. This information can increase the precision of impact estimates, especially for fixed-effect models, because pretests substantially reduce within-block random error in the outcome measure, 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 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.

Other Analytical Issues

Missing Covariates

For the baseline achievement (fall 2005) test, there are two missing cases (none for math and two for reading). For other covariates, there are nine percent or fewer missing cases.[3]

[3]In the two-year sample for the math analysis, eight students are missing a race/ethnicity indicator, 14 are missing a free-lunch status indicator, three are missing information about single-adult household, and 32 are missing information about mother's education. In the two-year sample for the reading analysis, five are missing (continued)

To keep the sample as complete as possible, the missing values were imputed with the mean value of the random assignment block and program group[4] to which the student belongs.[5] 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.

Weighting of Nonreturning Students

As explained in Chapter 2, not all Cohort 1 students applied to the second year of the study. In order to preserve the experimental design of the study, all Cohort 1 students were randomly assigned in the second year. Then, consent for follow-up data collection (Spring 2007) was sought from nonapplicants. Consent was obtained from 57 percent of nonapplicants in the full two-year math study sample, and 51 percent of students in the full two-year reading study sample. This means that nonapplicants are under-represented in the two-year sample used for analysis relative to applicants (as consent for follow-up data collection was obtained from all applicants).

If not corrected, this under-representation of nonapplicants will produce two-year impact estimates that are too large (biased upwards). In order to understand why this happens, notice that the impact of being assigned to the enhanced program for two consecutive years is a combination of:

1. the impact of two years of enhanced services on students who *applied* in the second year, and

2. the impact of one year of enhanced services on *nonapplicants* (students who did not apply in the second year and, thus, did not actually receive a second year of enhanced services).

Hence, if nonapplicants are under-represented in the analysis sample relative to those who applied, the estimated impact of assignment to two years of enhanced services will be too large because it will not fully account for all nonapplicants who only received one year of enhanced services.

---

a race/ethnicity indicator, 10 are missing a free-lunch status indicator, three are missing information about single-adult household, and 19 are missing information about mother's education. (No students are missing gender or age.)

[4]In other words, the mean value for students in program group *p* (enhanced or regular) in grade *j* in center *c* in fall 2005 (i.e., start of the study).

[5]Rather than imputing the missing reading or math SAT 10 total scaled score, the mean raw score for the missing subtest was imputed and then the subtest raw scores were added to obtain an imputed total raw score. The student was then assigned the scaled score associated with their imputed total raw score. This was done so that — if there is an actual score for one or more of the subtests — the imputed total score incorporates that information.

In order to account for nonapplicants who *did not* consent to follow-up data collection, nonapplicants who *did* consent to follow-up data collection are given a proportionately greater weight in the analysis.[6] This weighting ensures that *nonapplicants* are not under-weighted relative to students who applied and that the estimated impact of offering students the opportunity to enroll in the enhanced program for two school years is unbiased.

Additional Analyses for the Math Sample

This section presents additional impact findings for the enhanced math program. The section begins with a discussion of impacts on locally administered math assessments. This is followed by a presentation of impacts on the SAT 10 respondent sample. The section concludes by examining impacts based on an alternate specification of the statistical model.

Impact on State Assessments

Table H.1 presents estimated program impacts on students' performance on locally administered math tests. Because these test scores were standardized within each study site, all estimated impacts are in effect size units.[7] Also, because not all students in the two-year analysis have local assessment data (the sample size decreases by eight students), the table also shows program impacts on the study-administered SAT 10 tests for this same sample of students, for comparative purposes.

As shown in this table, the impact of the enhanced math program on the locally administered math test for this particular sample of students is positive though not statistically significant (0.15 standard deviation, p-value = 0.09). Impacts on SAT 10 total math scores for this same sample of students are also not statistically significant (0.05 standard deviation, p = 0.52).

Impact on the SAT 10 Respondent Sample

Impacts on student achievement were re-estimated 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 two observations. Table H.2 presents

[6]Specifically, in each after-school center *c*, nonapplicants who *did* consent to follow-up data collection are weighted up to account for the nonapplicants in that center who *did not* consent to data collection. Weights are then normalized to sum to the actual two-year sample.

An overall F-test indicates that within after-school centers, nonapplicants who consented to data collection are not systematically different from nonapplicants who did *not* consent to data collection, whether in terms of their background characteristics or their treatment status (enhanced or regular program group) (F-test = 0.89, p-value = 0.59 for math; F-test = 0.96, p-value = 0.51for reading) . Thus, it is appropriate to weight the nonapplicants who *did* consent to data collection to account for nonapplicants who *did not* consent to data collection.

[7]Appendix F describes the standardization of the test score variable.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.1

Impact of the Enhanced Math Program on Student Achievement in the Math Analysis Sample for Grades 3 to 5 (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact Effect Size	P-Value for the Estimated Impact
State test scaled scores	0.07	-0.08	0.15	0.09
SAT 10 math total scaled scores	619.09	617.12	0.05	0.52
Sample size (total = 359)	222	137		

SOURCES: MDRC calculations are from results on state tests administered in the 2006-2007 school year and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Each student's state test score was converted into a standardized score because school districts in different states administer different tests. See Appendix F for details.

Based on the SAT 10 national norming sample, math total scaled scores range from 428 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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: SAT 10 = 38.90; state test = 1.13. The standard deviation for a SAT 10 national norming sample with the same grade composition is 38.99.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.2

Impact of the Enhanced Math Program on Student Achievement for the SAT 10 Respondent Sample (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
SAT 10 math total scaled scores	618.18	616.67	1.52	0.04	0.62
Problem solving	619.88	617.53	2.36	0.06	0.45
Procedures	617.19	616.98	0.21	0.00	0.96
Sample size (total = 369)	228	141			

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

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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: 428 to 796, 444 to 776, and 466 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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

impacts on SAT 10 math test scores for this SAT 10 respondent sample. As seen in the table, the magnitude of the estimates changes very little relative to what was presented in Chapter 5 of the report, and the patterns of statistical significance are the same.

Model Specification Tests

All impacts were re-estimated with a model that has no covariates other than the random assignment block indicators and the treatment status indicator (i.e., without student pretests and background characteristics):

$$Y_{ik} = \beta_0 T_{ik} + \sum_k \gamma_{1k} B_{ik} + \varepsilon_{ik} \qquad (2)$$

Because this study is based on a randomized experiment, both sets of impact estimates — those that are and are not adjusted for student characteristics — should provide similar estimates of the treatment effect.

As can be seen in Table H.3, dropping the student characteristics from the statistical model and only controlling for the randomization strata does not change the conclusion; the estimated impact of being assigned to the enhanced program for two years is not statistically significant.

Additional Analyses for the Reading Sample

This section presents additional impact findings for the enhanced reading program. The section begins with a discussion of impacts on locally administered reading assessments. This is followed by a presentation of impacts on the SAT 10 respondent sample. The section concludes by examining impacts based on alternate specifications of the statistical model.

Impact on State Assessments

Table H.4 presents estimated program impacts on students' performance on locally administered reading tests. Because these test scores were standardized within each study site, all estimated impacts are in effect size units.[8] Also, because not all students in the two-year analysis have local assessment data (the sample size decreases by 39 students), the table also shows program impacts on the study-administered SAT 10 tests for this same sample of students, for comparative purposes.

[8] Appendix F describes the standardization of the test score variable.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.3

Impact of the Enhanced Math Program on Student Achievement for the Analysis Sample, with Random Assignment Indicators as the Only Model Covariates (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
SAT 10 math total scaled scores	618.27	612.53	5.74	0.15	0.13
Problem solving	620.09	613.67	6.42	0.16	0.08
Procedures	617.10	612.31	4.79	0.09	0.35
Sample size (total = 367)	227	140			

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

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: 428 to 796, 444 to 776, and 466 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 mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.4

Impact of the Enhanced Reading Program on Student on Achievement in the Reading Analysis Sample for Grades 3 to 5 (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact Effect Size	P-Value for the Estimated Impact
State test scaled scores	-0.10	-0.16	0.06	0.60
SAT 10 reading total scaled scores	598.40	602.72	-0.13	0.15
Sample size (total = 231)	148	83		

SOURCES: MDRC calculations are from results on state tests administered in the 2006-2007 school year and follow-up results on the Stanford Achievement Test Series, 10th ed. (SAT 10) abbreviated battery.

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Each student's state test score was converted into a standardized score because school districts in different states administer different tests. See Appendix F for details.

Based on the SAT 10 national norming sample, reading total scaled scores range from 416 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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: SAT 10 = 33.19; state test = 1.16. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

As seen in this table, the impact of the enhanced reading program on the locally administered reading test for this particular sample of students is not statistically significant. However, unlike the direction of the impact on SAT 10 total reading scores, the estimated impact on state tests is positive.

Impacts for the SAT 10 Respondent Sample

Impacts on student achievement were re-estimated 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 one observation. Table H.5 presents impacts on SAT 10 reading test scores for the SAT 10 respondent sample. As seen in the table, the magnitude of the estimates changes very little relative to what was presented in Chapter 9 of the report, and the patterns of statistical significance are the same.

Model Specification Tests and Other Sensitivity Tests

As noted in the introduction, randomization ensures that students assigned to the enhanced and program group are similar on average at baseline within random assignment block. Hence, the purpose of including student covariates in the impact model is simply to improve the precision of the impact estimates (reduce the standard error).

However, in the reading sample, randomization did not produce two statistically equivalent groups at baseline (see Chapter 9). Most notably, students in enhanced program group had lower pretest scores on average than students in the regular program group. Hence, in this situation, it is important to control for student background characteristics in the impact model, especially student pretests. Otherwise, the analysis may produce biased estimates of the program's impact. The three sensitivity analyses presented in this section confirm that including student pretests and background characteristics in the model effectively controls for baseline differences between the enhanced and regular program groups.

No Covariates Other Than Block

As a first step, all impacts were re-estimated with a model that has no covariates other than the "block" (random assignment unit) indicators and the treatment status indicator (see equation 2).

As seen in Table H.6, dropping all student covariates from the impact model does not affect the statistical significance of the impact estimates (i.e., impacts on SAT 10 total scores and two of the subtests are statistically significant). However, the magnitude of the findings is larger in absolute terms (they become more negative). This happens because the enhanced

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.5

Impact of the Enhanced Reading Program on Student Achievement for the SAT 10 Respondent Sample (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
SAT 10 reading total scaled scores	596.09	601.66	-5.57 *	-0.17	0.04
Vocabulary	590.22	597.89	-7.67 *	-0.17	0.05
Reading comprehension	597.06	604.42	-7.37 *	-0.20	0.02
Word study skills (grades 2-4)[a]	594.16	595.47	-1.31	-0.03	0.80
DIBELS					
Oral fluency score	88.20	88.04	0.16	0.00	0.96
Sample size (total = 271)	170	101			

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

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

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: 416 to 787, 464 to 777, 455 to 739, and 450 to 740. The DIBELS oral reading fluency scores have a minimum score of zero, but no set maximum 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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[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 H.6

Impact of the Enhanced Reading Program on Student Achievement for the Analysis Sample, with Random Assignment Indicators as the Only Model Covariates (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
SAT 10 reading total scaled scores	595.99	607.29	-11.30 *	-0.34	0.01
Vocabulary	590.26	603.50	-13.24 *	-0.30	0.01
Reading comprehension	596.83	609.97	-13.14 *	-0.36	0.00
Word study skills (grades 2-4)[a]	594.16	604.04	-9.88	-0.24	0.09
DIBELS					
Oral fluency score	87.89	93.78	-5.89	-0.18	0.18
Sample size (total = 270)	169	101			

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: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 416 to 787, 464 to 777, 455 to 739, and 450 to 740. The DIBELS oral reading fluency scores have a minimum score of zero, but no set maximum 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 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 mean covariate values for the enhanced program group as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[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.

reading group was lower-achieving on average before the start of the program, and the impact model no longer controls for this difference in prior achievement. These results confirm the importance of controlling for student background characteristics in the model.

No Covariates Other Than Block and Pretest

Impacts were also re-estimated based on a model that includes *prior achievement* as a student covariate (the variable on which the two research groups differed the most at baseline) but that does not include the set of student demographic characteristics:

$$Y_{ik} = \gamma_0 Y_{-1,ik} + \beta_0 T_{ik} + \sum_k \gamma_{1k} B_{ik} + \varepsilon_{ik} \quad (3)$$

As can be seen from Table H.7, the impact estimates produced by this model are *not* substantially different in magnitude than those presented in Chapter 9. This suggests that controlling for students' pretest scores effectively adjusts for observed differences between the enhanced and regular program groups at baseline.

Exclusion of Blocks with Baseline Differences

While controlling for students' pretest scores appears to adjust for *observed* baseline differences between the two research groups, it may not control for *unobserved* differences between the two groups, in which case the impact findings would be biased.

An additional sensitivity test was conducted to explore this possibility. Specifically, the three random assignment blocks with the largest differences in pretest scores between students in the enhanced and regular program groups were dropped from the analysis.[9] In the remaining sample of students, it was found that there was no longer a systematic difference between students in the enhanced and regular program group at baseline.[10] All impacts were therefore re-estimated using this restricted sample (which is 93 percent of the two-year analysis sample.) As seen in Table H.8, in general, impact estimates based on the restricted sample are similar in magnitude to those presented in Chapter 9. This suggests that including the baseline characteristics of students in the impact model effectively controls for observed and unobserved differences between the two program groups at baseline.

[9] In the two-year analysis sample, the most notable difference between the two program groups was in terms of their reading pretest score (which was lower on average in the enhanced program group). Thus, the difference in reading pretest scores between students in the enhanced and regular program group was calculated for each random assignment block (grade *j* within center *c*), and the 10 percent of blocks with the largest negative differences were dropped from the analysis (i.e., below the 10th percentile).

[10] F = 1.38, p-value = 0.14.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table H.7

Impact of the Enhanced Reading Program on Student Achievement for the Analysis Sample, Without Demographic Characteristics as Model Covariates (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
SAT 10 reading total scaled scores	595.99	601.74	-5.76 *	-0.17	0.05
Vocabulary	590.26	597.67	-7.41	-0.17	0.07
Reading comprehension	596.83	604.58	-7.74 *	-0.21	0.02
Word study skills (grades 2-4)[a]	594.16	597.08	-2.92	-0.07	0.57
DIBELS					
Oral fluency score	87.89	88.22	-0.33	-0.01	0.92
Sample size (total = 270)	169	101			

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: The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 416 to 787, 464 to 777, 455 to 739, and 450 to 740. The DIBELS oral reading fluency scores have a minimum score of zero, but no set maximum 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 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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[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 H.8

Impact of the Enhanced Reading Program on Student Achievement Based on a Reading Analysis Sample That Excludes the Random Assignment Blocks with the Largest Between-Group Differences in Baseline Characteristics (Offer of Two Years of Service)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	Estimated P-Value for the Impact
SAT 10 reading total scaled scores	597.05	603.31	-6.26 *	-0.19	0.05
Vocabulary	591.54	599.48	-7.94	-0.18	0.06
Reading comprehension	597.96	606.98	-9.02 *	-0.25	0.01
Word study skills (grades 2-4)[a]	595.23	595.94	-0.71	-0.02	0.90
DIBELS					
Oral fluency score	90.36	91.75	-1.38	-0.04	0.69
Sample size (total = 251)	156	95			

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: The restricted analysis sample excludes the 3 center-by-grade random assignment blocks with the largest differences in reading pretest scores at baseline (9 percent of blocks are excluded).

The enhanced program group includes students who were assigned to the enhanced after-school program in both years of the study. The regular program group includes students who were assigned to the regular after-school program in both years.

Based on the SAT 10 national norming sample, total, reading comprehension, vocabulary, and word study skills scaled scores, respectively, have the following possible ranges: 416 to 787, 464 to 777, 455 to 739, and 450 to 740. The DIBELS oral reading fluency scores have a minimum score of zero, but no set maximum 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.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the two-year sample regular program group. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50; word study skills = 41.65; oral fluency = 32.98. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

[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 I

Exploratory Analysis: The Association Between Receiving Two Years of Enhanced After-School Academic Instruction and Student Achievement

This appendix provides details on the analytical strategy used to estimate the association between receiving two years of enhanced after-school services and student outcomes. As explained in the report, not all students assigned to the enhanced program for two consecutive school years applied both years; some were nonapplicants, students who did not actually apply to the study in the second year.[1] Thus, the exploratory analysis in Chapter 6 (math) and Chapter 10 (reading) examine the association between receiving two years of enhanced services and student achievement, based on an instrumental variables analysis.

Instrumental Variables Analysis

The results presented in the report are based on an instrumental variables (IV) analysis, in which the number of years of enhanced after-school services received by students is instrumented using indicators of random assignment to treatment conditions. The analysis is based on students in three of the study's experimental groups (see Figure I.1):[2]

- **E_1E_2 group:** Students assigned to two years of enhanced services
- **R_1R_2 group:** Students assigned to two years of regular services
- **E_1R_2 group:** Students assigned to enhanced services in the first year of the study but not the second

As will be explained below, the latter group of students is included in the IV analysis because they provide an approximation of what happened to nonapplicants in the E_1E_2 group in the second year of the study (i.e., they received enhanced services in the first year but not the second).

Notice that students in these three groups received one of three possible amounts of "dosage": two years of enhanced services (i.e., applicants in the E_1E_2 group); one year of enhanced services (i.e., nonappliants in the E_1E_2 group as well as all students in the E_1R_2 group); and zero years of enhanced services (i.e., students in the R_1R_2 group).

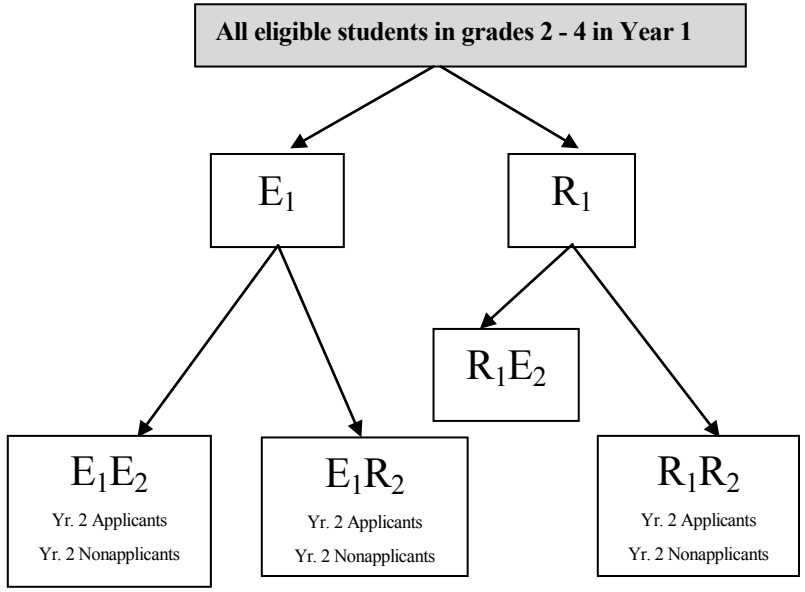
The statistical model used for the IV analysis — as well as the conditions and assumptions that underlie the analysis — are described in greater detail below.

[1]Specifically, in the math and reading analysis, respectively, 42 percent and 43 percent of students assigned to two years of the enhanced program did not apply in the second year and, therefore, did not receive a second year of enhanced after-school services.

[2]These groups include both applicants and nonapplicants.

Appendix Figure I.1

Sample Used to Estimate the Association Between Receiving Two Years of Enhanced After-School Services and Student Achievement (Instrumental Variables Analysis)



Sample sizes for Instrumental Variables sample		
	Math	Reading
E_1E_2	227	169
E_1R_2	167	138
R_1R_2	140	101

E_1 = Enhanced program group, Year 1
R_1 = Regular program group, Year 1
E_2 = Enhanced program group, Year 2
R_2 = Regular program group, Year 2

NOTES: The sample used in the analysis is limited to students with two-year follow-up data from both the evaluation-administered achievement test and the regular-school-day teacher survey.

[a]This sample includes the two-year instrumental variables sample: students from year 1 who applied to the second year of the study (applicants) as well as students from year 1 who had participated in the first year of the study, but did not apply to the second year of the study (nonapplicants). Random assignment was conducted separately for these two groups, stratified by grade and first year treatment status (that is, the enhanced program or regular program) within each after-school center. Test and survey data were collected at the end of Year 2.

Statistical Models

Student Achievement

The key research question that drives this analysis is whether receiving two years of enhanced after-school services affects student achievement. Thus, one might consider fitting the following statistical model using ordinary least squares (OLS):

$$Y_{ik} = \gamma_0 Y_{-1,ik} + \beta_1 REC2_{ik} + \beta_2 REC1_{ik} + \sum_k \gamma_{1k} B_{ik} + \sum_S \gamma_{2s} X_{sik} + \varepsilon \quad (1)$$

where:

Y_{ik} = Achievement for student i from random assignment block k (i.e., SAT 10 score).

$REC2_{ik}$ = Dummy variable equal to 1 if student i from random assignment block k received two years of enhanced after-school services (i.e., applicants in the E_1E_2 group), and zero otherwise.

$REC1_{ik}$ = Dummy variable equal to 1 if student i from random assignment block k received one year of enhanced after-school services (i.e., nonapplicants in the E_1E_2 group and all students in E_1R_2 group), and zero otherwise.

$Y_{-1,ik}$ = The pretest score for student i from random assignment block k before random assignment.

B_{ik} = Block dummy variable, equal to 1 if student i is in random assignment block k, and zero otherwise.[3]

$\sum_s X_{sik}$ = The set of s other student-level covariates for student i in random assignment block k.

ε = A student-level random error, assumed to be independently and identically distributed.

Notice that the impact of receiving two years of enhanced services is represented by β_1 in this model.

[3]Random assignment block is defined by students' grade j and center c at the start of the study (fall 2005). There are 46 random assignment blocks in the two-year sample for math and 34 blocks in the two-year sample for reading.

The problem with this approach, however, is that the number of years of enhanced services that students receive (***REC2***, ***REC1***) could be related to their experience in the enhanced program in the first year of the study. For example, students who chose to receive enhanced services for two school years (i.e., applicants in the E_1E_2 group) may be those who felt that they particularly benefited from the enhanced program in the first year. Conversely, students who chose to receive only one year of enhanced services (i.e., nonapplicants in the E_1E_2 group) could be students who felt that they did not benefit at all from the enhanced program in the first year. In other words, students *self-select* themselves into different amounts of enhanced instruction. As a result of this self-selection, students in the R_1R_2 group (who did *not* receive enhanced services) may no longer provide the right counterfactual for what would have happened to students who received two years (or one year) of enhanced services in the absence of the enhanced program. Nor is it possible to identify which students in the R_1R_2 group would have made similar participation decisions had they been invited to enroll in the enhanced after-school program in the first year.

Technically, this means that ***REC2*** and ***REC1*** (the amount of enhanced services received by students) in Equation (1) are *endogenous*, i.e., ***REC2*** and ***REC1*** could be correlated with unobserved student characteristics in the error term that are also associated with student outcomes. Thus, if Equation (1) were estimated using ordinary least squares (OLS), it could potentially produce a biased estimate of β_1 (the estimated impact of receiving two years of enhanced services).

A solution to this endogeneity problem is to use two-stage least squares (2SLS) to fit Equation (1). In the first stage, indicators of random assignment to the E_1E_2 group and to the E_1R_2 group are used as instrumental variables for the number of years that students received enhanced services (***REC2*** and ***REC1***), and each of the following first-stage equations is estimated:

$$REC2_{ik} = \tau_0 Y_{-1,ik} + \lambda_1 EE_{ik} + \lambda_2 ER_{ik} + \sum_k \gamma_{1k} B_{ik} + \sum_S \gamma_{2s} X_{sik} + \varepsilon' \qquad (2a)$$

$$REC1_{ik} = \delta_0 Y_{-1,ik} + \phi_1 EE_{ik} + \phi_2 ER_{ik} + \sum_k \gamma_{1k} B_{ik} + \sum_S \gamma_{2s} X_{sik} + \varepsilon' \qquad (2b)$$

where the instruments are:

EE_{ik} = Dummy variable equal to 1 if student i from random assignment block k was assigned to the enhanced program in both years of the study (E_1E_2 group), and zero otherwise.

ER_{ik} = Dummy variable equal to 1 if student *i* from random assignment block *k* was assigned to the enhanced program in the first year of the study but not the second), and zero otherwise.

In the second stage, Equation (1) is estimated but using as covariates the *predicted* values of ***REC1*** and ***REC2*** from the first-stage equations, rather than the observed values of ***REC2*** and ***REC1***. The resulting estimate of β_1 is the 2SLS estimate of the association between receiving two years of enhanced services and student achievement. This estimate is unbiased, provided that each instrument (***EE*** and ***ER***) has a unique relationship with each endogenous variable (***REC2*** and ***REC1***).[4] This condition is satisfied in this particular context, given the way in which random assignment was conducted.[5]

Hours of Academic Instruction

A simplified version of this model was also used to estimate the association between receiving two years of enhanced services and *hours of academic instruction* in the second year of the study (results are reported in footnotes in Chapter 6 and Chapter 10). For this outcome, the IV analysis is based on students in the E_1E_2 and R_1R_2 groups. In the first stage, the following model is estimated:

$$REC2_{ik} = \tau_0 Y_{-1,ik} + \gamma_1 EE_{ik} + \sum_k \delta_k B_{ik} + \sum_S \theta_s X_{sik} + \varepsilon \qquad (3)$$

The predicted values from this first stage model are then used to estimate the second stage:

$$Y_{ik} = \alpha_0 Y_{-1,ik} + \beta_1 \widehat{REC2}_{ik} + \sum_k \delta_k B_{ik} + \sum_S \theta_s X_{sik} + \varepsilon \qquad (4)$$

where:

Y_{ik} = The hours of after-school instruction received by a student in the *second year* of the study.

The resulting estimate of β_1 is the 2SLS estimate of the association between receiving two years of enhanced services and hours of after-school academic instruction in the *second*

[4]In other words, the condition is that $\lambda_1 / \lambda_2 \neq \gamma_1 / \phi$ (see Gennetian et al., 2005, for details on IV estimation with multiple endogenous variables and instruments).

[5]Specifically, assignment to two years of enhanced services (EE) is a relatively stronger predictor of receiving two years of enhanced services (***REC2***), while assignment to one year of enhanced services (ER) is a relatively stronger predictor of receiving one year of enhanced services (***REC1***).

year. In order to obtain the estimated association between receiving two years of enhanced services and the cumulative number of hours of instruction received across *both years* of the study, the estimate of β_1 from Equation 4 is added to the estimated between-group difference in instructional hours in the *first year* of the study (see Tables 5.2 and 9.2).

Conditions and Assumptions

In randomized experiments, it is often the case that some individuals assigned to the treatment group do not "take up" the treatment or program that is offered to them. These individuals are called "no-shows" in the program evaluation literature (Gennetian et al., 2005; Angrist et al., 1996).[6] When there are no-shows, the difference in outcomes between individuals in the treatment and control group provides an estimate of the impact of the "intent to treat" (ITT) with the program, rather than the impact of *receiving* the program (also called the impact of the "treatment on the treated").

While the impact of the treatment on the treated (TOT) cannot be estimated experimentally, several random assignment studies have made use of instrumental variables analysis to estimate the impact of the program on individuals who actually receive it. In the Moving to Opportunity (MTO) demonstration,[7] for example, housing vouchers of different types were randomly assigned to families in order to enable them to relocate to higher-income neighborhoods, which made it possible to estimate the impact on families of being assigned a housing voucher (i.e., the impact of the "intent to treat" with a housing voucher). However, not all families who received a housing voucher actually used it to relocate to another neighborhood. Thus, to estimate the impact on households of actually relocating to a higher-income neighborhood, Kling and others (2007) used individuals' treatment group status as an instrumental variable for whether or not they relocated.

In the context of experimental studies, IV analysis is a useful analytical tool for estimating the "treatment on the treated" because random assignment status (i.e., whether or not an individual is assigned to a treatment) meets all three conditions for an instrumental variable. First, it is correlated with receiving the treatment; second, it only affects student outcomes through *receipt* of the treatment; and, third, it is uncorrelated with individuals' unobserved

[6]Another type of noncompliance occurs when individuals assigned to the control group find a means of gaining admittance to the program (these individuals are typically called "cross-overs"). The latter form of noncompliance is not relevant to this analysis, so the term "non-compliers" in this appendix refers to "no-shows" only.

[7]Moving to Opportunity is a ten-year demonstration funded by the U.S. Department of Housing and Urban Development (HUD). Five public housing authorities (Baltimore, Boston, Chicago, Los Angeles, and New York City) administer HUD contracts under this demonstration. See http://www.hud.gov/progdesc/mto.cfm for more information.

characteristics (by definition). In the present study, for example, ***EE*** and ***ER*** (in Equations 2a and 2b) meet all three of these conditions.

Yet, it is also important to note that IV analysis is based on assumptions that may or may not be true depending on the context. In order to understand these assumptions, notice first that the impact of being *assigned to* a particular treatment or program can be decomposed into two types of impact:[8]

1. The impact of the program on the "compliers" (i.e., individuals in the treatment group who received the assigned treatment, who in this study are the *applicants* in the E_1E_2 enhanced program group)

2. The impact of the program on the "non-compliers" or "no-shows" (i.e., individuals in the treatment group who did not receive the assigned treatment, who in this study are *nonapplicants* in the E_1E_2 enhanced program group).

Based on this decomposition, one can see that if the impact of the program on the *no-shows* (second component) were known, then it would be possible to isolate the impact of the program on individuals who complied with random assignment and received the treatment (first component). Because the impact on the no-shows is not known with certainty, the IV approach makes assumptions about its magnitude based on the study design, which may or may not be accurate depending on the exact nature of the study design.

In a "simple" experimental study design, for example — in which individuals are assigned to a treatment or program only once — the IV approach assumes that the impact of the program on no-shows in the treatment group is zero. This is a reasonable assumption, given that non-compliers were not exposed to the program such that it could not have affected their outcomes. In this context, the IV approach produces a consistent estimate of the impact of receiving the treatment (Angrist et al., 1996; Gennetian et al., 2005).[9] This is the assumption that underlies the IV analysis that was used as part of the MTO evaluation.

In the present study, however, the random assignment design is less straightforward, because the "no-shows" (i.e., nonapplicants) in the enhanced program group received one year of the program rather than none at all, and the impact of these students' first year in the program may still exist at the end of the second year. Thus, it cannot be assumed that the enhanced program had no impact on these students. In this situation, rather than assuming a "zero impact," the IV approach exploits the experimental design of the study to estimate the impact of the

[8]This decomposition assumes that all non-compliers in the study are members of the treatment group (i.e., no control group members received treatment).

[9]Or, more specifically, it provides a consistent estimate of the local average treatment effect (LATE).

program on non-compliers (nonapplicants). As explained in Chapter 2, one group of students in the second-year study design was assigned to a treatment condition whereby they received enhanced services in the first year of the study but not the second (these students are the E_1R_2 group in Figure I.1), which is the same "level" of treatment that was ultimately received by nonapplicants in the enhanced program group. By comparing the outcomes of students in the E_1R_2 group to those of students assigned to the regular program in both years of the study (the R_1R_2 group in Figure I.1), one can obtain an internally valid estimate of the impact of receiving enhanced services in the first year of the study but not the second (see Tables I.1 and I.2 for these findings, for the math and reading samples, respectively). The IV approach as specified in Equations (1) and (2) utilizes these findings as an estimate of what happened to nonapplicants ("no-shows") in the second year of the study. Based on the assumption that these estimates are credible — and by using the "impact decomposition" above — the IV approach is able to estimate the impact of receiving two years of enhanced after-school services.[10]

The key limitation of the IV approach in this context, however, is that its underlying assumption about what happened to the nonapplicants may not be correct. This is because the majority of students in the E_1R_2 experimental group are students who *applied* in the second year,[11] who may be fundamentally different than those who did not apply in terms of their intrinsic motivation. In particular, nonapplicants in the enhanced program group *chose* to not reapply for the study in the second year and, therefore, had no interest in receiving a second year of enhanced services. Conversely, those who applied in the E_1R_2 experimental group were assigned to this condition but would have participated in a second year of the program had they been offered the opportunity to do so. Stated otherwise, the second year of enhanced services was withheld from applicants in the E_1R_2 group, while it was essentially rejected by the nonapplicants in the enhanced program group. In addition to likely having different types of intrinsic motivation, applicants and nonapplicants also differ in terms of their observed characteristics at baseline, as indicated by an overall F-test (for math, F-test = 3.06, p-value = 0.00; for reading, F-test = 3.49, p-value = 0.00). (See Appendix Tables I.3 and I.4 for a comparison of the baseline characteristics of these students in the two-year math and reading analysis samples, respectively.)

---

[10]Specifically, the IV (or 2SLS) estimate of the association between the receiving two years of enhanced services and student outcomes is: $\hat{\beta}_1 = \frac{\phi_2}{(\phi_2\lambda_1 - \phi_1\lambda_2)} EEvRR - \frac{\phi_1}{(\phi_2\lambda_1 - \phi_1\lambda_2)} ERvRR$

where:

EEvRR = Impact on achievement of assigning student to the enhanced program for two school years (E_1E_2 group vs. R_1R_2 group, see Tables 5.3 and 9.3).

ERvRR = Impact on achievement of assigning students the enhanced program in the first year of the study but not the second (E_1R_2 group vs. R_1R_2 group, see Tables I.1 and I.2).

[11]In the two-year math analysis sample, 59 percent of students in the ER group are applicants; in the two-year reading analysis sample, 60 percent are applicants.

As a result of these differences in sample composition, the estimated impact for the E_1E_2 group of having received enhanced services in the first year of the study but not the second — though internally valid for this group of students — may be a biased estimate of such an impact for nonapplicants. If so, then the IV approach will produce a biased estimate of the impact of receiving two years of enhanced services. Hence, the IV findings presented in the report are characterized as the *association* between receiving two years of enhanced after-school services and student achievement, rather than the impact of receiving two years of services.[12]

In terms of the IV analysis for *hours of academic instruction* (Equations 3 and 4), the noncompliance or IV adjustment is more straightforward because the "zero impact" assumption can be made for this outcome measure.[13] As explained earlier, the IV analysis focuses on instructional hours *in the second year only*. Specifically, the association between receiving two years of services and instructional hours in the *second* year of the study is estimated, based on the plausible assumption that the enhanced program had no impact on the amount of after-school instruction received by nonapplicants in the enhanced program group.[14] This IV estimate is then added to the service contrast for these students in the *first year* of the study in order to obtain the association between receiving two years of service and instructional hours *across both study years*.

Note this approach of adding the experimentally determined first-year estimate to a second-year impact estimate generated by confining the IV analysis to the second year only cannot be used for student achievement measures such as the SAT 10. Unlike the first-year hours of instruction received, which remain constant over time, the impact of students' first year of enrollment in the enhanced program may decrease or decay during the second year if they do not return to the program. Thus, the IV adjustment must use information from both years of the study and must rely on a more sophisticated assumption about the estimated impact on nonapplicants at the end of the second year of receiving enhanced services in the first year of the study but not the second, as described earlier in this section.

---

[12]Note that one disadvantage of using an instrumental variables approach is that it produces inefficient estimates. The standard errors of instrumental variables estimates are scaled up by (approximately) the proportion of individuals in the treatment group who received the treatment that they were assigned (in this case, the proportion of applicants in the enhanced program group).

[13]The IV (or 2SLS) estimate of the association between the receiving two years of enhanced services and hours of instruction in the second year is: $\hat{\beta}_1 = \frac{1}{\lambda_1} EEvRR$

where:

EEvRR = Impact of assigning student to the enhanced program for two school years on hours of instruction (E_1E_2 group vs. R_1R_2 group, see Tables 5.2 and 9.2).

[14]In other words, it can be assumed that nonapplicants in the enhanced program group received the same amount of after-school instruction in the second year of the study as did their counterparts in the regular program group.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table I.1

Impact of the Enhanced Math Program on Student Achievement (Service in the First Year but Not the Second)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
SAT 10 math total scaled scores	617.37	617.55	-0.18	0.00	0.96
Problem solving	621.10	618.99	2.10	0.05	0.52
Procedures	612.96	617.03	-4.07	-0.08	0.36
Sample size (total = 307)	167	140			

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

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in the first year of the study but not the second. The regular program group includes students who were assigned to the regular after-school program in both years.

Based on the SAT 10 national norming sample, total, problem solving, and procedures scaled scores, respectively, have the following possible ranges: 428 to 796, 444 to 776, and 466 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 means for students who were assigned to the enhanced program in the first year of the study but not the second. The values in column 2 (labeled "Regular Program") are the regression-adjusted means using the observed covariate values of the enhanced program group as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the regular program group. These standard deviations are: total score = 38.90; problem solving = 40.08; procedures = 51.79. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 38.99.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table I.2

Impact of the Enhanced Reading Program on Student Achievement (Service in the First Year but Not the Second)

Student Achievement Outcome	Enhanced Program	Regular Program	Estimated Impact	Estimated Impact Effect Size	P-Value for the Estimated Impact
SAT 10 reading total scaled scores	600.03	602.24	-2.21	-0.07	0.45
Vocabulary	596.60	597.07	-0.46	-0.01	0.92
Reading comprehension	600.45	605.39	-4.94	-0.14	0.14
Sample size (total = 239)	138	101			

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

NOTES: The enhanced program group includes students who were assigned to the enhanced after-school program in the first year of the study but not the second. The regular program group includes students who were assigned to the regular after-school program in both years.

Based on the SAT 10 national norming sample, total, reading comprehension, and vocabulary scaled scores, respectively, have the following possible ranges: 416 to 787, 464 to 777, 455 to 739.

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 means for students who were assigned to the enhanced program in the first year of the study but not the second. The values in column 2 (labeled "Regular Program") are the regression-adjusted means using the observed covariate values of the enhanced program group as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 for students in the regular program group. These standard deviations are: total score = 33.19; vocabulary = 44.63; reading comprehension = 36.50. The standard deviation in the total score for a SAT 10 national norming sample with the same grade composition is 39.08.

The Evaluation of Academic Instruction in After-School Programs

Appendix Table I.3

Baseline Characteristics of Student Applicants and Nonapplicants in the Math Analysis Sample (Offer of Two Years of Service)

Characteristic	Full Sample	Applicants	Non-Applicants	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Enrollment						
2nd grade	121	83	38			
3rd grade	134	95	39			
4th grade	112	84	28			
Total	367	262	105			
Race/ethnicity (%)						
Hispanic		30.53	33.71	-3.17	-0.06	0.54
Black, non-Hispanic		40.08	32.16	7.91	0.15	0.10
White, non-Hispanic		21.18	27.03	-5.85	-0.12	0.23
Other		5.49	7.49	-2.00	-0.08	0.44
Gender (%)						
Male		48.47	49.31	-0.84	-0.02	0.89
Average age (years)		8.12	8.18	-0.06	-0.10	0.38
Overage for grade[a] (%)		14.89	20.30	-5.42	-0.13	0.21
Free/reduced-price lunch (%)						
Eligible (among information providers)		77.09	78.96	-1.87	-0.04	0.66
No information provided		4.20	2.12	2.08	0.12	0.38
Average household size		1.95	1.88	0.07	0.07	0.54
Single-adult household (%)		33.43	38.28	-4.85	-0.09	0.40
Mother's education level (%)						
Did not finish high school		16.03	20.50	-4.47	-0.11	0.36
High school diploma or GED certificate		27.48	32.97	-5.49	-0.11	0.32
Some postsecondary study		47.71	36.30	11.41 *	0.21	0.05
No information provided		8.78	10.23	-1.45	-0.06	0.61
SAT 10 baseline math total scaled scores		553.29	552.24	1.06	0.03	0.76
Problem solving		559.14	558.60	0.54	0.01	0.89
Procedures		546.03	543.54	2.49	0.05	0.59
Sample size (total = 367)		262	105			

(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: Applicants applied for the opportunity to be assigned to the enhanced program in both years of the study. Nonapplicants applied for the enhanced program in the first year of the study but not in the second year.

Appendix Table I.3 (continued)

NOTES: Applicants applied for the opportunity to be assigned to the enhanced program in both years of the study. Nonapplicants applied for the enhanced program in the first year of the study but not in the second year.

The estimated differences are regression-adjusted using ordinary least squares, controlling for indicators of random assignment strata. The values in the column labeled "Applicants" are the observed mean for students who applied to the study in the second year. The "Nonapplicants" values in the next column are the regression-adjusted means using the observed distribution of applicants across random assignment strata as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined.

An F-test was calculated 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-value (F = 3.06) is significant at the 5 percent level.

[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 I.4

Baseline Characteristics of Student Applicants and Nonapplicants in the Reading Analysis Sample (Offer of Two Years of Service)

Characteristic	Full Sample	Applicants	Non-Applicants	Estimated Difference	Estimated Difference Effect Size	P-Value for the Estimated Difference
Enrollment						
2nd grade	100	76	24			
3rd grade	87	60	27			
4th grade	83	59	24			
Total	270	195	75			
Race/ethnicity (%)						
Hispanic		43.48	38.63	4.85	0.09	0.41
Black, non-Hispanic		38.55	36.07	2.47	0.05	0.61
White, non-Hispanic		9.38	19.20	-9.82 *	-0.28	0.02
Other		8.33	7.75	0.59	0.03	0.84
Gender (%)						
Male		49.74	55.82	-6.07	-0.11	0.38
Average age (years)		7.93	8.04	-0.11	-0.19	0.12
Overage for grade[a] (%)		10.26	18.42	-8.16	-0.22	0.12
Free/reduced-price lunch (%)						
Eligible (among information providers)		84.36	80.75	3.61	0.09	0.35
No information provided		3.59	5.43	-1.84	-0.09	0.52
Average household size		2.09	2.14	-0.05	-0.04	0.77
Single-adult household (%)		26.30	30.04	-3.74	-0.08	0.54
Mother's education level (%)						
Did not finish high school		28.21	26.00	2.21	0.05	0.73
High school diploma or GED certificate		31.28	22.79	8.49	0.17	0.16
Some postsecondary study		41.54	34.86	6.68	0.12	0.31
No information provided		4.10	14.12	-10.02 *	-0.38	0.03
SAT 10 baseline reading total scaled scores		556.13	548.88	7.25	0.22	0.08
Vocabulary/word reading[b]		548.66	540.34	8.32	0.19	0.14
Reading comprehension		554.77	551.72	3.04	0.08	0.52
Word study skills[c]		569.71	555.52	14.19 *	0.35	0.00
Sample size (total = 270)		195	75			

(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: Applicants applied for the opportunity to be assigned to the enhanced program in both years of the study. Nonapplicants applied for the enhanced program in the first year of the study but not in the second year.

Appendix Table I.4 (continued)

NOTES: Applicants applied for the opportunity to be assigned to the enhanced program in both years of the study. Nonapplicants applied for the enhanced program in the first year of the study but not in the second year.

The estimated differences are regression-adjusted using ordinary least squares, controlling for indicators of random assignment strata. The values in the column labeled "Applicants" are the observed mean for students who applied to the study in the second year. The “Nonapplicants” values in the next column are the regression-adjusted means using the observed distribution of applicants across random assignment strata as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

Among those who did not reapply to the study in the second year, nonresponse weights are used to account for those students for whom follow-up data were not collected.

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 effect size for each characteristic is calculated as a proportion of the standard deviation for students in the regular program group in both cohorts combined.

An F-test was calculated 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. The F-value (F = 3.49) is significant at the 5 percent level.

[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 vocatublary subtest.

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

Appendix J

Exploratory Analysis: Linking the Impact of One Year of Enhanced Services on Student Achievement with School and Program Characteristics

This appendix describes the strategy used to relate the size of a center's impact on the SAT 10 with the characteristics of (1) the schools housing the after-school program and (2) the implementation of the enhanced after-school program. The analysis was conducted by estimating the association between school-level characteristics and program impacts on SAT 10 total scores in each of the participating after-school centers in each study year (i.e., 30 center-level impacts in the math sample and 24 center-level impacts in the reading sample).[1]

This appendix begins by discussing the extent of variation in impacts across after-school centers and implementation years. This is followed by a description of the school-level characteristics and of the statistical model used in the exploratory analysis.

Variation in Impacts

Appendix Figure J.1 presents the average impact of the enhanced math program across both years of the study, as well as the distribution of impacts across after-school centers and study years.[2] Figure J.2 presents similar findings for the enhanced reading program. For each center and for the overall average, the figure displays impact estimates (represented by the circles for impacts in Year 1 and by triangles for impacts in Year 2) and the 95 percent confidence interval around the impact estimates (represented by the lines extending above and below the circles and triangles). Hence, the wider the confidence interval, the broader the margin of error and the greater the uncertainty around the impact estimate. Confidence intervals that do not include zero are statistically significant (p-value is less than or equal to 5 percent).

Math Centers

As seen in Figure J.1, the center-by-year impact estimates for math range from -10.1 scaled score points to 18.8 scaled score points. In all, 20 of the 30 center-level impact estimates are above zero (10 in Year 1 and 10 in Year 2), and 10 of the 30 are negative (five in Year 1 and five in Year 2). Five of the impact estimates (all positive) are statistically significant at the 5 percent level.

The variation in estimated impacts displayed in Figure J.1 overstates the true variation in impacts, however, because a large portion of the variation in estimated impacts is due to estimation error. To examine variability in impacts across centers and study years more systematically, a composite F-test was used to assess whether the center-level impacts in Figure J.1

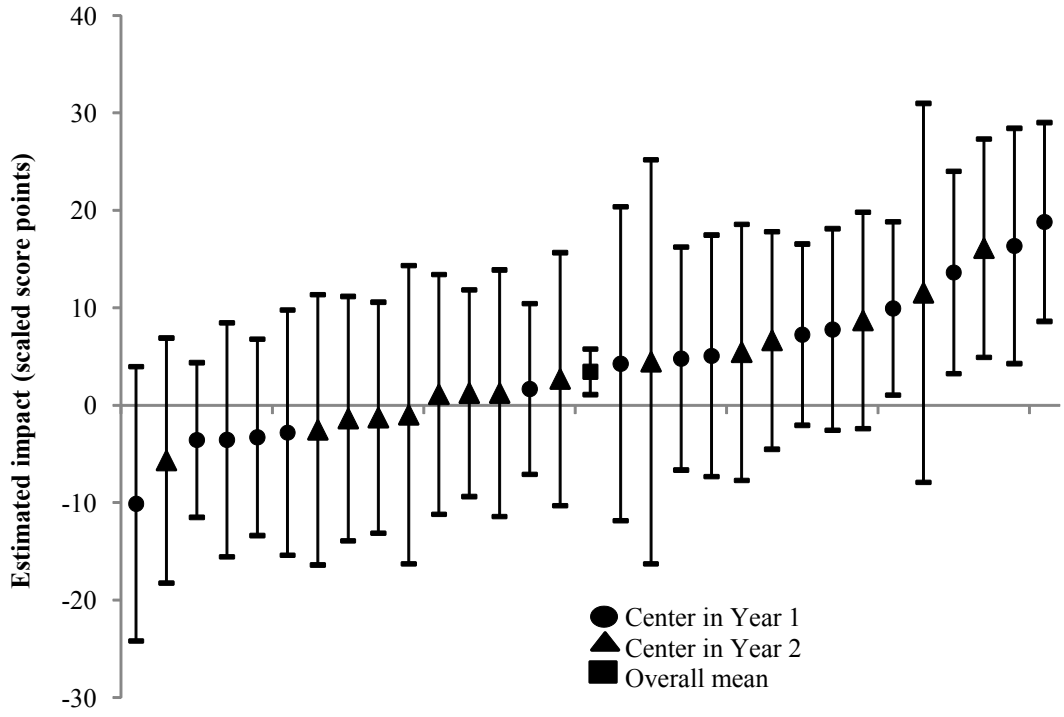
[1]For math, 15 centers * 2 implementation years = 30 center-level impacts; for reading, 12 centers * 2 implementation years = 24 center-level impacts.

[2]Center-level impacts were estimated by interacting the treatment indicator in the impact model with center-by-year dummies (30 for math and 24 for reading).

The Evaluation of Academic Instruction in After-School Programs

Appendix Figure J.1

Impact of One Year of the Enhanced Math Program on Student Achievement and Its Distribution Across Centers and Implementation Years



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

NOTES: Confidence intervals are based on a statistical significance level of 5 percent.

The figure shows the estimated program impact for the student-level analysis sample on students' SAT 10 total math scores (the square; p-value = 0.004) and how that impact is distributed across the 30 center-by-year estimates in the analysis sample (each circle or triangle). The center-by-year impacts (presented ordinally) are estimated by interacting the treatment indicator with center and implementation-year 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.

are statistically equivalent to each other; this test accounts for estimation error in center-level impacts and provides an indication of the confidence one might have that there is true variation in impacts across the centers. This test indicates that the overall variation in impacts across centers and implementation years is not statistically significant at the 5 percent level (p-value = 0.07).[3] This indicates that the overall variation in center-level impacts could be due to chance.

Reading Centers

As seen in Figure J.2, the center-by-year impact estimates for the reading program range from -11.1 scaled score points to 7.2 scaled score points. In all, 11 of the 24 center-level impact estimates are above zero (four in Year 1 and seven in Year 2), and 13 of the 24 impact estimates are negative (eight in Year 1 and five in Year 2). One of the negative impact estimates is statistically significant at the 5 percent level.

To examine variability in impacts across centers and study years more systematically, a composite F-test was used to assess whether the center-level impacts in Figure J.2 are statistically equivalent to each other; this test accounts for estimation error in center-level impacts and provides an indication of the confidence one might have that there is true variation in impacts across the centers. This test indicates that the overall variation in impacts across centers and implementation years is not statistically significant at the 5 percent level (p-value = 0.55).[4] This indicates that overall variation in center-level impacts could be due to chance.

Nonetheless, the relationship between impacts and school-level characteristics carries practical relevance for policymakers and practitioners because these relationships may provide information that can be used to improve the design and implementation of the program. Thus, the next section describes the school-level measures and the statistical model used to examine the extent to which variation in impacts is related to the local context in which the program was implemented.[5]

[3]There is statistically significant variation in impacts across centers in the first year of the study (p-value = 0.01) but not in the in the second year of the study (p-value = 0.69).

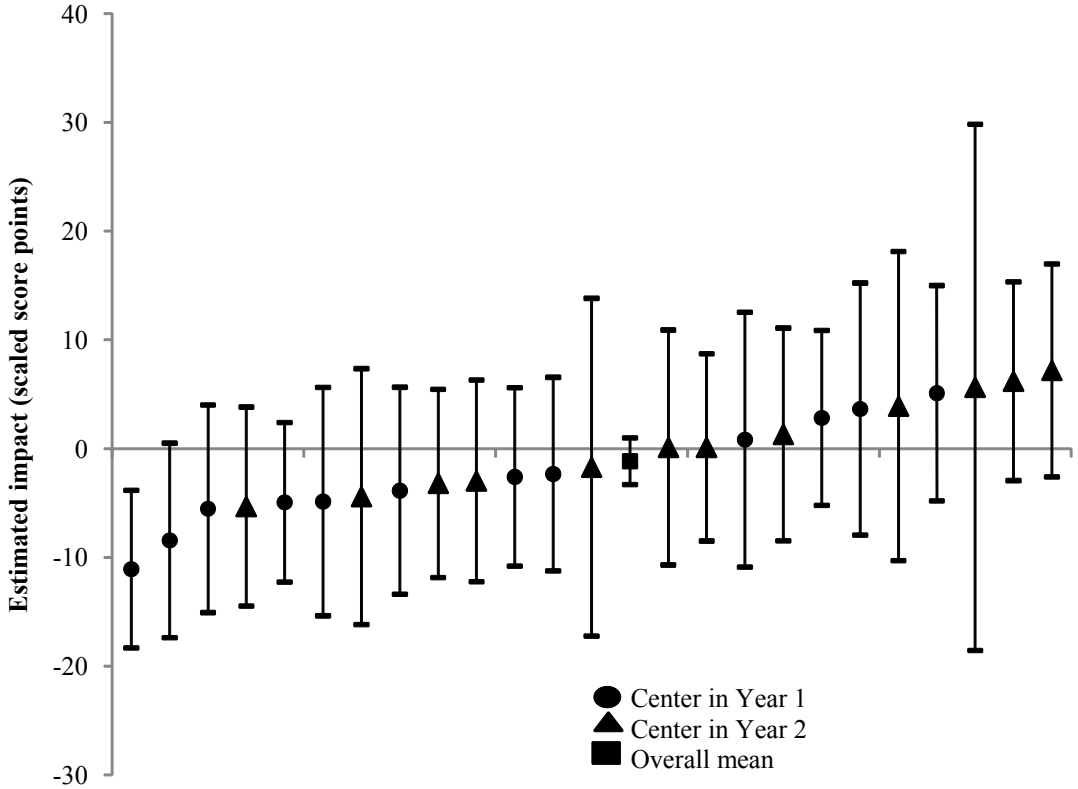
[4]Nor is there statistically significant variation in impacts across centers in the first year of the study (p-value = 0.27) or the second year of the study (p-value = 0.89).

[5]As part of the analysis, it was also found that that after-school centers' year-to-year *change* in impacts is not statistically significant. This suggests that centers' impact from one year to the next is stable (from a statistical perspective).

The Evaluation of Academic Instruction in After-School Programs

Appendix Figure J.2

Impact of One Year of the Enhanced Reading Program on Student Achievement and Its Distribution Across Centers and Implementation Years



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

NOTES: Confidence intervals are based on a statistical signficance level of 5 percent.

The figure shows the estimated program impact for the student-level analysis sample on students' SAT 10 total reading scores (the square; p-value = 0.29) and how that impact is distributed across the 24 center-by-year estimates in the analysis sample (each circle or triangle). The center-by-year impacts (presented ordinally) are estimated by interacting the treatment indicator with center and implementation-year indicators in an ordinary least squares regression model that also controls 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.

Analytic Approach

School-Level Characteristics

Three measures of *program implementation* are included in the correlational analyses: the number of days over the course of the school year that the enhanced math program was offered (included as a measure of program dosage); whether one or more teachers teaching the enhanced program left during the school year (included as a measure of disruption in instruction); and the difference between the total hours of after-school academic instruction received by students in the enhanced program group relative to students in the regular program group (a measure of service contrast).[6]

The correlational analyses also include five measures of the *local school context* that capture the characteristics of the regular school day, as well as the characteristics of the school's student body. These are: the instructional approach of the school-day curricula (available for the math sample but not for the reading sample);[7] how much time is spent in the regular school day on instruction in math or reading;[8] whether the school meet its Adequate Yearly Progress (AYP) goals;[9] what proportion of students in the school receive free or reduced-price lunch;[10] and

[6]This difference is obtained from a regression model that estimates the impact of the enhanced program on the number of hours of after-school academic instruction received by students, controlling for random assignment strata and student characteristics. This regression model is estimated for each center in each year of the study.

[7]Based on their instructional approaches, school-day curricula were categorized into two groups. The first group 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). The reference 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.

This information is not available for the reading sample because 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.

[8]School administrators were asked how many minutes teachers spend per day teaching math 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, which for math was 60 minutes and for reading, 90 minutes.

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

[10]Data on the student-teacher ratio and the proportion of students receiving free or reduced-price lunch come from the National Center for Education Statistics' Common Core of Data (CCD), which compiles school-level demographic information. At the time of writing, 2006-2007 data (corresponding to the second

(continued)

whether the in-school student-to-teacher ratio is greater than the student-teacher ratio in the after-school program (13:1[11]).[12]

Statistical Model

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.[13] Specifically, the analysis estimates whether the size of the impact is related to school-level characteristics by fitting the following linear interaction model:

$$Y_{ik} = \gamma_0 Y_{-1,ik} + \beta_0 T_{ik} + \sum_g \tau_g T_{ik} W_{gc} + \sum_k \gamma_{1k} B_{ik} + \sum_S \gamma_{2s} X_{sik} + \varepsilon_{ik} \quad (1)$$

year of the study) were not yet available. Given that these two characteristics are unlikely to have changed substantially in one year, schools in the second year of the study were assigned their value from the prior year (2005-2006).

[11]As noted in Chapter 2, the planned student-teacher ratio was 10:1; however, up to 13 students were randomly assigned to each class, in order to account for the possibility that some students might not attend on a given day.

[12]Three additional school-level measures were available for the second year of program implementation. The first is the average yearly achievement gain of students in the school, which serves as a proxy for the level and quality of instruction and leadership at the school. The purpose of this "school quality" measure is to examine whether schools that work harder to improve the math skills of their students (i.e., with greater achievement gains) do not benefit as much from the enhanced program because "business as usual" is stronger in these schools and hence the service contrast between the enhanced and regular program group is diminished. Conversely, schools that are chaotic and provide fewer instructional supports to students (i.e., with lower achievement gains) may benefit more from enhanced after-school programs because the instruction otherwise received by students is of lower quality.

The second measure is the percentage of enhanced program teachers in the second year of the study who also taught during the first year (i.e., "returning" teachers). This measure is intended to gauge program implementation strength, since one would expect returning teachers to be better able to deliver the enhanced curriculum than new teachers.

The analysis based on math centers also includes a third additional measure: an indicator of whether, on average, students in the enhanced program spent fewer than four days on each math skill pack assigned by the teacher (where four days is the center-level average in the sample, and "day" is defined as a 45-minute session; adjustments are made for centers that implemented the enhanced program based on a 60-minute session rather than a 45-minute session). This indicator serves as a measure of teachers' instructional pacing — the fewer days spent on each skill pack, the faster the teacher is progressing through the curriculum. Information on skill packs was obtained by Bloom Associates from teacher records.

Given the availability of these additional measures, a separate analysis was conducted focusing on the second year of the study only (i.e., 15 center-level impacts in the Cohort 2 sample for math and 12 center-level impacts in the sample for reading) and using all available school-level characteristics in the second year of the study.

[13]Variation in each of the program implementation and local school context measures across centers and years is statistically significant (p-value for variation of each measure is 0.00).

where:

T_{ik} = Indicator of program group membership (treatment status). This indicator is equal to 1 if student i from random assignment block k was assigned to the enhanced program and zero otherwise.

Y_{-1ik} = The pretest score for student i from random assignment block k before random assignment.[14]

B_{ik} = Random assignment block indicators; equal to 1 if student i is in random assignment block k and zero otherwise.[15]

X_{sik} = The set of s other student-level covariates for student i in random assignment block k.[16]

ε_{ik} = A student-level random error term assumed to be independently and identically distributed.

and where W_{gc} (g = 1, 2,…, G) is a set of G school/center-level characteristics for school/center c, which are hypothesized to be associated with program impacts. W_{gc} includes the following characteristics of *program implementation*:

$TLEFT_c$ = A dummy equal to 1 if one of the instructors teaching the enhanced after-school program left the program during the school year, and zero otherwise.

$SRVDIFF_c$ = The difference between the total hours of after-school academic instruction received by students in the enhanced program group relative to students in the regular program group.

[14]Pretest scores are scaled scores from the SAT 10 tests in reading and math administered in the fall of 2005 (for Cohort 1) and fall of 2006 (for Cohort 2), before the start of the after-school program. Total scores for math are used in the math analysis, and total scores for reading are using in the reading analysis.

[15]In the Cohort 1 sample, random assignment block is defined by grade j and center c in 2005-2006. In the Cohort 2 sample, random assignment block is defined by first-year treatment status (regular program, or new to study) by grade j and center c in 2006-2007. As noted in Table 2.1, fifth-grade students in one site are housed in middle schools. For analysis at this site, the average across fifth-grade students, by first-year treatment status, is taken and then attributed to each of the after-school centers housed in elementary schools. This provides each after-school center in the elementary schools with a fifth-grade random assignment block. And the after-school centers in the middle schools, which only serve fifth-graders, are not included in this analysis as independent centers.

[16]This includes race/ethnicity categories, gender, free-lunch status, age, overage for grade, single-adult household, and mother's education.

$TOTDYS_c$ = The number of days that the enhanced after-school program was offered.

as well as the following characteristics of the *local school context*:

$Group1_c$ = 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 zero otherwise.

$PERIODlong_c$ = 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 zero otherwise.

$NOAYP_c$ = A dummy equal to 1 if the school did not meet its AYP requirements, and zero otherwise.

S/T_c = A dummy equal to 1 if the student-to-teacher ratio in school *k* is greater than the student-to-teacher ratio in the after-school program (13:1).

$\%FRL_c$ = The percentage of students in school *k* who receive free or reduced-priced lunch centered on the grand mean of all schools in the sample.

In Equation (1), τ_g represents the association between the intervention's impact and school characteristic *g*, controlling for other characteristics included in Equation (2).[17] Or, more specifically, it is the change in an after-school center's impact predicted by a one-unit increase in school-level characteristic *g*. If τ_g is statistically significant and positive, it means that a higher level of school-level characteristic *g* is associated with a bigger program impact.[18]

[17]For math centers: *g* = 1 to 8 for the correlational analysis based on the Cohort 1 and Cohort 2 samples; *g* = 1 to 11 for the analysis based on the Cohort 2 only sample, with additional measures (see footnote 12). For reading centers: *g* = 1 to 7 for the correlational analysis based on the Cohort 1 and Cohort 2 samples; *g* = 1 to 9 for the analysis based on the Cohort 2 only sample.

[18]The standard errors used in statistical tests (and associated p-values) are adjusted to account for student-level clustering, due to the fact that students in the regular program group in Cohort 1 were randomized again as part of Cohort 2. The Cohort 2 sample is weighted to reflect the distribution of students across grades at random assignment.

References

Angrist, J.D., G.W. Imbens, and D.R. Rubin. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91 (434):444-455.

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.

Black, A. R., Doolittle, F., Zhu, P., Unterman, R., and Grossman, J. B. (2008). *The Evaluation of Enhanced Academic Instruction in After-School Programs: Findings After the First Year of Implementation* (NCEE 2008-4021). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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.

Durlak, J. A., and Weissberg, R. P. (2007). *The Impact of After-School Programs that Promote Personal and Social Skills.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

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.

Gennetian, L.A, Morris, P.A., Bos, J.M., and Bloom, H.S. 2005. Constructing instrumental variables from experimental data to explore how treatments produce effects. In *Learning More from Social Experiments*, edited by H. S. Bloom. New York, NY: Russell Sage.

Granger, R. C., (2008). After-School Programs and Academics: Implications for Policy, Practice, and Research. Society for Research in Child Development, *Social Policy Report*, Volume XXII, Number 2.

Harcourt Assessment. (2007). *Stanford Achievement Test Series, Tenth Edition – Abbreviated Battery, Areas of Assessment.* 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.

James-Burdumy, S., Dynarski, M., Moore, M., Deke, J., and Mansfield, W. (2005). *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: Final Report*. U.S. Department of Education, National Center for Education Evaluation and Regional Assistance. Washington, DC: U.S. Government Printing Office.

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.

Kling, J. R., Liebman, J. B. , and Katz., L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica* 75 (1):83-119.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., and Martin-Glynn, M. L. (2006). Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students. *Review of Educational Research,* 76 (2):275-313.

Little, P., Wimer, C., and Weiss H. B. (2008, February). *After-School Programs in the 21st Century: Their Potential and What It Takes to Achieve It*. Harvard Family Research Project, Issues and Opportunities in Out-of-School Time Evaluation, Number 10.

Miller, B. M. (2003). *Critical Hours: Afterschool Programs and Educational Success*. Brookline, MA: Miller Midzik Research Associates.

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 8, 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.

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.

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. (2007). *Program Description*. Retrieved September 7, 2007, from
http://www.ed.gov/programs/21stcclc/index.html.

Vandell, D. L., Reisner, E. R., and Pierce, K. M. (2007). *Outcomes linked to high-quality after-school programs: Longitudinal findings from the study of promising after-school programs.* California After-School Network. Web site: http://www.afterschoolnetwork.org/node/759.

Zief, S.G, Lauver, S., Maynard, R.A. (2006). *Impacts of After-School Programs on Student Outcomes: A Systematic Review for the Campbell Collaboration.* Campbell Collaboration Policy Review (Oslo, Norway). Web site:
http://www.campbellcollaboration.org/review_list/categoryPrinterPage.shtml