NCEE 2012-4053

U. S. DEPARTMENT OF EDUCATION

Impacts of Title I Supplemental Educational Services on Student Achievement

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

Impacts of Title I Supplemental Educational Services on Student Achievement

May 2012

John Deke
Lisa Dragoset
Karen Bogen
Brian Gill
Mathematica Policy Research

Yumiko Sekino
Project Officer
Institute of Education Sciences

NCEE 2012-4053
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
Rebecca Maynard
Commissioner

May 2012

The report was prepared for the Institute of Education Sciences under Contract ED-04-CO-0112/0008. The project officer is Yumiko Sekino in the National Center for Education Evaluation and Regional Assistance.

Deke, John, Lisa Dragoset, Karen Bogen, and Brian Gill (2012). *Impacts of Title I Supplemental Educational Services on Student Achievement* (NCEE 2012-4053). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

To order copies of this report,

- Write to ED Pubs, P.O. Box 22207, Alexandria, VA 22304.
- Call in your request toll free to 1-877-4ED-Pubs (1-877-433-7827). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 1-877-576-7734. If 877 service is not available in your area, call 1-800-872-5327 (1-800-USA-LEARN).
- Fax your request to 703-605-6794.
- Order online at www.edpubs.gov.

This report also is available on the Department's Web site at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-0852 or 202-260-0818.

ACKNOWLEDGEMENTS

Many people have contributed in significant ways to the study of Supplemental Educational Services. First, we wish to thank members of the evaluation's Technical Work Group—Thomas Cook, Erick Dorris, Drew Gitomer, Erica Harris, Robert Linn, Steven Ross, Jeffrey Smith, and Ronald Zimmer—who imparted valuable input at critical junctures.

At Mathematica Policy Research, important contributions were made by Laura Kalb, who was the initial survey director for the project; Eric Zeidman, who led the effort to recruit school districts into the study; Lara Hulsey and Kevin Booker who helped recruit school districts; Kristin Hallgren, who helped with early design work and with recruiting; Peter Schochet, who provided advice on technical issues and who provided helpful comments on the draft report; Mason DeCamillis, who programmed the impact models; Jennifer McGovern, Theresa Boujada, Rosalie Malsberger, and Jessica Droz, who were key to the success of the SES provider survey; and Jennifer Baskwell, Cindy McClure, Cindy George, and John Kennedy, who expertly produced and edited the report.

Finally, we would like to extend a special thanks to the staff at each participating district. Without their strong support and participation, this study would not have been possible.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The research team for this evaluation consists of a prime contractor, Mathematica Policy Research of Princeton, New Jersey. Neither the organization nor the key staff members have financial interests that could be affected by findings from the evaluation. None of the Technical Working Group, convened by the research team to provide advice and guidance, has financial interests that could be affected by findings from the study.

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

As one of the parental choice provisions implemented with Title I funds under the No Child Left Behind Act of 2001 (NCLB), parents of low-income students in low-performing schools are offered a choice of Supplemental Educational Services for their children. Supplemental Educational Services (SES) include tutoring or other academic support services offered outside the regular school day, at no charge to students or their families, by public or private organizations that have been approved by the state as SES providers. School districts are required to offer SES to all students from low-income families attending a Title I school that for three consecutive years did not make Adequate Yearly Progress (AYP) toward meeting state standards regarding the percentage of students (or subgroups of students) who have achieved proficiency in reading and math.

Implementation of SES has grown quickly and is now widespread. The number of students participating more than doubled (to 530,000) from 2004–2005 to 2006–2007 (Stullich et al. 2009) and continued growing (to 672,101) in 2008–2009 (www.eddataexpress.ed.gov). Whether ESEA will require SES in its next reauthorization is unclear. The Department of Education's blueprint for reauthorization of ESEA, released in March 2010, recommends that chronically low-performing schools should no longer be required to fund SES but instead should be required to implement any of a number of "data-driven interventions," which could include "expanded learning time, supplemental education services, public school choice, or other strategies" (http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf).

Over the years, SES funding in a few districts has fallen short of what was needed to serve all eligible youth who were interested in accessing services. Federal regulations governing the implementation of SES require prioritization of services to the lowest achieving eligible students when resources are constrained and funds are insufficient to provide SES to all eligible students who request services (Section 116[b][10][C]; 34 CFR 200.45 [d]). A few prior studies have estimated the effect of SES for students receiving services and provide some preliminary evidence of the effectiveness of these tutoring services in raising student achievement (for a summary of some of these studies, and a discussion of implementation issues, see Springer et al. 2009b).

This report presents the findings of an evaluation sponsored by the Institute of Education Sciences (IES) at the U.S. Department of Education (ED) and conducted by Mathematica Policy Research (Mathematica) that uses a regression discontinuity (RD) design to assess the potential benefits of offering SES in districts that have unmet need. Specifically, the study focuses on six school districts in which more eligible students applied for SES than could be served with available funds (i.e., oversubscribed districts), and which therefore allocated scarce SES spaces by giving priority to lower-achieving students among the eligible applicants. The current study is the first to use a research design that can directly account for selection into services and therefore (if reasonable assumptions are met) support drawing causal inferences about the impact of SES on students at the cutoff of receiving services (Hahn, Todd, & Van Der Klaauw, 2001; Shadish, Cook, & Campbell, 2002; Imbens & Lemieux, 2008). The study's RD design estimates the impact of SES on academic achievement by comparing the post-test scores of students offered services with those for students whose prior achievement scores were slightly too high to be offered services, with the comparison adjusting for the prior achievement scores that were used to determine assignment to services. The estimates apply to students in the study's school

districts who are on the cusp of receiving services. Because different cut points were used in different grades across the six districts, these estimates show the effect on students on the cusp of participating in the program in multiple contexts in oversubscribed districts.

Research Questions

The evaluation study's design and data collection plans were developed to provide the most methodologically rigorous answers to the study's primary research question:

1. What is the average impact of *offering* SES to eligible applicants who are on the cusp of having access to services, in school districts where services are oversubscribed?

The study also addresses an additional research question:

2. What are the characteristics of SES provided to students in oversubscribed districts? Are the characteristics of services, providers, or practices in host school districts correlated with the estimated impacts?

Key Findings

Findings from this evaluation are based on six, nonrepresentative school districts in three states (Connecticut, Ohio, and Florida) where more eligible students applied for SES than could be served with available funds. Across the six oversubscribed districts included in our study, there were 50,843 applicants to SES, of which 30,673 were in study grades (3–8). Among these applicants, 24,113 were assigned to services based on a measure of prior achievement. These 24,113 constitute the study population. Among the students in the study population, 19,750 students (82 percent of the applicants) were offered SES based on their prior scores while 4,363 students were not offered SES, and 16,954 (86 percent of those offered services) participated in SES (i.e., reported any tutoring services). Almost all participating students (98 percent) were offered services with their first choice providers, with most students being served by a few dominant providers.

The answers to the key questions addressed in the study are as follows:

- ***What is the average impact of offering SES to eligible applicants who are on the cusp of having access to services, in school districts where services are oversubscribed?*** Across the six oversubscribed districts, we find no evidence of impacts of offering SES to students near the cut point for an offer. For students in these oversubscribed districts in grades 3–8 at the cusp of receiving an offer of services, we find no statistically significant impact of *offering* SES on student achievement in reading or in mathematics.[1] The point estimate of the average impact on reading is -0.03 standard deviations, and that for mathematics is 0.05 standard deviations. Furthermore, there is no evidence of potential benefits for at-risk

[1] Note, however, that the study did not explicitly test the expansion or contraction of SES. Estimated impacts are based on the assumption that the number of students offered services would change at the cutoff while all other relevant factors are unchanged.

subgroups of students. Similarly, we find no statistically significant impact of *participating* in SES on student achievement in reading or mathematics. The estimated impact of participating (which involved an average of 21 hours of services) is -0.10 standard deviations for reading and 0.11 standard deviations for math (again estimated for students in grades 3-8 near the cutoff for an offer, in these oversubscribed districts).

- ***What are the characteristics of SES provided to students in oversubscribed districts?*** Across districts participating in this study, services averaged 21.2 hours per student for the school year (standard deviation of 8.8), with over a third of the students (36 percent) receiving tutoring in both reading and math, 55 percent receiving tutoring in only reading, and 9 percent receiving tutoring in only math. For students receiving reading services, the mean was 17.2 hours of tutoring (standard deviation of 9.2). For students receiving math services, the mean was 12.5 hours of tutoring (standard deviation of 8.2). In the study districts, most providers (70 percent) were for-profit firms. On average, 60 percent of providers' instructional staff were regular schoolteachers working in the local district. Most providers (64 percent) offered services at the schools of their students. Providers reported that group sizes of 2–5 students were most frequently used, with most other sessions provided individually (in one-on-one sessions). On average, 44 percent of provider services were in groups of 2–5 students, 34 percent in one-on-one sessions, and 21 percent in groups of 6–10.
- ***Are the characteristics of Supplemental Educational Services, providers, or practices in host school districts correlated with the estimated impacts?*** There is variation across providers in the average number of hours of math and reading services received by students, with average hours in services focused on math ranging from 0 to 27 across providers and average hours in services focused on reading ranging from 0 to 43. However, the intensity of services is not significantly related to the estimated size of impacts on math or reading achievement (for these eligible students near the cutoff for having been offered SES). We also found no evidence that any other observed provider characteristics were significantly associated with stronger impacts.

Background on Title I SES

Districts are required to offer SES to all low-income students in schools that have fallen short of Adequate Yearly Progress (AYP) proficiency standards for a third time. (After missing AYP for two consecutive years, schools receive the "identification for improvement" designation.) As one of the options available to parents under NCLB, the SES program permits eligible families to choose from a wide variety of state-approved SES providers, including national for-profit firms, local nonprofits, faith-based organizations, institutions of higher education, and local school districts. School districts are permitted to become approved providers unless they are themselves identified for improvement under NCLB. As of September 2010, more than 3,000 providers across the country had been approved to offer SES.

Federal regulations govern the funding of SES. State education agencies allocate Title I funds to districts with eligible students. Those districts are then required to make available up to 20 percent of their total Title I, Part A funds for a combination of SES and transportation for

students using NCLB's school-choice option. In addition to regulations for district-wide SES expenditures, there is a restriction on the district's maximum expenditures per student. NCLB requires provision of SES to all eligible students if expenditures are less than 20 percent of their Title I, Part A funds. However, if funds are insufficient to provide SES to all students who request them, non-regulatory guidance from the U.S. Department of Education (ED) indicates that priority must be given to the lowest-achieving students (*Section 1116[b][10][C]; 34 C.F.R. §200.45[d]*[2]*)* based on "fair and equitable procedures in determining which students are the lowest achieving." This prioritization rule creates the opportunity to conduct a study using an RD design in oversubscribed school districts.

What Kinds of Districts and Students Were Included in the Study?

Oversubscription for SES is unusual among school districts. In 2008-09, when we were drawing our study sample, ED's Office of Innovation and Improvement (OII) identified only forty school districts that might have more eligible applicants for SES than they could serve. We ultimately found only nine of these assigned districts that were actually oversubscribed and that intended to ration scarce spaces based on a "cut-off" value on students' prior year's test score. Eight of the nine districts agreed to participate, of which two districts were subsequently excluded from the study because assignment practices did not support the RD design.

The study's six school districts are by no means nationally representative. They are in three states: Connecticut, Ohio, and Florida. As a result of statewide rules regarding unspent SES funds, Florida districts had a stronger incentive than districts in other states to ensure that as many students as possible were participating in SES. Unlike most other states at the time, Florida had in place funding rules that did not permit unused SES funds to be automatically returned to a school district's overall Title I allocation (http://www.fldoe.org/flbpso/pdf/finalpro.pdf). Aside from being oversubscribed for SES, the six study districts also differed from the average characteristics of all Title I school districts in having higher proportions of economically disadvantaged students. Study districts were largely comprised of urban and suburban schools (38 percent and 56 percent, respectively), with 66 percent of district students eligible for free/reduced-price lunch (as compared with 56 percent of students in all Title I schools nationally). These six districts, in comparison with Title I schools nationwide, also had higher percentages of black students (34 percent versus 20 percent) and Hispanic students (46 percent versus 28 percent).

Across the six study districts, applicants to SES numbered 50,843 students, 30,673 of whom were in study grades (3–8). Of those, 24,113 were assigned to treatment and comparison groups using a continuous measure of prior achievement (while the rest were in different categories, such as prioritized grade levels in which all eligible applicants were admitted). These 24,113 students constitute the study population. Among the study population, 19,750 students (82 percent) were offered SES based on their prior scores, 4,363 students were not offered SES, and 16,954 (86 percent of those offered services) participated in SES (i.e., reported any tutoring services) Parents of eligible students were given a choice of approved providers, and the overwhelming majority of students (98 percent) were assigned to their first choice.

[2] See http://www2.ed.gov/nclb/choice/help/ses/guidance.html for non-regulatory guidance.

Five percent of SES participants in our sample were white, 47 percent were black, and 46 percent were Hispanic. Sixteen percent of participants were English language learner (ELL) students, and 21 percent of participants were students with a disability (those in special education).

What Were the Study's Data Sources?

The data collection had two distinct sources that are integrated to answer the study's research questions. The first is student-level data on all SES applicants from the districts, including demographics, SES provider choices, number of hours of tutoring received, and standardized test scores in math and reading from 2007, 2008, and 2009. Though collected from districts, the standardized tests used as outcomes are the same tests used for state accountability purposes. The second component is a survey of SES providers in the six districts that gathered information on provider characteristics and practices.

What Is the Potential Achievement Benefit of Offering SES to Applicants at the Cusp of Eligibility in Oversubscribed Districts?

This study uses an RD analysis to assess the effectiveness of SES implemented in real-world settings among school districts that lacked the funds needed to serve all SES applicants eligible for services.

Applications for SES were received by districts in the fall of 2008 and students received services during the 2008–2009 school year. Students were offered SES based on a continuous measure of academic need (such as a state achievement test score). In each district, students with prior test scores below a specific cutoff value were offered SES (the treatment group), whereas students above the cutoff value were not offered SES (the comparison group). Under this design, estimates of the impacts of SES on students at the cutoff for being offered services can be obtained by comparing the outcomes of students just below and above the cutoff value, after adjusting for the score on the test used to assign students to treatment and comparison groups (see, for example, U.S. Department of Education 2010; Shadish, Cook, & Campbell, 2002). Cutoff scores varied across districts and often across grade levels within districts, creating 42 "mini-studies" with unique cut points. Because different cut points were used, these estimates show the effect on students in multiple contexts (assuming other factors are unchanged).

The study had an estimated 80 percent or greater chance of detecting an overall impact of SES if the true impact is 0.12 standard deviations or larger. An impact of 0.12 standard deviations is equivalent to moving from the 50th percentile to the 55th percentile on a test with normally distributed scores.

For our primary impact analysis, we found no statistically significant effect of offering SES to eligible applicants near the cutoff, either on reading scores or math scores. Point estimates were -0.03 standard deviations in reading and 0.05 standard deviations in math, but neither approached statistical significance.

Because not all students with prior scores below the relevant cutoff value actually received services (2,796 students, or 14 percent of those students offered SES, did not participate), we also estimated the impact of participation. For applicants near the cutoff, there were no statistically significant impacts of participation in SES on either subject. Other variants of the

impact of participation were estimated using alternative definitions of SES participation, including the impact on the specific subject in which services were received. We found no statistically significant impacts.

What Were the Key Characteristics of Services, and Were Any Characteristics Related to Estimated Impacts?

The study examined characteristics of SES providers and the services they provided to students. Understanding the variation in provider services and students' access to SES is important context for understanding the impact analyses. In addition, we capitalize on variation in program implementation across districts and providers to look at the relationship between such characteristics and program impacts. None of these relationships were stronger than what we would expect to see by random chance alone.

The first descriptive question of interest relates to the intensity of services. In fact, over the course of an academic year, providers reported that a complete course of services was about 28 hours, on average. This is substantially less than the 45-hour average found in a previous survey of providers in 16 districts across the country (Vernez et al. 2009). The length of a full course of services varied across providers, from 13 hours to 108 hours, with approximately 70 percent of providers reporting that a full course of services was 30 hours or less. However, there was no statistically significant relationship found between the length of the course of services and the estimated impacts of provider services on students at the cutoff for receiving SES.

Average student participation in our sample was even less—at 21.2 hours—than the average full course. Across the six districts, 36% of SES participants (6,058) received tutoring in both math and reading, 9 percent (1,560) received tutoring in only math, and 55 percent (9,336) received tutoring in only reading. For students who received tutoring in math, mean math hours were 12.5 (standard deviation of 8.2). For students who received tutoring in reading, mean reading hours were 17.2 (standard deviation of 9.2). Thus, time spent in SES was only a small fraction of the time that students typically spend in school. For example, if a student received 12.5 hours of SES math instruction (the mean hours received), it would equal approximately 7 percent of a full year of math instruction, assuming that students spend one hour a day on math as part of their regular curriculum. Likewise, if a student received 17.2 hours of reading instruction, it would equal approximately 10 percent of a full year of reading instruction. The average number of hours of services that students received was not associated with the estimated impacts of provider services on students at the cutoff for receiving SES.

Various other characteristics of services and providers were potentially relevant to their effectiveness in raising student achievement, and therefore potentially of interest to state authorizing agencies and to the Department of Education. For example, provider size varied widely. There were a few dominant providers in each district; in our sample, 9 providers served 50 percent of the students, with the remaining half of students served by the remaining 179 providers. Providers had, on average, at least two years of experience providing SES or other educational services. The majority of providers (69 percent) were for-profit organizations that relied extensively on local school teachers to serve as their SES instructors. Almost all services were provided face-to-face (92 percent of providers), most commonly in small group sessions of 2 to 5 students (44 percent in groups of 2–5; 34 percent in one-on-one sessions; and 21 percent in groups of 6–10). Larger providers were more likely to rely on slightly larger group sessions (6 to

10 students). A majority of SES providers (63 percent) administered multiple diagnostic assessments to students. Although providers reported their programs to be aligned with state standards, they typically were unable to identify the particular instructional programs used in the schools of the local district.

We examined relationships between provider characteristics and impacts, but none of these relationships were stronger than what we would expect to see by random chance.

I. INTRODUCTION

A. Background

Under the No Child Left Behind Act (NCLB, the latest reauthorization of the Elementary and Secondary Education Act), more than 50,000 high-poverty schools received Title I funds. These funds can be used in a variety of ways to help schools ensure that all their students meet state academic standards. Recipients generally must spend the funds on services aimed at failing or at-risk students, although schools with 40 percent or more of their students from low-income families are eligible to spend them on school-wide programs. One of the provisions being implemented with Title I funds is Supplemental Educational Services (SES), which are tutoring or other academic support services offered outside the regular school day, at no charge to students or their families, by public or private providers that have been approved by the state. Districts are required to offer SES to all students from low-income families in Title I schools that have not made Adequate Yearly Progress (AYP) toward proficiency standards for a third year. The program is one of the parental choice provisions created by NCLB: Eligible students and their parents are permitted to choose among a wide variety of state-approved SES providers, including national for-profit firms, local nonprofits, faith-based organizations, institutions of higher education, and local school districts. School districts are permitted to become approved providers unless they are themselves identified for improvement under NCLB. As of September 2010, more than 3,000 providers across the country had been approved to offer SES.[3]

Implementation of SES has grown quickly and is now widespread. The number of students participating more than doubled (to 530,000) from 2004–2005 to 2006–2007 (Stullich et al. 2009) and continued growing (to 672,101) in 2008–2009 (www.eddataexpress.ed.gov). The potential market for SES is substantially larger: fewer than one in six of the 4.3 million students eligible for SES participated in 2008–2009. Participation rates across states vary: 15 states have enrollment rates below 10 percent of eligible students, while 6 enroll more than 30 percent of their eligible students (www.eddataexpress.ed.gov). These numbers may grow as the number of schools missing AYP for a third year increases and as parents become better informed about SES availability.[4]

As student participation in SES has increased over the years, SES funding in some districts has fallen short of what was needed to serve all eligible youth interested in accessing services. Federal regulations governing the funding of SES require state education agencies to allocate Title I funds by formula to districts with high numbers of students from low-income families. Districts are then expected to spend up to 20 percent of their Title I, Part A funds on SES and transportation for students who choose to attend a better-performing public school. If there is sufficient demand, districts are required to spend the full 20 percent on these services. If funds are insufficient to provide SES to all eligible students who request them, non-regulatory guidance from the U.S. Department of Education (ED) indicates that priority must be given to the lowest-achieving students (*Section 1116[b][10][C]; 34 C.F.R. §200.45[d])* based on "fair

[3] This is based on a survey of state education agency websites. Approved provider lists were not available for New Mexico, South Carolina, and South Dakota.

[4] For more background information on SES, see Springer et al., 2009b.

and equitable procedures in determining which students are the lowest achieving." Services are targeted to students who are underachieving generally or in a specific subject area, or who are in specific grades (http://www2.ed.gov/policy/elsec/guid/suppsvcsguid.pdf).

In addition to the regulation regarding district-wide SES expenditures, there is a formula specifying the maximum amount that can be spent on each student: the district's SES allocation divided by the number of its low-income students (http://www2.ed.gov/about/overview/budget/titlei/fy05/index.html). The 2009 non-regulatory guidance called for states to report this maximum per-student amount on their websites. The amounts spent per pupil on SES vary greatly: in 2006–2007, nearly one-third of districts spent less than $600 per pupil, on average, while 17 percent spent more than double that amount (Vernez et al. 2009).

In 2008, ED published additional regulations on coordination of SES. These regulations stipulate that districts must (1) provide timely and clear notification to parents about the option of SES, and (2) post online the number of eligible and participating students, as well as a list of SES providers and locations. Districts can spend 0.2 percent of their Title I, Part A allocation on parent outreach and assistance. State education agencies must also post the following information online: the maximum amount each district is expected to spend on SES; the maximum per-pupil expenditure; and a list of all approved SES providers (http://www2.ed.gov/policy/elsec/reg/ proposal/strengthening-choice.pdf).

Whether the Elementary and Secondary Education Act (ESEA) will require SES in its next reauthorization is unclear. The Department of Education's blueprint for reauthorization of ESEA, released in March 2010, recommends that chronically low-performing schools should no longer be required to fund SES but instead should be required to implement any of a number of "data-driven interventions," which could include "expanded learning time, supplemental education services, public school choice, or other strategies" (http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf). Lessons learned from the evaluation of SES about academic support provided outside the school day may be relevant to other methods of extending learning time.

B. Prior Research on Effectiveness of SES

During the past few years, several studies have used longitudinal, nonexperimental methods to examine the relationship between SES participation and student achievement. A handful of studies have examined the average effect of SES for students receiving services by comparing the achievement trajectories of SES participants before and after participation with those of eligible nonparticipants (Zimmer et al. 2007; Heinrich, Meyer, & Whitten, 2010; Springer, Pepper, & Ghosh-Dastidar, 2009; Ross et al. 2009; Munoz & Ross, 2009). They all rely on variants of a difference-in-differences approach, in which the achievement gains of SES participants are compared to the gains of nonparticipants in the same districts. The largest of these studies, conducted by the RAND Corporation (Zimmer et al. 2007) for the Policy and Program Studies Service at ED, found evidence of positive achievement gains for students participating in SES in five of seven large urban districts where the analysis was conducted (Zimmer et al. 2007). Across the seven districts, about 50,000 student observations contributed to the estimates. The majority of districts provided data for grades 3–8 for most years, and some provided data for grades 1 and 2 and 9–11. The achievement trajectories of SES students before and after participating were compared with those of nonparticipating students. The authors used

a student-level fixed-effects model to control for time invariant characteristics such as parent education, income, and underlying student ability/motivation. They found that math test scores of SES students were 0.09 standard deviations higher than math test scores of non-SES students (with a 95 percent confidence interval ranging from 0.03 to 0.14) averaged across seven districts.

Another study has reported some evidence of positive relationships between SES participation and student achievement. Springer, Pepper, and Ghosh-Dastidar (2009) looked at roughly 10,000 students in grades 3–8 in a large urban school district. For each student who attended SES tutoring, the authors compared actual performance to expected performance, which was calculated using performance of non-SES students. The authors controlled for student- and school-level covariates, actual attendance of students enrolled in SES, the subject in which the student received the most tutoring, and the number of years of attendance. They found significantly higher test scores in math (0.09 standard deviations higher) for students attending SES, but no significant differences in reading scores. Two alternate approaches—using a comparison group of future SES participants and using propensity-score matching—produced similar results.

Meanwhile, several studies have found no support for any positive achievement gains associated with participating in SES. Heinrich, Meyer, and Whitten (2010) found no significant changes in the student achievement test scores before and after SES for students in the Milwaukee public schools. The researchers used math and reading test score data for grades 5–10 for most of the school years, and the sample sizes ranged from 400 to 2,000 over the two subjects and the multiple school years. Propensity scores were calculated to predict SES participation, and both difference-in-differences and fixed effects models were used to examine differences in student achievement changes associated with SES attendance. The researchers estimate both the changes associated with *any* SES attendance on student achievement as well as the difference in changes related to different levels of hours of SES attendance among those who attend. Heinrich, Meyer and Whitten (2010) also found no statistically significant differences in achievement gains associated with SES in middle school or high school. Munoz and Ross (2009) looked at test scores on math assessments for about 560 students and reading assessments for about 2,500 students, in grades 3–8 and grade 10 in public schools in Jefferson County, Kentucky. Examining test scores on state-mandated assessments in math and reading, they also found no significant differences between SES attendees and a comparison group of non-attendees matched on demographic and achievement indicators. Ross et al. (2009) examined the relationship between SES attendance and student achievement in two districts in Tennessee, using value-added methodology to control for student ability and teacher effects in two different models. They included only students who completed 50–100 percent of their contracted SES hours, and they matched these students with non-SES students with similar predicted achievement scores. The researchers found no significant improvement in reading assessments among roughly 350 students receiving SES tutoring in grades 4–8. Among 250 students receiving math services, they found that services were associated with a moderate but statistically significant shortfall in math assessment results.

These prior studies provide preliminary evidence of the effectiveness of SES. The current study is the first to use a research design that can directly account for selection into services and therefore (if reasonable assumptions are met) support drawing causal inferences about the impact of SES on students at the cutoff for receiving services (Hahn, Todd, & Van Der Klaauw, 2001; Shadish, Cook, & Campbell, 2002; Imbens & Lemieux, 2008).

C. Study Overview and Research Questions

This report presents the findings of an evaluation sponsored by the Institute of Education Sciences (IES) at ED and conducted by Mathematica Policy Research (Mathematica) that uses a regression discontinuity (RD) design to assess the potential benefits of offering SES to additional eligible applicants in six school districts in which more students applied for SES than could be served with available funds. If funds are not sufficient to provide SES to all eligible students, NCLB requires districts to prioritize services to the lowest-achieving eligible students. This allocation of SES when resources are constrained creates the opportunity for an RD analysis. Using an RD design, this evaluation estimates the impact of SES on the academic achievement of students in the study's school districts at the cutoff for services (i.e., in the vicinity of the achievement score cutoff value for assigning students to SES), provided that all relevant assumptions of the RD design are met. Among students at the cutoff for SES, the post-test scores of students whose applications to SES were accepted were compared to those of students whose applications were not accepted, after adjusting for the prior achievement scores that were used to determine assignment to services.

An earlier feasibility study determined that there were, in fact, oversubscribed districts that could support an evaluation of SES using an RD design although oversubscription is not common among school districts. In that study, we contacted school districts identified by the Office of Innovation and Improvement (OII) as being potentially oversubscribed. Through conversations with district officials, we determined that enough districts were already oversubscribed (during the 2007–2008 school year) or close to being oversubscribed to support this evaluation (Hallgren, Deke, & Angus, 2007). We worked with six school districts that were oversubscribed for SES during the 2008–2009 school year and that collectively enrolled more than 27,000 SES participants. These six districts prioritized services to the lowest achieving eligible students using a quantifiable measure of academic need on which a cutoff score could be determined.

The evaluation study's design and data collection plans were developed to provide causal answers to the primary research question:

1. What is the average impact of *offering* SES to eligible applicants who are on the cusp of having access to services, in school districts where services are oversubscribed?

The answers to this question represent our assessment of whether providing access to more eligible applicants in the study's oversubscribed school districts would result in achievement gains for those students. In this context, students are offered SES if they are below the cutoff value on the continuous assignment variable. The analysis produces an *intent-to-treat* (ITT) estimate for students in the vicinity of this cutoff value for services. We restrict the analysis to students in grades 3–8 because all students in those grades take state achievement tests. Because not all students whose SES applications are accepted actually participate, we use a "fuzzy" RD analysis to assess the potential benefits of *participating* in SES. The analysis produces an estimate of the effect of *treatment-on-the-treated* (TOT) that is derived from the potential effect of offering SES to students at the cutoff for being offered services.

The evaluation also addressed an additional research question:

2. What are the characteristics of SES provided to students in oversubscribed districts? Are the characteristics of services, providers, or practices in host school districts correlated with the estimated impacts?

We conducted descriptive analyses to describe the providers and the characteristics of the services offered students in the six study districts. Examining the relationships between impact estimates and SES provider characteristics in research question 2 will provide evidence that potentially can inform the refinement of SES programs. Unlike the first research question, however, research question 2 cannot be answered using methods that support causal inferences.

One research question regarding SES impacts that is *not* addressed in this study merits mention. The study is designed to assess the potential benefits of offering SES to more eligible applicants in oversubscribed sites, by producing an unbiased estimate of the impact of SES on the achievement of students at the cutoff for being offered services. However, it is not designed to estimate the net policy impact of shifting Title I resources to SES from other purposes determined by school districts (i.e., the study does not assess how students in these districts would have been affected had the SES funding been spent on other educational interventions). In the absence of the SES requirement, school districts might have been permitted to use the SES funds for other educational purposes they deemed appropriate. The SES policy mandates a reallocation of resources in districts required to offer SES; it does not provide districts with new resources to support SES.

D. Report Outline

Chapter II describes our study design in detail, and also describes the data we collected. Chapter III summarizes descriptive information about the characteristics of SES providers and their services in the six study districts. In Chapter IV we present our primary findings addressing the first research question described above regarding the potential benefits of offering SES to more applicants in these oversubscribed districts. Chapter IV also summarizes findings from a series of exploratory analyses, including TOT impact estimates to address the potential benefits of participating in SES for students eligible for but not able to access services, and the relationships between provider characteristics and impact estimates. The appendices provide additional details about methods and results.

II. STUDY DESIGN AND DATA COLLECTION

This study was designed to assess the effectiveness of Supplemental Educational Services (SES) implemented in real-world settings by school districts that lacked the funds needed to serve all SES applicants eligible for services. Because NCLB requires prioritizing SES to the lowest achieving eligible students when resources are constrained, we used a regression discontinuity (RD) design to examine the potential benefits of offering SES to more applicants in six school districts in which more students applied for SES than could be served with available funds. Using an RD design, we estimated the impact of SES on student outcomes for those students at the cutoff for being offered services, specifically in the vicinity of the achievement score cutoff value for assigning students to SES. Applications for SES were received by districts in the fall of 2008, and students received services during the 2008–2009 school year. Applicants were offered SES based on a measure of academic need such as their previous score on a state achievement test. The eligible applicants who were offered SES constitute our treatment group; students who applied but were not offered SES (based on the prior test score) constitute the comparison group. The total analysis sample included 24,113 students who were assigned to treatment and comparison groups.

This chapter presents information about the research design and data collection for the SES study, focusing on the way the study was structured to provide answers to the key questions of interest. Sections A through D present the characteristics of the participating districts, the design of the study, the types of data collected, and the analytic approach used to estimate program impacts.

A. Characteristics of Study Districts

We sought to include in our study every school district in the country that was oversubscribed for SES in 2008–2009. Potentially oversubscribed districts were identified by ED's Office of Innovation and Improvement (OII) through its work with state SES coordinators and contacts with local school districts. Existing Government Accountability Office (GAO) research on SES was also reviewed to identify additional districts that might be eligible for the evaluation. In 2008, OII identified a total of forty districts that might be oversubscribed. Of these forty potential districts, nine met all the study criteria (namely, being oversubscribed for the 2008–2009 school year and assigning students to treatment and comparison groups using a specific procedure of a cutoff value on a continuous measure of prior achievement). Of those nine, eight agreed to participate in the study. As part of their agreement to participate in the study, each district provided Mathematica with information on all applicants, including their standardized test score data for three years. Two of the eight participating districts were excluded from the study because assignment practices could not support the RD design.

The six school districts ultimately included in the study are in three states: Connecticut, Ohio, and Florida. Unlike most other states at the time, Florida had in place funding rules that did not permit unused SES funds to be automatically returned to a school district's overall Title I allocation. In Florida, districts that did not spend 20% of Title I funds on a combination of school

choice and supplemental educational services could lose the unspent portion of the funds.[5] They therefore had a stronger incentive than districts in other states to ensure that as many students as possible were participating in SES. The fact that four of the six districts in the study are from Florida is likely the result of the state's more aggressive efforts to promote SES participation.

As SES implementation is targeted for districts with Title I schools that did not make AYP for a third year, we compared the characteristics of the Title I schools in the six study districts with other Title I schools. These six districts differ specifically from other U.S. school districts offering SES in that they are oversubscribed. In addition, the Title I schools in the study districts differ from all Title I schools nationwide. Table II.1 shows characteristics of Title I schools in the study districts, including the percentage of students eligible for free/reduced-price lunch (FRPL), the percentage of Hispanic students, the percentage of black students, and the percentage of city/suburban/rural/town schools. Student-level variables (FRPL and race/ethnicity) are weighted using the number of students in each school. Across all Title I schools in the study districts, the average percentage of students eligible for FRPL was 66.1 percent. For comparison, Table II.1 also shows the same characteristics for all Title I schools nationwide. Compared to Title I schools nationally, the study Title I schools had a greater percentage of black students (34 percent versus 20 percent) and Hispanic students (46 percent versus 28 percent), and a larger percentage of city and suburban schools (39 and 56 percent, respectively, versus only 28 and 25 percent for Title I schools nationwide).

Table II.1. Characteristics of Title I Schools in Study Districts Compared to All Title I Schools

Characteristic	Average Across Study Title I Schools	National Average for Title I Schools
Percentage free/reduced-price lunch	66.1	55.6
Percentage black	34.3	19.2
Percentage Hispanic	45.7	27.8
Percentage white	15.9	47.0
Percentage of city schools	38.2	27.8
Percentage of suburban schools	55.5	24.7
Percentage of rural schools	2.5	31.8
Percentage of town schools	3.8	15.4

Source: Common Core of Data (http://nces.ed.gov/ccd/).

While school districts' SES expenditures are capped at 20 percent of total Title I funds, each district has a different maximum per-pupil spending allotment for students participating in SES. As noted in Chapter I, the maximum per-pupil allocation is calculated by dividing the amount set aside for SES and school choice transportation (20 percent of Title, Part A funds) by the number of low-income students in the district (http://www2.ed.gov/about/overview/budget/titlei/fy05/index.html). The median maximum per-pupil allocation across the six study districts was $1,450, with a range from $1,100 to $2,300. In

[5] Although the default practice would cause a district to lose left-over funds, there was an option for districts to apply to reallocate unspent SES funds (see http://www.fldoe.org/flbpso/pdf/finalpro.pdf). The districts in the study, however, were rationing spaces because demand for services involved costs that would have exceeded their required expenditures.

contrast, the median maximum per-pupil allocation nationally was $1,240. Five of the six districts in the study had maximum per-pupil allocations exceeding the national median.

Federal regulations governing district-wide funding of SES and maximum per-pupil expenditures are likely to limit the service hours that each student could be offered. Table II.2 shows an estimation of the SES dosage per student that would be implied in each district if the maximum per-pupil dollar amounts were spent and services were charged at the hourly rate that was the average for providers operating in the district (as reported on state websites). The hours are calculated by dividing the district's per-pupil maximum by the average hourly rate for local providers. The number of hours a student could theoretically receive based on this calculation varies across districts, from a low of 18.3 to a high of 49.0. The median theoretical dosage across districts was 23.4 hours. In contrast to this theoretical dosage, a prior study conducted by ED surveyed providers across 16 districts to examine state and local implementation of SES across the country and found an average of 45 hours of SES annually (Vernez, Naftel, Ross, Le Floch, Beighley, & Gill, 2009).

Table II.2. Theoretical SES Dosage per Student Based on Maximum Per-Pupil Expenditure and Average Provider Hourly Rate

District	Hours Per Student Implied by Maximum Per-Pupil Expenditure and Average Hourly Rate of Local SES Providers
1	49.0
2	31.5
3	18.3
4	22.4
5	22.7
6	24.1

Source: State websites and http://www2.ed.gov/about/overview/budget/titlei/fy09/index.html.

Note: The theoretical number of hours per student is calculated as the maximum per-pupil Title I allocation for SES in FY 2008 (available at http://www2.ed.gov/about/overview/budget/titlei/fy09/index.html) divided by the average hourly provider rates (available on state websites) for providers in the district.

B. Study Design: A Study of Many Mini-Studies

We conducted a separate RD study for each outcome-grade-district combination (a total of 42 mini-studies). As described above, school districts' SES expenditures are capped at 20 percent of total Title I funds. When the number of students participating in SES is large enough that the costs reach this cap, NCLB permits districts to ration services, with priority given to the lowest-achieving students among applicants. The prioritization rule creates an opportunity to conduct a study in oversubscribed school districts using an RD design. When funds were not sufficient to provide SES to all eligible students, the six study school districts used a continuous measure of students' academic need (the "assignment variable") to determine the allocation of services. While all six districts applied this same rule, the cutoff value differed in each district, and some districts used different cutoff values for different grade levels. Appendix A contains details on the rules used in each district to determine who is offered SES and explains how SES and non-SES students were identified in the analysis.

Under the RD design, estimates of the impacts of SES on students at the cutoff can be obtained by comparing the outcomes (follow-up test scores) of students below and above the cutoff value, after adjusting for the score on the assignment variable. Unlike an RCT, in which the estimated average treatment effect applies to all students in the study, the RD impact applies

only to students near the cutoff value of the assignment variable. This estimate does not necessarily represent the impact of SES on students far away from the cutoff value of the assignment variable. In other words, the RD analysis examines whether there is a discontinuity in the relationship between the assignment variable (prior achievement) and the outcome (subsequent achievement) at the prior achievement level that is the cutoff for assignment to services. Under plausible assumptions, mean differences in outcomes between students below and above the cutoff value, controlling for pretest scores, generate unbiased estimates of the impacts of access to services for students near the cutoff for being offered services in these oversubscribed districts (see, for example, U.S. Department of Education 2010; Shadish, Cook, & Campbell, 2002).[6]

We refer to each unique combination of outcome, grade, school district, and cutoff value of the assignment variable as an RD mini-study.[7] Because studies must be conducted separately for each grade and subject, the six districts produce a total of 42 RD mini-studies, 21 corresponding to the math test score outcome and 21 to the reading test score outcome. Every mini-study involves its own independent RD analysis (including bandwidth selection, diagnostic analyses, and sensitivity analyses). Every mini-study has its own cutoff point for the assignment of services. In consequence, these estimates show the effect on students in multiple contexts. An overview of these analytic techniques is provided in Section C of this chapter, with additional details provided in Appendices B, C, and E.

Aggregating Impact Estimates. The overall estimated impact of SES is calculated as a weighted average of all of the mini-study impact estimates. Our benchmark approach is to use a fixed-effects meta-analytic approach (Cooper, Hedges, & Valentine 2009) in which we weight each impact estimate by the inverse of the variance of each impact, meaning that more weight is given to impacts with smaller variances. This approach provides greater statistical precision than other weighting approaches (such as equal weighting or sample size weighting). If impact estimates across mini-studies vary systematically, then alternative weighting approaches might yield a substantively different aggregate impact. To assess the sensitivity of our findings to our choice of aggregation weights, we also report overall impact estimates giving each mini-study equal weight and using sample size as the weight.

We do not make a multiple comparisons adjustment to the impact estimates of SES on math and reading test scores as we treat math and reading as separate domains and estimate a single impact within each domain. In doing so, we view the investigation of whether SES has an effect on math scores as a separate research question from whether SES has an effect on reading scores.

[6] As with an RCT, one of the assumptions involved in an RD analysis is the stable unit treatment value assumption (SUTVA): nonparticipating students must not be affected by the treatment. It is theoretically possible that nonparticipating students could have benefited from positive peer effects if SES had a large impact on participants, but given that participants constituted a relatively small proportion of students in their grade levels, the existence of peer spillovers seems unlikely.

[7] We analyze each grade in a school district as a separate mini-study even when a single cutoff value of the assignment variable is used across all grades. This is because the relationship between the outcome (a test score which varies by grade) and the assignment variable can vary across grades.

Power to Detect Effects. The study had an 80 percent chance of detecting an overall impact of SES of 0.12 standard deviations or greater (Table II.3); this is very close to the minimum detectable effect size of 0.14 that was planned for the study (as specified in the design report).[8] An impact of 0.12 standard deviations is equivalent to moving from the 50th percentile to the 55th percentile on a test with normally distributed scores.

Table II.3. Minimum Detectable Effect Sizes

Aggregate Impact Estimate	Standard Error	MDES	Smallest Possible Effect[a]
RD Mini-Studies	0.04	0.12	0.08

Source: School district records.

Note: The minimum detectable effect size value is expressed in effect size units and was calculated assuming (1) a two-tailed test; (2) a 0.05 percent significance level, α ; (3) an 80 percent level of power, β ; and (4) the estimated standard error of the RD impacts, $\hat{\sigma}_{impact}$ which was adjusted for clustering of students within unique values of the assignment variable (Lee and Card 2008). The impact was calculated using the following formula:

$$f\left(\alpha,\beta\right)*\hat{\sigma}_{impact}$$

where *f* is a function equal to the sum of two critical values (corresponding to α and β) from the normal distribution (the value of *fct* is approximately 2.8).

[a]The smallest possible effect is the minimum detectable effect size assuming a 50 percent level of power β .

C. Data Collected

The data collection has two distinct quantitative components that are integrated to answer the study's research questions. These components are (1) student-level data from the districts, used primarily for the regression discontinuity (RD) analysis; and (2) information from SES providers, used to describe provider characteristics and practices. In addition, we collected information from districts on prioritization procedures (for example, identifying the RD assignment variable). Table II.4 is a summary of the sources of data and how they are used.

Table II.4. Data Sources and Uses

Data	Source	How the Data Are Used
District prioritization procedures	Districts	Regression discontinuity design requires a continuous measure of assignment to services. These data ensure the design can be implemented.
Student-level information from application data and other school records	Districts	Test scores are the main outcome variable for this study. They are combined with descriptive information about SES enrollment in multivariate analyses regarding the success of SES programs.
SES provider survey	Providers	The provider survey data allow a description of the services and providers and make it possible to relate information about provider characteristics and practices to provider effectiveness.

Source: School district records and SES provider survey.

[8] The minimum detectable effect size (MDES) values reported in Table II.3 are based on the calculated standard errors from the impact estimates in this study, where impact estimates are aggregated using inverse variance weights.

Student-Level Data Collected from the Districts. We collected information from SES applications and from other school records provided by the participating school districts. These records have information about provider preferences (i.e., first-choice provider) and basic demographics, including each student's name, student identification number, gender, grade level, school, and SES provider choices (first, second, and third). We also collected information about student attendance at SES providers, including the number of hours of tutoring (by subject, where available) and the number of tutoring sessions attended. For the main outcome measure of this study, we collected the standardized test scores in math and reading in 2007, 2008, and 2009.[9] Though collected from districts, the tests used as outcomes are the same tests used for state accountability purposes.

Each district also provided data on the following student-level characteristics: eligibility for free or reduced-price lunch, race and ethnicity, English proficiency, and disability status.

The final analysis sample sizes used to estimate RD impacts equal the number of students in grades 3–8 who (1) met statutory requirements for participation, thus making them eligible for services, (2) were assigned to treatment and comparison groups using an RD design, and (3) who have at least one non-missing outcome variable (either a reading or math test score from 2009). Among the 50,843 students who had applied for SES across the six study districts, 30,673 students were in the study grades 3–8, of which 24,113 students were eligible for services, having met statutory requirements for participation, and were assigned to treatment and comparison groups (see Appendix A for additional details). Table II.5 shows the student sample sizes for each mini-study (i.e., for each district/grade level combination). RD impact estimates could not be calculated for some of the mini-studies because of inadequate sample sizes; the footnotes in Table II.5 provides additional details about these mini-studies. A formal analysis of missing student achievement data (i.e., students who do not have post-test scores) is provided in Appendix C. This analysis shows that for the study as a whole, the attrition rates for math test scores were four percent in the treatment group and two percent in the comparison group. Attrition rates for reading test scores were three percent in the treatment group and two percent in the comparison group. Although some mini-studies had attrition rates that were statistically significantly different for the treatment and control groups, the sensitivity analyses conducted in Appendix E show that the study's main findings are robust to the exclusion of those mini-studies.

We tested for baseline equivalence of the treatment and comparison group analysis samples by examining whether there is evidence of spurious RD impacts (which should not exist) on pre-intervention characteristics of students (including past test scores that were not used as the assignment variable). In other words, we estimated a regression of each pre-intervention characteristic on the assignment variable, separately on either side of the cutoff value.[10] The impact estimates were then calculated as the difference in intercept terms from the regressions on

[9] A limitation of this study is that the precise timing of services is unknown. The concern with the lack of timing information is that there is unknown variation in the lag between services and the tests used as outcome measures.

[10] The analysis was conducted using only data within the optimal bandwidth identified by an algorithm described in Imbens & Kalyanamaran (2009).

Table II.5. Student Sample Sizes, by District and Grade Level (21 Mini-Studies with Sufficient Sample Sizes Are in Bold)

	Number of Students that				
			AND		
Mini-Study Description	Applied to SES	AND Met Statutory Requirements for SES Participation and Were Assigned to Treatment and Comparison Groups	Have Non-Missing Math Outcomes	OR	Have Non-Missing Reading Outcomes
District 1 / grade 3	**320**	**282**	**265**		**264**
District 1 / grade 4	**251**	**251**	**233**		**222**
District 1 / grade 5	**184**	**184**	**160**		**154**
District 1 / grade 6	**150**	**150**	**131**		**130**
District 1 / grade 7	**108**	**108**	**98**		**95**
District 1 / grade 8	**97**	**97**	**89**		**90**
District 2 / grade 3	**495**	**305**	**276**		**276**
District 2 / grade 4	**491**	**394**	**367**		**366**
District 2 / grade 5[a]	485	393	358		359
District 2 / grade 6[a]	422	333	311		310
District 2 / grade 7[a]	352	281	254		254
District 2 / grade 8[a]	327	259	238		239
District 3 / grade 3[b]	220	0	n.a.		n.a.
District 3 / grade 4	**197**	**157**	**156**		**156**
District 3 / grade 5	**161**	**136**	**136**		**136**
District 3 / grade 6	**93**	**65**	**64**		**64**
District 3 / grade 7[c]	71	45	45		45
District 3 / grade 8[c]	72	46	46		46
District 4 / grade 3	**4,391**	**4,265**	**4,261**		**4,262**
District 4 / grade 4	**3,985**	**3,395**	**3,302**		**3,299**
District 4 / grade 5	**3,603**	**3,030**	**2,956**		**2,953**
District 4 / grade 6	**2,530**	**2,468**	**2,405**		**2,401**
District 4 / grade 7	**2,169**	**2,081**	**2,019**		**2,025**
District 4 / grade 8	**1,794**	**1,674**	**1,594**		**1,597**
District 5 / grade 3	**1,767**	**1,505**	**1,461**		**1,459**
District 5 / grade 4[b]	1,445	0	n.a.		n.a.
District 5 / grade 5[b]	1,273	0	n.a.		n.a.
District 5 / grade 6[b]	967	0	n.a.		n.a.
District 5 / grade 7[b]	656	0	n.a.		n.a.
District 5 / grade 8[b]	552	0	n.a.		n.a.
District 6 / grade 3	**969**	**907**	**878**		**880**
District 6 / grade 4	**710**	**672**	**650**		**650**
District 6 / grade 5	**669**	**630**	**605**		**611**
District 6 / grade 6[d]	Unknown	0	n.a.		n.a.
District 6 / grade 7[d]	Unknown	0	n.a.		n.a.
District 6 / grade 8[d]	Unknown	0	n.a.		n.a.

Source: School district records.

Note: The 21 mini-studies that had sufficient sample sizes to conduct the analysis are in bold.

[a]Impact estimates for these mini-studies could not be calculated because there was no comparison group as all students were below the cutoff point.

[b]Impact estimates for these mini-studies could not be calculated because the district did not use a cutoff point to assign students to receive services.

[c]Impact estimates for these mini-studies could not be calculated because of inadequate sample sizes.

[d]The school records from District 6 did not include any information for grades 6–8.

n.a. = not applicable.

the right and left sides of the cutoff value. Across all RD mini-studies, we conducted 213 tests for spurious impacts on pre-intervention student demographic characteristics and 76 tests for spurious impacts on pretest scores. For the overall study, the aggregate spurious impacts on the 2008 math and reading test scores across all mini-studies showed no spurious impacts. Specifically, the aggregate pre-test difference on 2008 math test scores was 0.02 and was not statistically significant (p-value of 0.80), and the aggregate pre-test difference on 2008 reading test scores was 0.02 and was not statistically significant (p-value of 0.78).

Outcome Measures. The outcome measures for analyzing the effects of SES were the 2009 math and reading standardized achievement test scores for the state accountability test in each district, collected from the district as part of the student-level data. The testing program and content standards for state accountability tests vary across states (Bandeira de Mello et. al, 2009). In order to aggregate effect size impact estimates across grade levels and districts, we standardized the outcome variables in each mini-study by subtracting the mean and dividing by the standard deviation. To ensure that differences in effect size impact estimates across mini-studies are due to differences in the effect of SES, rather than differences in the standard deviations of the outcomes, we use the state-grade-level means and standard deviations rather than the means and standard deviations from the mini-study samples. See May and Perez-Johnson (2009) for a broader discussion of issues related to the use of state test scores in evaluations.

SES Provider Survey. In June 2009, we distributed a mail survey to SES providers that served students in the six participating school districts during the 2008–2009 school year.[11] The sample for the data collection included 186 providers who had served students in the study.[12] As shown in Table II.6, 42 providers that served students in the analysis sample did not complete the survey. Thus, the final sample size for providers in the survey analyses was 144. The unweighted response rate was 77.4 percent and the student-weighted response rate was 88.8 percent. The survey methodology is described in detail in Appendix G.

Table II.6. Provider Sample Sizes

	Number	Percent
Served students in study sample	186[a]	
Did not complete the survey	42	23 percent
Completed the survey	144	77 percent

Source: SES provider survey.

[a]This number differs from the total number of providers listed in Table III.1. Providers that had not been identified by the districts, but were later determined to be providers, did not receive a survey but are included in the sample for the achievement analysis.

[11] None of the six study districts served as SES providers.

[12] Initially, there were 238 providers whose names were given by the districts. District records on SES participation were examined to identify providers who had served students in the study. As a result, 52 of the providers were excluded because they did not serve any of the students in the analysis sample (i.e., they may have served students in other grades).

D. Analytic Methods

In the remainder of this chapter, we present our benchmark analytic approaches to estimating the impacts of SES for the RD mini-studies. We then discuss our methods for estimating student subgroup impacts and conducting diagnostic analyses for RD mini-studies. Finally, we describe the sensitivity analyses used to test the robustness of the main impact findings to the choice of benchmark analytic methods.

Benchmark Approach. The main impact findings of this report (presented in Chapter IV) are based on the study team's benchmark approach for estimating program impacts. The study followed the What Works Clearinghouse (WWC) standards for regression discontinuity designs (U.S. Department of Education 2010). Our benchmark approach of estimating impacts using a linear functional form within an optimal bandwidth was selected before data analysis began. This approach was chosen because it performed better than alternative approaches in analyses of data from earlier studies and in simulations. In Appendix E, we describe alternative analytic methods and show the robustness of the study's findings to those methods. In contrast to an RCT, in which the validity of findings depends primarily on successful random assignment, the validity of findings from an RD study depends heavily on the statistical analysis, not just the design. Specifically, with an RD design, the relationship between the outcome and the assignment variable must be modeled appropriately in order to obtain rigorous impact estimates.

Our benchmark approach to estimation uses a linear regression with an appropriate bandwidth (as opposed to a linear regression using all data or an optimal polynomial regression using all data). Specifically, we estimate a linear regression on either side of the cutoff using only data within a bandwidth. That bandwidth was selected using the optimal bandwidth selection algorithm described in Imbens and Kalyanaraman (2009); we refer to our benchmark approach as the IK bandwidth approach.[13] Appendix B provides additional details on the RD benchmark approach and describes the method used for choosing the IK bandwidth.

Imputation of Missing Data. Missing covariate data were multiply imputed using the MIX package in R, which is based on algorithms from Schafer (1996). Imputations were conducted separately for the treatment and comparison groups, creating 10 multiply imputed datasets. All of the covariates (Z_i) included in the main RD impact regressions were included in the imputation models. RD impact estimates and standard errors were calculated separately for each of the 10 imputed datasets and, we then combined impact estimates and standard errors using the formulas in Rubin (1996).

For the benchmark analytic approach, missing data on outcomes were not imputed; rather, students with missing outcomes were excluded from the analysis. Appendix C provides additional details regarding student attrition from the study.

Subgroup Impact Estimates. We calculated impact estimates for subgroups of interest by applying the impact estimation methods described above separately for each subgroup. This approach allows for maximum flexibility in estimation, which may be particularly important in

[13] Calculating an impact within an optimally selected bandwidth is an evolution of the local linear estimator described in Hahn, Todd, & Van Der Klaauw (2001).

an RD study if the relationship between the outcome and assignment variable differs across subgroups. Subgroup impact estimates are aggregated across mini-studies using the same weighting scheme described above; each mini-study's impact estimate for a particular subgroup is weighted by the inverse of the variance of that impact estimate. Thus, it is possible that mini-studies received different weights in the main impact analysis and each subgroup analysis.

Depending on data availability and the characteristics of SES applicants in each mini-study, we estimated impacts for some of the key subgroups defined by NCLB as those for which Title I schools are expected to meet Adequate Yearly Progress (AYP) proficiency standards: racial/ethnic minorities, students with limited English proficiency (LEP), and students with disabilities. Finally, we estimated impacts for a subgroup that consists of the students in each school who are members of any groups that did not meet AYP targets in that school. In some schools this may consist of only one subgroup, but many schools have multiple subgroups (e.g., African-American students and special education students) falling short of AYP.

Diagnostic Analyses. We conducted diagnostic analyses pertaining to five key issues associated with RD designs (U.S. Department of Education 2010): (1) continuity of the assignment variable (see Appendix A), (2) integrity of the assignment variable (see Appendix C), (3) continuity of the relationship between the assignment variable and the outcome (Appendix C), (4) student attrition from the study sample (Appendix C), and (5) choice of an appropriate bandwidth and functional form (Appendix B and E).

Sensitivity Analyses. To test the sensitivity of the impact findings to our choice of benchmark analytic methods, we used two alternative analytic methods to estimate the impacts of SES on math and reading test scores. First, we estimated RD impacts with a linear regression using all of the data (instead of only data within a bandwidth).[14] We then estimated RD impacts with an optimal polynomial regression (instead of a linear regression) using all of the data. The sensitivity of findings to variations on the benchmark approach is summarized in Chapter IV and described in more detail in Appendix E.

[14] While we agree that this approach is not on as solid a theoretical foundation as the other approaches, we believe that it may be of interest to some readers who might argue that a linear functional form is plausibly correct when both the outcome and the assignment variable are measures of student achievement.

III. SES PROVIDER CHARACTERISTICS AND STUDENTS' SES EXPERIENCES

Based on the data from the provider surveys and student records (of the 187 providers that served students in the analysis sample, 144 providers completed the survey), we can describe the average SES experience in the six study districts.

- There were a few dominant providers in each district, with 9 providers across the six districts that served 50 percent of the students and 179 providers serving the other half. On average, 60 percent of providers' instructional staff (standard deviation of 40) was regular schoolteachers currently employed in the district.
- The average full course of SES, as reported by providers serving the six study districts, was about 28 hours (standard deviation of 11), lasting an average of 13 weeks, encompassing two sessions per week.
- Over half the providers (64 percent) offered services at the student's school, with 44 percent of providers offering sessions in groups of 2-5 and 34 percent in one-on-one sessions. Nearly all providers (99 percent) offered services in reading/language arts/English, and 87 percent offered services in math.
- On average, students received 21.2 hours of services (standard deviation of 8.8), with over a third of the students (36 percent) receiving tutoring in both reading and math, 55 percent receiving tutoring only in reading, and 9 percent receiving tutoring only in math.
- According to student records, 98 percent of students offered SES in these six districts were assigned to their first choice provider, and 86 percent of applicants who were offered SES received services.

In describing the providers (Section A), we include information about the number of providers in the study districts, and the characteristics of the organizations and their staffs. We describe the services providers offered in Section B, presenting provider reports on the grade levels served, the content areas covered, and duration and mode of their service. In Section C we provide information about activities that are required by regulation, including development of individual Student Learning Plans (SLPs), use of assessments, coordination with schools and curriculum alignment, and communication with parents. In Section D, we describe the demographic and academic characteristics of SES applicants and SES participants, the experiences of participants (specifically, the number of hours of math and reading services they received), and information about potential alternative services that were available to students in the study districts.

We describe the providers and their services using both weighted and unweighted data. For calculations based on unweighted data, each provider is given an equal weight to portray the characteristics and practices of the average provider. We also show student-weighted results to portray the characteristics and practices of the services received by the average SES participant. These results will differ in circumstances where the practices of large providers differ from those of small providers. The student-based weighting highlights providers that serve a large number of students, and the methodology is described in Appendix G.

A. SES Providers Serving Students

Although most study districts had a large number of SES providers, there were a few dominant providers in each district. Table III.1 shows the number of providers in each study district. The number of providers varied from 4 in District 1 to 61 in District 4 (column 1); only District 1 had fewer than 10 providers. The substantial number of providers in five of the six districts distinguishes these districts from other districts offering services: nationally, one-third of districts offering SES had four or fewer providers in 2005–2006 (Vernez et al. 2009).

Within each district there were a few dominant providers. To assess whether a minority of providers served a majority of students, we ranked (in descending order) providers in each district by the number of students served. The percentage of providers serving at least 50 percent of students ranged between 8 percent and 25 percent of the providers in each district. For the study as a whole, 9 out of 187 providers served at least 50 percent of the students.

Table III.2 shows the number of states and districts in which providers offered SES. While the majority of providers (60 percent) worked in just one state, 42 percent worked in more than five school districts. Thirty-nine percent of providers worked in more than one district and more than one state.

Provider Types and Experiences. The majority of providers (69 percent) were for-profit organizations (Table III.3) with a range of experience providing SES and other educational services in the study districts.[15] About one-fourth of providers (23 percent) were in their first year offering SES, while nearly a third (30 percent) had at least four years of experience.

Table III.1. Number of SES Providers by District

	Number of Providers That Served Students in the Analysis Sample	Number of Providers That Served Students in the Analysis Sample and Completed the Provider Survey	Percent of Students (in analysis sample) Served by Providers that Completed the Provider Survey
District 1	4	4	3.7%
District 2	37	31	10.1%
District 3	18	12	1.3%
District 4	61	45	71.8%
District 5	39	31	4.8%
District 6	28	21	8.3%
Total	**187[a]**	**144**	**100%**

Source: School district records and SES provider survey.

[a]This number differs from the total number of providers listed in Table II.6 and Appendix G, Table G.2 (186). Providers that had not been identified by the districts, but were later determined to be providers, did not receive a survey but are included in the sample for the achievement analysis.

[15] None of the six study districts served as SES providers.

Table III.2. Number of States and Districts in Which SES Providers Offered Services

	Unweighted Percentage of Providers
One State	60.4
2–5 states	20.8
More than 5 states	18.8
Total	100.0
One District	27.1
2–5 districts	31.3
More than 5 districts	41.7
Serves in One District/One State	26.6
Serves in more than one district/one state	34.3
Serves in more than one district/more than one state	39.2
Total Number of Providers	**144**

Source: SES provider survey.

Table III.3. Provider Organization Type

	Unweighted Percentage of Providers
For-profit	69.4
Nonprofit	11.1
Community-based	11.1
Faith-based	4.2
College or university	1.4
Public school	+
Private school	+
Other	1.4
Total Number of Providers	**144**

Source: SES provider survey.

+ Suppressed to protect respondent confidentiality.

Table III.4 shows the average provider's SES experience to be 2.9 years, with little variation across the districts. The means ranged from 2.5 years in District 1 to 3.2 years in District 2. District 1 providers served only as SES providers, but providers in other districts have offered other educational services to students in the district for an average of 3.6 years (range of 1.3 to 4.3 years), indicating that they were serving students in the district prior to doing so via the mechanism of SES. These experienced providers are now making their educational services accessible through SES.

Table III.4. Mean Number of Years Provided SES in This District and Mean Number of Years Provided Other Educational Services in This District (Standard deviations in parentheses)

	Mean Number of Years Provided SES in This District (Unweighted)	Mean Number of Years Provided Other Educational Services in This District (Unweighted)
District 1	2.5 (1.3)	0
District 2	3.2 (1.8)	4.3 (5.0)
District 3	2.9 (1.4)	1.3 (2.6)
District 4	2.6 (1.4)	4.0 (5.4)
District 5	3.1 (1.3)	3.8 (6.4)
District 6	2.9 (1.5)	3.5 (4.4)
Overall Average	**2.9 (1.5)**	**3.6 (5.2)**

Source: SES provider survey.

Staff Characteristics. Over a third of providers (37 percent) reported requiring tutors to be certified schoolteachers. As shown in Table III.5, providers indicated on the survey that those not requiring teacher certification reported a range of other minimum educational qualifications, such as two years of college (18 percent of providers), a four-year college degree related to content (13 percent of providers), any four-year degree (11 percent), or some other less-rigorous certification (22 percent).

Table III.5. Instructional Staff Certification Requirements

	Unweighted Percentage of Providers
Required to Be a Certified Schoolteacher	37.3
Minimum Requirement If Not Required to Be a Certified Schoolteacher	
4-year degree related to content	12.7
Any 4-year degree	10.6
2-year college/60 credit hours/associate's degree/paraprofessional certification	17.6
Other minimum requirement	21.8
Total Number of Providers	**142**

Source: SES provider survey.

Providers' descriptions of characteristics of tutoring staff were consistent with provider requirements: providers reported that, on average, 82 percent of their instructors were schoolteachers, including 60 percent currently employed in the district (Table III.6). When provider size is considered (using student-weighted data), 80 percent, of students had instructors who were teachers currently employed in the district, indicating that large providers hired more local teachers than small providers. The fact that a large majority of SES instructors in these districts were full-time classroom teachers in the districts may be of particular interest given that the districts themselves were not permitted to act as providers: federal policy precludes a district that is itself identified for improvement from acting as SES providers unless it applies for and receives a waiver from the Department.

Table III.6. Current Status of Instructional Staff (Standard deviations in parentheses)

	Unweighted	Student-Weighted
Mean percentage of instructors who are schoolteachers currently employed in the district	60.1 (39.7)	80.4 (30.4)
Mean percentage of instructors who are schoolteachers not currently employed in the district	21.6 (30.2)	9.6 (20.7)
Mean percentage of instructors who are enrolled as college students	6.8 (17.2)	5.7 (13.6)
Mean percentage of instructors who are enrolled as high school students	0.0 (0.4)	0.1 (0.8)

Source: SES provider survey.

Almost all providers (99 percent) reported training their staff. These providers reported having given an average of about 10 hours of training, though large providers gave somewhat less, as indicated by the student-weighted average of 7 hours (Table III.7). Providers whose staff had the least teaching experience (1–4 years) offered the most hours of training, on average (15.7 hours; see Table III.7).

Table III.7. Mean Hours of Training, by Years of Teaching Experience (Standard deviations in parentheses)

Teaching Experience	Unweighted Mean Hours of Training	Student-Weighted Mean Hours of Training
1–4 years	15.7 (24.1)	10.1 (8.5)
5–9 years	8.9 (8.6)	7.6 (6.7)
10–14 years	10.0 (6.4)	6.5 (3.9)
15+ years	7.0 (6.5)	5.9 (4.0)
Overall Average	10.3 (13.4)	7.0 (5.8)

Source: SES provider survey. n=136 providers.

B. Characteristics of Supplemental Educational Services Offered to Students

Provider services varied across grade levels, content areas, location, hours, duration, and a number of other dimensions. Almost all providers (96 percent) reported offering SES to students in primary elementary grades, while fewer served students at intermediate elementary (73 percent) and middle school (81 percent) grades.

Offered SES Content Areas. Nearly all providers (99 percent) offered services in reading/language arts/English, henceforth referred to simply as *reading* (Table III.8). Most (87 percent) also offered services in math. At all grade levels included in the study, reading was most widely offered, followed by math, then test-taking skills and study/organizational skills. A previous ED study conducted a survey of providers in 16 districts across the country and likewise found that providers were most likely to offer reading, followed by math, with other subjects far less common (Vernez et al. 2009). Student-weighted results are even more skewed toward reading and math, which suggests that smaller providers were more likely than large providers to offer tutoring in subject matter outside reading and math (see Appendix Table H.1).

Table III.8. Content Areas Offered by Grade Level

	Unweighted Percentage of Providers		
	Primary Elementary Grades (e.g., K-2)	Intermediate Elementary Grades (e.g., 3-5)	Middle School Grades (e.g., 6-8)
Reading/Language Arts/English	97.8	89.7	92.3
Mathematics	83.3	80.4	82.1
Writing	17.4	13.1	15.4
Homework help	21.0	17.8	19.7
Test-taking skills	43.5	41.1	46.2
Study/organizational skills	39.9	32.7	39.3
Other subjects	8.0	5.6	6.0
Any Content Area	99.3	92.5	94.0
Total Number of Providers	**138**	**107**	**117**

Source: SES provider survey.

Note: Grade configurations of schools varied across districts, making it impossible to define grade ranges that consistently divided school levels across districts. The survey therefore permitted providers to define grade ranges (Primary Elementary, Intermediate Elementary, and Middle School) as appropriate for the school district for which they were reporting.

Location of Offered Services. Table III.9 shows that about two-thirds of providers (64 percent) reported supplying SES in the school the student attended—a location that was even more prevalent among large providers: providers serving 90 percent of students were offering services at the students' schools. Smaller providers were also likely to offer their services in the students' homes (40 percent) and at their own facilities (27 percent). This result is consistent with prior work showing that students' schools were the sites most commonly used for services (Vernez et al. 2009).

Table III.9. Where SES Was Provided

	Unweighted Frequency	Unweighted Percentage	Student-Weighted Percentage
At the school of the served student	92	63.9	90.2
At student's home, in-person tutoring	58	40.3	27.7
At organization's facility	39	27.3	15.4
At other school district facilities	13	9.0	15.8
At student's home, online services	9	6.3	3.2
Some other location	71	50.0	46.0

Source: SES provider survey.

Format of Offered Services. Most providers (92 percent) reported that all services were provided face to face rather than by distance education (such as video, telephone or computer). Group size in SES sessions varied (Table III.10). Overall, providers reported that, on average, 34 percent of their services were in one-on-one sessions, 44 percent were in groups of 2 to 5 students, and 21 percent were in groups of 6 to 10. While fewer than a quarter of providers report *delivering* SES using groups larger than 5, the student-weighted data show that over half of students *receive* SES in groups larger than 5, indicating that larger providers offer larger group sizes (p-value 0.03). Large group instruction was reported as infrequent among all providers, with 97 percent saying they never offered instruction in groups of more than 10.

Table III.10. How SES Was Provided

	Unweighted	Student-Weighted
Mean Percentage Provided in One-on-One Sessions	34.3	18.5
Mean Percentage Provided in Groups of 2–5	43.8	29.4
Mean Percentage Provided in Groups of 6–10	20.9	51.2
Mean Percentage Provided in Groups of More than 10	1.0	0.8

Source: SES provider survey.

Length of Offered Services. A typical full course of SES, as reported by providers, was 13.4 weeks long (standard deviation 5.5). The average number of sessions per week was 2.1 (standard deviation 0.5), 2.1 (standard deviation 0.5), and 2.1 (standard deviation 0.6) for primary elementary, intermediate elementary, and middle school grades respectively. The average session length was 78.8 minutes (standard deviation 26.1), 83.3 minutes (standard deviation 28.3), and 84.1 minutes (standard deviation 27.4) for primary, intermediate, and middle grades. The average number of hours that a student could attend in a year was 28.4 (standard deviation 11.4), 28.0 (standard deviation 8.4) and 28.5 (standard deviation 11.8) for primary, intermediate, and middle grades (Table III.11). The length of a full course of services varied across providers, from 13 hours to 108 hours, with approximately 70% of providers reporting that a full course of services was 30 hours or less.

Table III.11. Times per Week a Typical Student Attended SES, Duration of Typical SES Session, Total Number of SES Hours a Student Could Attend per Year (Standard deviations in parentheses)

	Unweighted Percentage of Providers		
	Primary Elementary Grades	Intermediate Elementary Grades	Middle School Grades
Times per Week a Typical Student Attended SES			
1	8.7	5.0	10.0
2	76.8	79.2	71.8
3 or more			
Mean times per week	2.1 (0.5)	2.1 (0.5)	2.1 (0.6)
Total Number of Providers	**138**	**101**	**110**
Duration of Typical SES Session (Minutes)			
45-60	50.0	40.6	37.3
75-90	38.4	43.6	47.3
105-110	1.4	1.0	0.9
120-270	10.1	14.9	14.5
Mean minutes	78.8 (26.1)	83.3 (28.3)	84.1 (27.4)
Total Number of Providers	**138**	**101**	**110**
Total Number of SES Hours a Student Could Attend per Year			
20 or fewer	21.3	19.8	18.5
21–25	28.7	25.0	30.6
26–30	22.8	27.1	22.2
31+	27.2	28.1	28.7
Mean hours per year	28.4 (11.4)	28.0 (8.4)	28.5 (11.8)
Total Number of Providers	**136**	**96**	**108**

Table III.11 (continued)

	Unweighted Percentage of Providers		
	Primary Elementary Grades	Intermediate Elementary Grades	Middle School Grades
Mean Number of Weeks for Full Course of SES—All Grade Levels Combined			
4-10 weeks		30.0 (42)	
11-15 weeks		45.0 (63)	
16 or more weeks		25.0 (35)	
Mean		13.4 weeks (5.5)	

Source: SES provider survey.

The mean of about 28 hours constituting a full treatment of SES, based on provider reports (student-level data from district administrative records on hours attended are in Section B), is somewhat less than the average of 45 hours of SES annually from a prior study (Vernez et al. 2009).

C. SES Provider Requirements

SES regulations mandate that in the provision of services, all providers establish individualized student learning plans (SLPs), use assessments to measure student progress, coordinate with the students' schools and make sure the curriculum is aligned, and communicate with the parents. Through their survey responses, the SES providers offered a self-reported perspective on implementation of the regulations.[16]

Student Learning Plans. Student learning plans (SLPs) are written, systematic plans that describe instruction tailored to each student's individual needs. Nearly all providers (98 percent) reported that no students were served without first having an SLP in place. Indeed, in all participating districts, payment to providers was dependent on the existence of a parent-signed SLP.

Providers also reported who contributed to the SLP development in any way, ranging from substantive contributions to simply approving the plan (Table III.12). Three-fourths of providers (75 percent) reported that most parents played a limited role in the SLP process, approving the plan (which is required) but not contributing to its development. Nearly two-thirds of providers (69 percent) reported at least a minimal level of involvement by parents in developing or approving the SLP (that is, they reported that no parents had *no* involvement).

[16] Self-reports of implementation of the SES regulations may be subject to response bias, in particular the tendency for respondents to report in a socially desirable way, thus over-reporting adherence to regulations. Self-reports should be interpreted with caution.

Table III.12. Parent and Teacher Contribution to Development of the Student Learning Plan

	Unweighted Percentage of Providers Reporting Parental and Teacher Depth of Contribution		
	Contributed Substantively to Content of the Plan	Approved Plan but Did Not Contribute Substantively to Development	No Involvement in Developing or Approving Plan
Level of Parental Contribution (Percentage of Parents)			
Most (>50%)	18.7	74.5	4.5
Some (25%–50%)	15.8	13.1	3.8
Few (<25%)	50.4	6.6	22.6
None	15.1	5.8	69.2
Total Number of Providers	**139**	**137**	**133**
Level of Teacher Contribution (Percentage of Teachers)			
Most (>50%)	15.2	20.9	33.8
Some (25%–50%)	17.4	11.2	8.8
Few (<25%)	28.3	22.4	16.9
None	39.1	45.5	40.4
Total Number of Providers	**138**	**134**	**136**

Source: SES provider survey.

Providers reported that teachers were more likely than parents to have no involvement in the SLP process. That is, two-fifths of the providers reported that none of their students' teachers had any involvement in developing or approving the SLP (60 percent of providers reported at least a minimal level of involvement by teachers). The student-weighted data (in Appendix Table H.2) show that large providers were more likely to report plan approval or substantive contribution by some teachers and less likely to report that teachers had no involvement at all with the SLPs.

Using Assessments. SES providers are required to measure student's progress (see J-1 of 2009 non-regulatory guidance; district agreement with provider 34 CFR 200.46 [b][2]). Specifically, they are to describe how the student's progress will be measured, a timetable for improving achievement, and procedures for regularly informing the student's parents and teachers of the student's progress. Therefore, we collected information on provider access to and use of student assessments.

Providers had access to some information about their SES students before they began services, and they reported administering assessments a number of times over the course of providing services. Nearly two-thirds of providers (65 percent) reported having previous state assessment scores for all or nearly all of their students prior to the start of tutoring (Table III.13).

Table III.13. Percentage of Participating Students for Which Organization Received State Assessment Scores, Report Card Grades, or Student Work Prior to Providing SES

	Unweighted Percentage of Providers		
	State Assessment Scores Received	Report Card Grades Received	Examples of Student Work Received
Received for…			
All or nearly all	64.5	24.8	8.0
More than half	12.8	7.1	3.6
About half	5.0	3.5	2.2
Less than half	8.5	7.1	9.4
None	9.2	57.4	76.8
Total Number of Providers	**141**	**141**	**138**

Source: SES provider survey.

The providers typically conducted their own baseline assessments: nearly all (85 percent) reported conducting diagnostic assessments of *all* students at the outset of providing services. In addition to the baseline assessment, 80 percent of providers reported conducting assessments after the tutoring services were completed, and nearly two-thirds (63 percent) said they administered tests between the pre- and post-service assessments (Table III.14). This is consistent with previous findings that providers administer multiple assessments of their own (Vernez et al. 2009). The average number of assessments reported was 4.3 and 4.1 per year for reading and math, respectively, for a provider's typical SES student.

Table III.14. Timing and Number of SES Student Assessments (Standard deviations in parentheses)

	Unweighted Percentage of Providers
Timing of SES Student Assessments	
Used prior scores provided by district	66.2
Administered tests "pre" services	82.4
Administered tests "post" services	78.9
Administered one or more tests between "pre" and "post"	63.4
Number of Reading/Language Arts/English Assessments Conducted per Year	
None	5.8
1	3.6
2	39.6
3	17.3
4	16.5
5 or more	17.3
Total Number of Providers	**139**
Number of Math Assessments Conducted per Year	
None	5.0
1	2.5
2	39.7
3	19.0
4	15.7
5 or more	18.2
Total Number of Providers	**121**
	Unweighted
Mean number of reading/language arts/English assessments conducted per year	4.3 (8.9)
Mean number of math assessments conducted per year	4.1 (8.0)

Source: SES provider survey.

Alignment with School Curriculum. To understand efforts to align SES curriculum with the regular school curriculum, we asked providers to identify the important considerations in choosing or developing curriculum and materials for SES instruction and to choose the single most important consideration. Table III.15 shows that state standards were reported as a major consideration. Most providers (94 percent) selected it as an important consideration, and nearly three-fourths (72 percent) selected it as the single most important. Providers claimed that their services were closely aligned with state standards: on a 10-point scale, the provider-reported mean level of alignment to state standards was 9.5 (standard deviation 0.9).

Table III.15. Important Considerations in Providers' Decisions About Choosing or Developing a Curriculum

	Unweighted Percentage of Providers	
	Important Considerations (Mark All That Apply)	Single Most Important Consideration
State Standards	94.4	72.1
Nationally Recognized Standards	82.5	16.4
District Curriculum Materials	68.5	3.6
Common Standards Across States in Which Provider Operates	53.1	3.6
Other	14.7	4.3
Total Number of Providers		**140**

Source: SES provider survey (unweighted percentages with unweighted frequencies in parentheses).

Providers also claimed alignment to the district's instructional program, reporting a mean level of alignment to district standards at 9.1 (standard deviation 1.2) on a 10-point scale. On the one hand, hiring local school teachers and maintaining regular communications with the students' teachers could potentially contribute to alignment. On average, 60 percent of the SES instructors were teachers currently employed in the district (Table III.6). Nearly half of the providers (49 percent, Appendix Table H.3) reported communicating "at least once a month" with teachers about student progress related to classroom instruction, and over a third of providers (39 percent, Appendix Table H.3) reported communication "at least once a month" with teachers about upcoming academic topics to be covered in school (again, somewhat less so for large providers, with monthly communication reported by a student-weighted 24 percent of providers). However, providers reported limited teacher involvement in SLP development (Table III.12—on average, 15 percent of the providers reported that more than half of participants' teachers "contributed substantively to the content of the plan").

As an alternative perspective about alignment with district instructional programs, we asked providers about the curriculum for two grades and two subjects (3rd grade reading and math and 7th grade reading and math) in the districts they served. We compared their answers to what the districts reported and found that providers were rarely able to identify the specific curriculum used in the district's schools. Table III.16 shows the percentage selecting the correct answers. Summarizing across districts, nearly three-fourths of providers (71 percent) failed to correctly identify even one of the four curriculum programs. Only 3.5 percent correctly identified all four curricula. The larger providers were more likely to name the district's curriculum correctly with 25 percent of students served by providers that correctly identified all four (Appendix Table H.4). Even so, providers serving nearly half (48 percent) of the participating students incorrectly reported all four curricula or reported not knowing the curricula in the district.

Table III.16. Did the SES Provider Know the District Curriculum?

	Unweighted Percentage of Providers
Correctly Selected at Least One Publisher Title	
Grade 3 reading	16.7
Grade 3 math	20.1
Grade 7 reading	8.3
Grade 7 math	10.4
Summary Across All Districts	
One correct	13.9
Two correct	7.6
Three correct	4.2
Four correct	3.5
All wrong or don't know	63.9
All missing	6.9
Total Number of Providers	**144**

Source: SES provider survey.

Parent Communication. Under NCLB, the SES program was designed to empower parents in selecting services to improve their children's learning; communication between providers and parents could therefore be an important feature. Table III.17 shows that providers reported regular communication with parents using various methods of communication. Email communication was least common (13.5 percent of providers reported communicating once a month or more), while communications by telephone, by regular mail, in person, and by notes were more common (the percentage that reported communicating once a month or more using these other modes ranged from 52.8 to 72.1). Larger providers were less likely than smaller ones to use in-person communication (student-weighted percentage reporting in-person communication at least once a week was half the unweighted percentage: 17 versus 34 percent, p-value < 0.01, Appendix Table H.5).

Table III.17. Frequency of Communication to Parents about Child's Progress, by Mode

	Unweighted Percentage of Providers				
	Telephone	Email	In Person	Notes Sent Home with Students	Postal Mail or Other Mode
At Least Once a Week	10.6	3.5	34.3	8.6	4.9
A Few Times a Month	21.8	4.3	18.2	17.9	11.3
At Least Once a Month	26.1	5.7	19.6	35.0	36.6
A Few Times a Year	29.6	16.3	21.7	22.1	26.8
Did Not Use/Do	12.0	70.2	6.3	16.4	20.4
Total Number of Providers	**142**	**141**	**143**	**140**	**142**

Source: SES provider survey.

Providers reported that bad contact information was the most common challenge to successful communication with parents (87 percent of providers identified as a moderate or serious challenge, Table III.18).

Table III.18. Moderate and Serious Challenges to Successful Communication with Parents

	Unweighted Frequency	Unweighted Percentage	Student-Weighted Percentage
Language Barriers			
Moderate challenge	20	13.9	17.2
Serious challenge	6	4.2	10.4
Work Schedule of Parent			
Moderate challenge	56	39.2	44.0
Serious challenge	15	10.5	25.7
Work Schedule of Instructional Staff			
Moderate challenge	5	3.5	1.2
Serious challenge	0	0.0	0.0
Bad Contact Information for Parents			
Moderate challenge	45	31.3	18.3
Serious challenge	80	55.6	77.5
Parent Lack of Email Access			
Moderate challenge	38	26.6	46.7
Serious challenge	44	30.8	33.4
District Discourages Communication with Parents			
Moderate challenge	3	2.1	0.3
Serious challenge	0	0.0	0.0

Source: SES provider survey.

D. Student Experiences with SES

In this section we describe the students who participated in SES, including who was selected to receive SES, how many and to what extent students participated, and the demographic and academic characteristics of SES applicants and SES participants. We then provide some information on the potential experiences of comparison students who did not participate in SES.

SES Student Applicants and Participants. Across the six study districts, 50,843 students applied to SES. Sixty percent of them (30,673) were in study grades 3-8. Of those, 24,113 (79 percent) met statutory requirements for participation. Among those eligible, a total of 19,750 students (the treatment group, 82 percent of the analytic sample) were offered SES and 4,363 (the comparison group, 18 percent of the sample) were not offered SES. Of the 24,113 students that were assigned to treatment and comparison groups, 755 of them had missing math outcomes and 770 had missing reading outcomes (see Table II.5 for these calculations). Among those who were offered services, 16,954 (or 86 percent of those offered services) received tutoring services (that is, hours of service were greater than 0), and 2,796 (or 14 percent of those offered services) did not participate in SES. Almost all students (98%) who were offered SES in these districts were assigned to their first-choice provider. We define a student as having participated in SES if he or she received any tutoring services. Participation rates for individual districts are described in Appendix Table A.1.

Table III.19 shows average demographic and academic characteristics for SES participants and nonparticipating applicants, including gender, race/ethnicity, English-language learner (ELL) status, disability status, special education status, grade level, and baseline test scores (standardized in each of the three states where our study districts are located to have a mean of 0 and standard deviation of 1). The distribution of characteristics for participants is similar to that

for nonparticipating applicants. About half of SES participants are male (52 percent), 5 percent are white, 47 percent are black, and 46 percent are Hispanic. ELL students account for 16 percent of participants, and students with a disability (those in special education) were 21 percent. The average grade level of participants was 4.6, the average standardized math test score at baseline was -0.57, and the average standardized reading test score at baseline was -0.69. All results are student-weighted rather than district-weighted (each student counts equally across districts).

Table III.19. Average Characteristics of SES Applicants and Participants

Characteristic	Nonparticipating Applicants	Participants
Male	54%	52%
White	6%	5%
Black	44%	47%
Hispanic	39%	46%
English-Language Learners (ELLs)	20%	16%
Disability/Special Education	12%	21%
Grade Level	4.7	4.6
Baseline Math Score (Standardized)	-0.39	-0.57
Baseline Reading Score (Standardized)	-0.48	-0.69

Source: School district records.

Note: Participants are defined as students who received any tutoring services (hours >0).

According to non-regulatory guidance from the Department (see C-31, C-32 of 2009 non-regulatory guidance), districts are to make services and necessary accommodations available to eligible students with disabilities and LEP students. However, districts do not necessarily have to meet these requirements with each provider. For this reason, we conducted some descriptive analysis to shed light on the SES experiences of these two subgroups of SES students. First, within each district, we tested whether the variation across providers in the percentage of ELL and SPED students served was greater than would be expected to occur by chance. For Districts 2 and 4, we found that the variation across providers in the concentration of ELL students was more than would be expected by random chance, consistent with these districts meeting the requirement with a portion of the providers. For District 4, we found that the variation across providers in the concentration of SPED students was more than would be expected by chance. In other districts, participating ELL and SPED students appeared to be distributed randomly across providers.

Second, we assessed whether the percentages of ELL and SPED students served by a provider were associated with any provider characteristics found in the provider survey, consistent with districts' requirement to make accommodations for the needs of these subgroups. We found no evidence to support this hypothesis. Specifically, within the districts in which the variation across providers was more than would be expected by chance, we tested for differences in 179 provider characteristics between providers with low and high concentrations of ELL or SPED students, in an effort to assess whether particular characteristics are associated with increased or reduced likelihood of serving either group of students. Results are shown in Appendix A, Tables A.10–A.12. We found no evidence that any particular provider

characteristic is related to the proportion of ELL or SPED students served. Across the 179 tests in each district for ELL and SPED, we found some statistically significant relationships, but in both districts, the number of significant results was no greater than would be expected to occur by chance.[17]

Hours of Service Received. Students received fewer average hours of service than providers offered. Recall from Section B that a complete course of services was about 28 hours, on average, according to providers (see Table III.11). School district records (typically derived from bills submitted by providers based on student attendance at services) indicated that the mean hours of student SES participation for the overall study was 21.2 (standard deviation 8.8), and this did not differ significantly across student subgroups. There was variation across districts in the mean number of tutoring hours provided, which ranged from 16.5 (standard deviation 5.2) in District 3 to 41.3 (standard deviation 14.3) in District 1.

For the study overall, 36 percent of SES participants (6,058) received tutoring in both math and reading, 9 percent (1,560) received tutoring only in math, and 55 percent (9,336) received tutoring only in reading. For students who received tutoring in math, mean math hours were 12.5 (standard deviation 8.2). For students who received tutoring in reading, mean reading hours were 17.2 (standard deviation 9.2).

Time spent in SES was only a small fraction of the students' time spent in school. To compare the average SES dosage with classroom instruction time in reading and math, suppose students spent one hour a day on reading and one hour a day on math as part of their regular curriculum. This would imply that students spent about 180 hours on math and 180 hours on reading in a school year. Thus, if a student received 12.5 hours of SES math instruction (the mean hours received), it would equal approximately 7 percent of a full year of math instruction. Likewise, if a student received 17.2 hours of reading instruction, it would equal approximately 10 percent of a full year of reading instruction.

Potential Experiences of Comparison Students: Alternatives to SES. Alternative after-school programs were available in all districts, but public information about them is limited, and most districts were unable to provide information on whether students participated in these alternatives, which typically operate independently of district supervision. Based on discussions with school districts, we know that alternative services, such as those supported by 21st Century Learning Centers or YMCA/Boys and Girls Clubs, often included an academic component. We also know that several districts provided lists of alternative tutoring providers to parents of students not selected for SES. Some providers on these lists offered free services; others charged a fee. A web search revealed additional non-SES after-school programs in several districts that had not been identified by the districts.

[17] Specifically, three were statistically significant in District 2 for ELL students (shown in Appendix A, Table A.10), four were statistically significant in District 4 for ELL students (shown in Table A.11), and four were statistically significant in District 4 for SPED students (shown in Table A.12).

IV. ESTIMATED IMPACTS OF SES

The primary focus of this evaluation is to assess the potential benefits of offering SES to more applicants in districts not able to serve all eligible students due to limited resources. The study examines the impacts of SES on students' math and reading test scores for those students at the cutoff for services in six school districts where more eligible students applied for SES than could be served with available funds. We found no statistically significant impacts on math or reading test scores for students at the cutoff for being offered SES (defined as the set of students with assignment variable values in the vicinity of the cut point). In the rest of this chapter we describe our findings in detail. First we present impact findings using our benchmark analytic approach, and describe the robustness of the findings to changes in our analytic approach and to particular subgroups of students. We then summarize the results of exploratory analyses conducted to address the study's second and third research questions.

A. No Statistically Significant Impacts on Math or Reading Test Scores

Table IV.1 presents the overall intent-to-treat (ITT) impact estimates, which estimate the effect of SES for students in the study's school districts at the cutoff for being offered SES on students' math and reading test scores. Offering SES to students does not indicate placement in and receipt of services. In other words, these impact estimates ignore whether or not students *participated* in SES, how many tutoring hours participants received, and whether those hours were in math or reading. Section B of this chapter presents the treatment-on-the-treated (TOT) impact estimates, which are the ITT estimates rescaled to adjust for the difference in *participation* rates between the treatment and comparison groups. Section C summaries analyses that examine whether impact estimates varied based on the *number of service hours* provided in a specific subject.

Table IV.1. Overall Intent to Treat (ITT) Impact Estimates of SES on Math and Reading Test Scores on Students Near the Cutoff for Services

	Adjusted Mean Outcomes		ITT Impact Estimate (Effect Size Units)	
Outcome	Treatment	Comparison		*p*-Value
Math test scores	0.20	0.15	0.05	0.25
Reading test scores	0.18	0.21	-0.03	0.48

Source: Math and reading test scores from school district records.

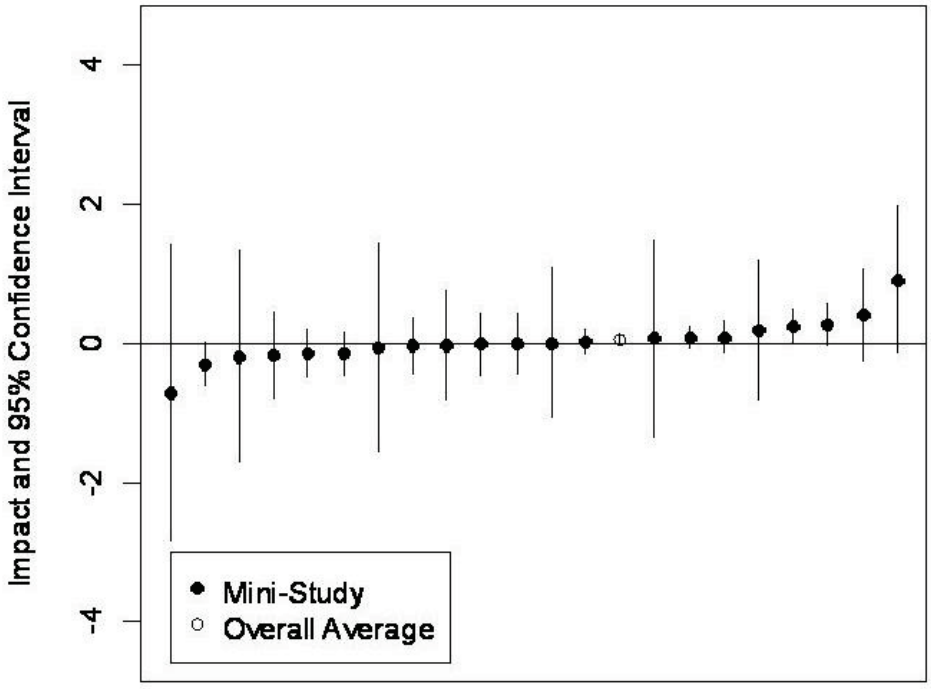
Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cut point/grade combination. All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect size units. Adjusted mean outcomes for the treatment and comparison groups are equal to the estimated intercept terms from the regressions of the outcome (student test scores) on the assignment variable and other covariates that were estimated separately on either side of the cutoff value. The outcomes were standardized using state-grade-level means and standard deviations. Aggregate impact estimates are calculated as a weighted average of the impact estimates for each mini-study, where the weight is equal to the inverse of the variance of that mini-study's impact estimate. For each RD mini-study, the ITT impact estimate is calculated using a linear functional form within an optimal bandwidth selected using the Imbens & Kalyanaraman (2009) algorithm, and standard errors are adjusted to account for clustering of observations within unique values of the assignment variable, as in Lee & Card (2008).

*Significantly different from zero at the 0.05 level, two-tailed test.

There were no statistically significant impacts on math or reading test scores of students at the cutoff for being offered services when aggregating across mini-studies using the variance-minimizing weight described in Chapter II. Table IV.1 shows the adjusted mean outcomes for the treatment and comparison groups; these are simply the predicted values of the outcome at the RD cut point for the students to the left (treatment) and right (comparison) of the cut point, represented by the intercept terms in equations (5) and (6) in Appendix B.

The ITT impact estimate is the difference between these two intercept terms. The overall math impact estimate, aggregated across all RD mini-studies, is 0.05 standard deviations and is not statistically significant (*p*-value = 0.25). The overall reading impact estimate is -0.03 standard deviations and is not statistically significant (*p*-value = 0.48). These are estimates of the effect of SES for students in the vicinity of the cutoff value of the assignment variable—those on the cutoff for being offered SES. Because the study consists of a large number of mini-studies, impacts are estimated at a diversity of cutoff values, increasing the generalizability of the findings beyond a single assignment variable and a single cutoff value. In Appendix D, we report that the variation in impact estimates across mini-studies is no larger than one would expect by chance, meaning that there is no systematic variation in impact estimates across the diversity of cutoff values in the study. Appendix A describes the various assignment variables used in the study and presents figures showing the cutoff values for each assignment variable.

Figure IV.1. Estimated Impacts on Math Test Scores on Students Near the Cutoff for Services, with 95 Percent Confidence Intervals, in Effect Size Units



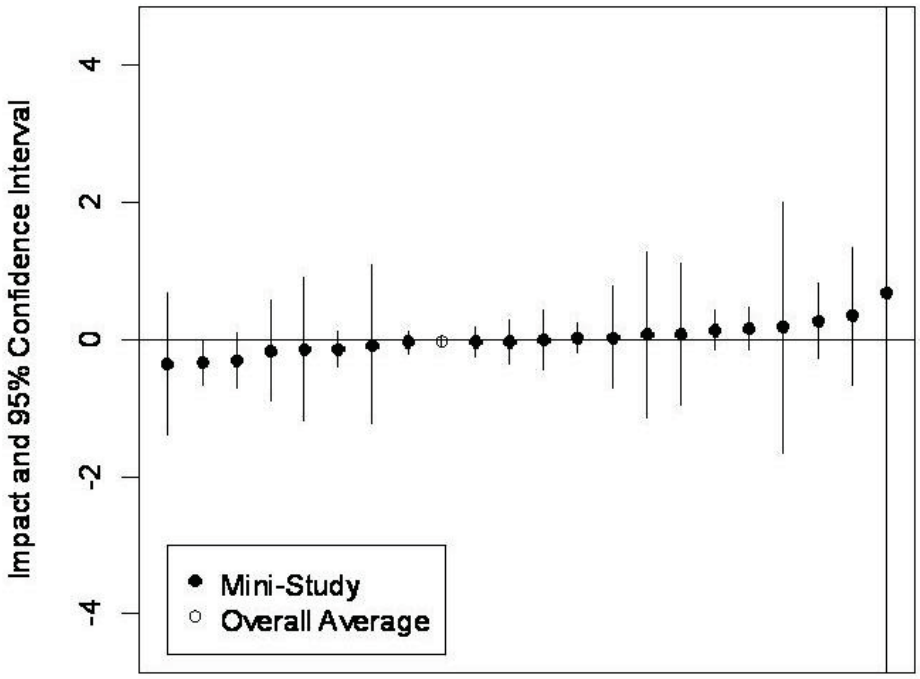
Source: School district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cut point/grade combination. Twenty-one of the mini-studies correspond to math outcomes and 21 correspond to reading outcomes. This figure shows the distribution of ITT impact estimates of SES on math test scores by mini-study. Each dot in the figure represents one of the 21 mini-studies with 95 percent confidence intervals indicated by vertical lines running through each impact estimate. For each mini-study, the ITT impact estimate was calculated using a linear functional form within an optimal bandwidth selected using the Imbens & Kalyanaraman (2009) algorithm, and standard errors were adjusted to account for clustering of observations within unique values of the assignment variable, as in Lee & Card (2008). All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect size units.

Figures IV.1 and IV.2 show the distribution of ITT impact estimates on math and reading test scores (respectively) by mini-study; each dot in the figures represents one mini-study. Impact estimates are plotted on the vertical axis and 95 percent confidence intervals are indicated by vertical lines running through each impact estimate. The overall aggregate estimate (0.05 for math and -0.03 for reading) is indicated by a white dot included in each figure. See Appendix F for ITT impact estimates presented separately by mini-study.

Figure IV.3 shows aggregated ITT impact estimates separately for each district. Impact estimates are plotted on the vertical axis and 95 percent confidence intervals are indicated by vertical lines running through each impact estimate. None of the district-specific impact estimates are statistically significant.

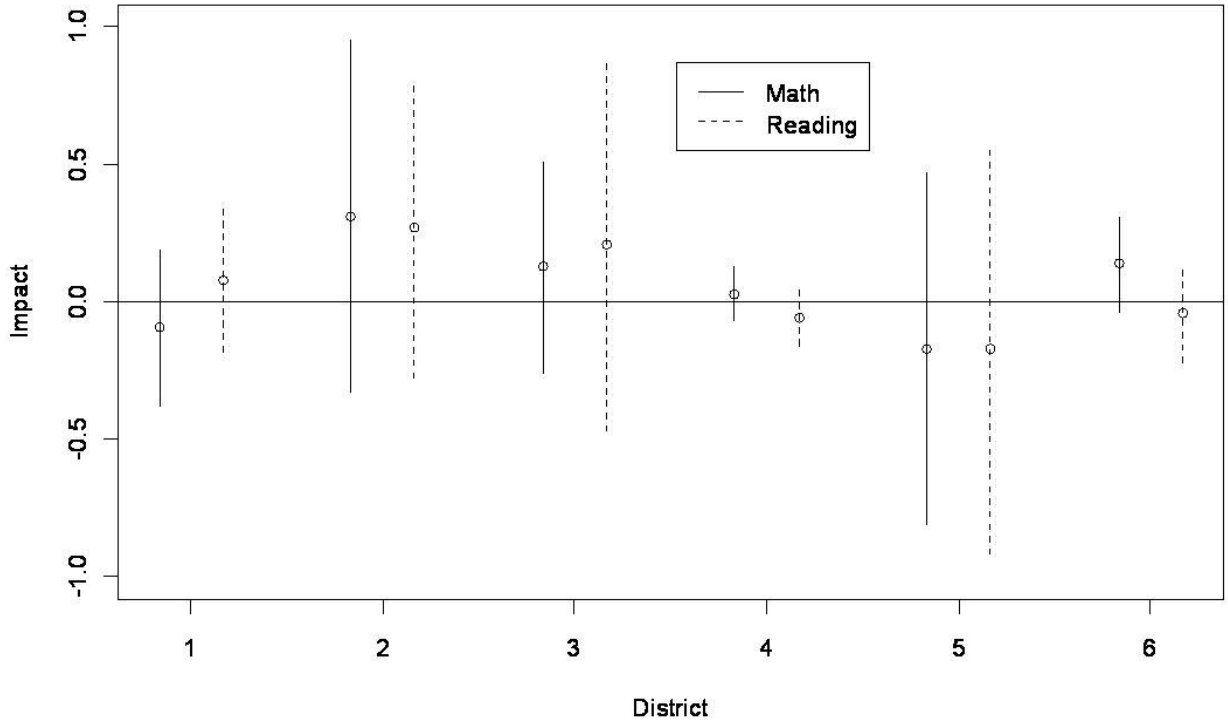
Figure IV.2. Estimated Impacts on Reading Test Scores on Students Near the Cutoff for Services, with 95 Percent Confidence Intervals, in Effect Size Units



Source: School district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cut point/grade combination. Twenty-one of the mini-studies correspond to math outcomes and 21 correspond to reading outcomes. This figure shows the distribution of ITT impact estimates of SES on reading test scores by mini-study. Each dot in the figure represents one of the 21 mini-studies with 95 percent confidence intervals indicated by vertical lines running through each impact estimate. The 21st mini-study had a confidence interval larger than the vertical scale of the figure. For each mini-study, the ITT impact estimate was calculated using a linear functional form within an optimal bandwidth selected using the Imbens & Kalyanaraman (2009) algorithm, and standard errors were adjusted to account for clustering of observations within unique values of the assignment variable, as in Lee & Card (2008). All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect size units.

Figure IV.3. District-Specific Impact Estimates on Students Near the Cutoff for Services, in Effect Size Units



Source: School district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cut point/grade combination. Twenty-one of the mini-studies correspond to math outcomes and 21 correspond to reading outcomes. This figure shows aggregated ITT impact estimates separately for each district. Impact estimates are plotted on the vertical axis and 95 percent confidence intervals are indicated by vertical lines running through each impact estimate. All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect size units.

B. Sensitivity, Subgroup, and SES Participation Analyses

The impact findings presented in this chapter are robust to changes in our analytic approach. Specifically, we estimated impacts using different weighting methods, applying different functional forms, imputing missing outcome values, excluding covariates, and excluding districts with high levels of differential attrition.[18] Appendix E describes the sensitivity analyses in detail, and Appendix Table E.1 presents results from these analyses. None of the overall ITT impact estimates on math and reading test scores are statistically significant, and the size of the impact estimates ranged from -0.05 to 0.09 across all the sensitivity analyses.

We also conducted an alternate analysis that examined only the four Florida districts. Recall from Chapter II that Florida districts had a stronger incentive than districts in other states to ensure that as many students as possible were participating in SES as a result of statewide rules regarding unspent SES funds—rules that are similar to new national regulatory guidance from

[18] We examine the sensitivty of impacts to excluding distircts with high levels of differential attrition because a large difference in attrition rates could lead to biased impacts due to unobserved differences in the treatment and comparison groups, although removing districts from the study as the result of an impact finding could also lead to bias. We show in Appendix Table E.1 that excluding districts with high rates of differential attrition does not change the sign or significance of impacts, suggesting that the attrition issues are not affecting our findings.

the U.S. Department of Education. The Florida results were consistent with the overall results: Neither the reading or math average impact estimates for students on the cutoff for services across the four Florida districts were statistically significant.

We examined whether offering SES to more applicants in oversubscribed districts would potentially benefit student subgroups that are targeted by NCLB. Consistent with Title I's long-maintained focus on improving the educational outcomes of at-risk students, NCLB mandated that each school meet AYP standards for a variety of different subgroups of students, including racial/ethnic minorities, English-language learners, and students with disabilities. Students in a school become eligible for SES if the school falls short of AYP for three consecutive years for any one or more of these subgroups, even if the schoolwide population as a whole is meeting the proficiency target. Improving outcomes for these subgroups is thus an explicit goal of NCLB.

Out of 12 comparisons of impacts for student subgroups, only one was statistically significant: the difference in math impact estimates for special education/non-special education students (-0.25 for special education students, 0.05 for non-special education students, p-value of 0.01 on the difference). A similar difference was not observed for the same subgroup in reading, where estimated impacts were -.09 for special education students and -0.05 for non-special education students. Moreover, one difference out of 12 is about what we would expect to see as a result of random chance.

More details on the analyses for the Florida districts and student subgroups are included in Appendix D.

Finally, we examined impacts of *participation* in SES on students on the cutoff for receiving services; this is called a treatment-on-treated (TOT) impact analysis. As described in Chapter III, not all applicants whose values of the RD assignment variable are below the cutoff value actually participate in SES. Across the six study districts, 14 percent of applicants below the cutoff on the RD assignment variable did not participate in SES. It is therefore useful to estimate impacts for the subset of students who not only are below the cutoff value but actually attended services. The TOT analysis relies on a "fuzzy" RD method and has potential to support causal inferences under appropriate assumptions.[19] Details are provided in Appendix D. The overall TOT impact estimate on reading test scores is -0.10 and not statistically significant (p-value = 0.25) and the TOT impact estimate on math test scores is 0.11 and not statistically significant (p-value = 0.21). In sum, we do not find that participation in SES improved student achievement, just as we do not find that offering SES improved student achievement.

C. Exploratory Findings

In this section we summarize findings from exploratory analyses conducted to answer the study's second research question, regarding the association between estimated impacts and

[19] The assumptions are those needed for a valid "instrument" (as in instrumental variables analysis) and are described in Angrist, Imbens, & Rubin (1996). Briefly, the assumptions are that: (1) the relationship between the instrument and program participation is monotonic, (2) the correlation between the instrument and program participation is sufficiently large, and (3) that the only effect of the instrument on outcomes is through program participation (that is, not through some other unobserved factor that is correlated with program participation).

characteristics of services, providers, or practices. The correlational analyses were conducted for the purposes of developing hypotheses and identifying potential considerations for program improvement, and they cannot support causal inferences. Additional information on these analyses and detailed results are provided in Appendix D.

For these correlational analyses, we examined a variety of provider characteristics that could be used to select providers and monitor their services. As noted in Chapter III, we found variation across providers in many of the characteristics, including hours of service, provider size, tutoring practices, and instructional qualifications.

Prior to running the correlational analyses, we first examined whether the variation in impact estimates across providers was greater than one would expect by chance. We observed no more variation in impact estimates across providers than what we would expect to see by chance. Specifically, we conducted a test for the homogeneity of impact estimates and failed to reject the null hypothesis of impact homogeneity (math impact p-value = 0.94; reading impact p-value = 0.88).

The inability to confirm that any of the providers produced statistically significant achievement impacts does not bode well for an exercise of correlating estimated impacts with provider characteristics. Nonetheless, the failure to reject the test of impact homogeneity across providers does not tell us that there is no true variation in impacts—only that we could not confirm the existence of variation. Prior research has found that the homogeneity test can have low power to detect true heterogeneity in impacts (Huedo-Medina et al. 2006).[20] We therefore proceeded to examine whether characteristics of services (e.g., intensity of services) or providers (e.g., provider size, instructional staff qualifications, tutoring practices) were related to impacts.

None of these relationships were stronger than what we would expect to see by random chance alone. Detailed results are provided in Appendix D.

[20] The homogeneity test that we used likely has even less power than the Q-statistic test because the Q-statistic test assumes normality of the individual impacts while our test accounted for the small sample sizes of many of the provider impacts and treated the impacts as following the *t*-distribution with appropriate degrees of freedom for each impact.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Angrist, J., Imbens, G., & Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*, 444–472.

Bandeira de Mello, V., Blankenship, C., & McLaughlin, D.H. (2009). Mapping State Proficiency Standards Onto NAEP Scales: 2005-2007 (NCES 2010-456). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Black, A. R., Somers, M.-A., Doolittle, F., Unterman, R., & 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.

Bloom, H. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8*(2), 225–246.

Center on Education Policy. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act.* Washington, DC: Center on Education Policy.

Cooper, H. M., Hedges, L. V., & Valentine, J. (Eds.). (2009). *The handbook of research synthesis and meta-analysis (2nd ed.)*. New York: The Russell Sage Foundation.

Gill, B., Lockwood, J. R., III, Martorell, F., Setodji, C. M. & Booker, K. (2009, August). *An exploratory analysis of adequate yearly progress, identification for improvement, and student achievement in two States and three cities.* Washington, DC: U.S. Department of Education, Policy and Program Studies Service.

Hahn, J., Todd, P., & Van Der Klaauw, W. (2001). Regression discontinuity. *Econometrica, 69*(1), 201– 209.

Hallgren, K., Deke, J., & Angus, M. H. (2007). *Design considerations for an evaluation of supplemental education services.* Princeton, NJ: Mathematica Policy Research, Inc.

Heckman, J., Smith, J., & Taber, C. (1994). *Accounting for dropouts in evaluations of social experiments* (NBER Working Paper No. 166). Cambridge, MA: National Bureau of Economic Research.

Heinrich, C. J., Meyer, R. H., & Whitten, G. (2010). Supplemental education services under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis*, *32(2),* 273–298.

Huedo-Medina, T. B., Sanchez-Meca, J., Botella, J., & Marin-Martinez, F. (2006). Assessing Heterogeneity in Meta-Analysis: Q Statistic or I^2 Index? *Psychological Methods*, 11(2), 193–206.

Imbens, G., & Kalyanaraman, K. (2009, February). *Optimal bandwidth choice for the regression discontinuity estimator* (NBER Technical Working Paper Series: Technical Working Paper 14726). Cambridge, MA: National Bureau of Economic Research.

Imbens, G. & Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142 (2), 615–635.

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142, 655–674.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics*, 122(1), 159–208.

May, H., Perez-Johnson, I., Haimson, J., Sattar, S. & Gleason, P. (2009). *Using state tests in education experiments: A discussion of the issues* (NCEE 2009-013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714.

Muñoz, M. A., & Ross, S. M. (2009). *Supplemental education services as a component of No Child Left Behind: A mixed-methods analysis of its impact on student achievement.* Jefferson, KY: Accountability, Research, and Planning Department, Jefferson Public Schools.

Ross, S. M., Neergaard, L. L., Harrison, L., Ford, J., & Paek, J. (2009). *Implementation and outcomes of supplemental education services: The 2007-2008 Tennessee state-wide evaluation study.* Memphis, TN: Center for Research on Education Policy, The University of Memphis.

Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91, 473–489.

Schafer, J. L. (1996). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Springer, M. G., Pepper, M. J., & Ghosh-Dastidar, B. (2009a). *Supplemental education services and student test score gains: Evidence from a large, urban school district.* Paper presented at the *NCLB: Emerging Findings Research Conference* at the Urban Institute, Washington, DC on August 12, 2009.

Springer, M. G., Pepper, M. J., Gardner, C. D., & Bower, C. B. (2009b). "Supplemental educational services under No Child Left Behind," in *Handbook of Research on School Choice,* Berends, M., Springer, M. G., Ballou, D,. & Walberg, H. J., eds. Routledge.

Stullich, S., Abrams, A., Eisner, E., & Lee, E. (2009). *Title I implementation–update on recent findings.* Washington DC: U.S. Department of Education, Office of Planning, Evaluation and Development, Policy and Program Studies Service.

Trochim, W. M. K. (1984). *Research design for program evaluation.* Beverly Hills, CA: Sage Publications.

U.S. Government Accountability Office. (2006). *No Child Left Behind Act: Education actions needed to improve local implementation and state evaluation of supplemental education services.* Washington, DC: U.S. GAO.

Vernez, G., Naftel, S., Ross, K. E., Le Floch, K. C., Beighley, C., & Gill, B. (2009). *State and local implementation of the No Child Left Behind Act: Volume VII—Title I school choice and supplemental education services: Final report.* Washington DC: U.S. Department of Education, Office of Planning, Evaluation and Development, Policy and Program Studies Service.

Wood, S. N. (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99, 673–686.

Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood, J. R. III. (2007). *State and local implementation of the No Child Left Behind Act: Volume I—Title I school choice, supplemental education services, and student achievement.* Washington DC: U.S. Department of Education, Office of Planning, Evaluation and Development, Policy and Program Studies Service.

APPENDIX A

SES ASSIGNMENT AND PARTICIPATION

This appendix presents detailed information on the study's sample sizes, the methods used in each district to assign students to treatment and comparison groups, the cutpoints and bandwidths used in the regression discontinuity (RD) analyses, the hours of student SES participation, and the RD "impacts" on participation by district and by mini-study. The first section provides student sample sizes by school district. The second section describes the assignment procedures in each district. The third section describes our method for choosing cutpoints and presents the cutpoints and bandwidths used in each RD mini-study. The fourth section presents treatment and comparison group sample sizes by mini-study. The fifth section presents hours of student SES participation by district and by mini-study. The sixth section presents RD "impacts" on participation.

A. Student Sample Sizes by School District

Table A.1 shows the student sample sizes by school district, including how many students applied to SES, were offered SES, and participated in SES. We define a student as having participated in SES if he or she received any tutoring services (that is, hours of service were greater than 0). Not all students who were selected to receive services participated. For the overall study, 30,673 students applied to SES and were in tested grades. Of that number, 24,113 met statutory requirements for SES participation and were assigned to treatment and comparison groups using a continuous measure of prior achievement, 19,750 were offered SES (4,363 were not offered SES), and 16,954 participated in SES (2,976 did not participate in SES).

B. Assignment Procedures

In this section, we describe the methods used by each study district to determine who is offered SES. We also note that certain groups of students were excluded from the analysis, and we list the reasons why they were excluded.

District 1. The assignment variable and cutpoint differed by grade group: the first group included only kindergarteners, the second group included grades 1 and 2, the third group included grade 3, and the fourth group included grades 4–8. The assignment variables for the first three groups were grade-specific, non–high-stakes reading tests. The assignment variable for grades 4–8 was the 2008 composite score on the state's high-stakes test. Only grades 3–8 were included in our analysis; earlier grades did not have outcome data on the high-stakes test. Cutpoints were chosen in the manner described in Section D of this appendix.

District 2. Students were first grouped according to grade and proficiency category. Within each group, students were then sorted by their 2008 reading score and then their 2008 math score; scores came from either the state's high-stakes test, or another test, depending on grade level. The groups were then stacked in the order listed below and each student was given a rank that served as the assignment variable; ranks ranged from 1 to 3855. The stacked groups and associated assignment variable values were:

- 8th graders who failed both math and reading: 1–157
- 7th graders who failed both math and reading: 158–343
- 6th graders who failed both math and reading: 344–625
- 5th graders who failed both math and reading: 626–852

- 4th graders who failed both math and reading: 853–1086
- 3rd graders who failed both math and reading: 1087–1230
- 2nd graders who failed both math and reading: 1231–1310
- 1st graders who failed both math and reading: 1311–1390
- 8th graders who failed either math or reading: 1391–1484
- 7th graders who failed either math or reading: 1485–1548
- 6th graders who failed either math or reading: 1549–1620
- 5th graders who failed either math or reading: 1621–1744
- 4th graders who failed either math or reading: 1745–1853
- 3rd graders who failed either math or reading: 1854–2079
- 2nd graders who failed either math or reading: 2080–2159
- 1st graders who failed either math or reading: 2160–2164
- 8th graders who were proficient in math and reading: 2165–2227
- 7th graders who were proficient in math and reading (proficient 7th graders were split into two groups): 2228–2247
- Students with severe disabilities who took alternate assessments: 2248–2321
- Kindergarteners who were not on track: 2322–2563
- 7th graders who were proficient in math and reading (proficient 7th graders were split into two groups): 2564–2628
- 6th graders who were proficient in math and reading: 2629–2679
- 5th graders who were proficient in math and reading: 2680–2790
- 4th graders who were proficient in math and reading: 2791–2926
- 3rd graders who were proficient in math and reading: 2927–3088
- 2nd graders who were proficient in math and reading: 3089–3312
- 1st graders who were proficient in math and reading: 3313–3676
- Kindergarteners who were on track: 3677–3855

The district used a single cutpoint to divide students into those who were offered SES and those who were not. We inferred the cutpoint in the manner described in Section D of this appendix. Only grades 3 and 4 were used in the analysis; all other grades fell completely to one side or the other of the chosen cutpoint, leaving no comparison group.

Before conducting the RD impact analyses for grades 3 and 4, we transformed the district-assigned ranks for each grade into a continuous ranking that ranged from one to the number of students in the analysis. The reason for this was to avoid "banding" of the assignment variable values. For example, the district-assigned ranks for students in grade 3 range from 1087–1230, 1854–2079, and 2927–3088, leaving gaps from 1231–1853 and 2080–2926. We assigned new

ranks that did not contain any gaps, but that kept students in the same order as the district-assigned ranks. This was done to facilitate modeling of the relationship between the outcome and the assignment variable.

District 3. The assignment variable for grades 3–11 was the 2008 score on the state's high-stakes reading test. We inferred grade-specific cutpoints using the procedure described in Section D of this appendix. All third graders were assigned to receive SES, leaving no comparison group; thus, only grades 4–8 were used in the analysis.

District 4. Assignment to SES was based on a single district-wide ranking of all students in the district (not just those who applied to SES). A student's rank was determined by prior test scores and other characteristics, such as free or reduced-price lunch (FRPL) status. A single cutpoint was used to divide students into those who were offered SES and those who were not. We inferred the cutpoint using the procedure described in Section D of this appendix. Only grades 3–8 were included in our analysis; other grades did not have outcome data on the high-stakes test.

We excluded some data far from the cutpoint because they were sparse and were indicative of a relationship between the outcome and the assignment variable that was fundamentally different from the relationship nearer to the cutpoint (this is the only district that experienced this issue). We also assigned new ranks that ranged from one to the number of students in the analysis. The reason for this was to avoid "banding" of the assignment variable values. Specifically, the district-assigned ranks contained large gaps when the sample was restricted to a single grade. Thus, for each grade, we assigned new ranks that did not contain any gaps, but that kept students in the same order as the district-assigned ranks.

District 5. Assignment procedures differed by grade. Only third-grade students were assigned to SES based on an RD design. The assignment variable was the 2008 score on a diagnostic reading test. We inferred the cutpoint using the procedure described in Section D of this appendix.

District 6. The assignment variable for students in grades K–5 was the 2008 test score transformed into a z-score within each grade; scores came from either the state's high-stakes test or another test, depending on grade level. All students who were retained in their grade from last year were automatically offered SES, regardless of their assignment variable value. We are unable to identify these students in the data, so they could not be removed from the analysis (this issue does not apply to other districts in the study). Although removing them from the analysis might have improved statistical power, failing to remove them does not bias our impact estimates because these students still have valid assignment scores (that is, this is a compliance issue, not a manipulation issue). We inferred the cutpoint using the procedure described in Section D of this appendix. Only grades 3–5 were included in our analysis; earlier grades did not have outcome data on the high-stakes test.

C. Sample Sizes by District and Mini-Study

For each district, Table A.2 shows the number of students on either side of the cutpoint, and the number of students selected by the district for the treatment and comparison groups. The last two columns show the proportion of students on either side of the cutpoint who attended SES. For the overall study, 82 percent of students were assigned to treatment (that is, they were

offered SES) and 18 percent were assigned to the comparison group (and were not offered services). Among the treatment group, 80 percent of students participated (that is, received at least one hour of services). Among the comparison group, 4 percent of students participated and 96 percent did not. Overall, only 16 percent of students did not comply with treatment status.

Table A.3 presents similar statistics for each mini-study and shows the total number of study-eligible SES applicants for each mini-study. An applicant is study-eligible if he or she met statutory requirements for SES participation and was assigned to the treatment or control group using an RD design.

D. Cutpoints and Bandwidths for RD Mini-Studies

In all districts, the cutpoints were the result of the rule-based application process playing out. The rule in that process that makes the RD approach possible is that districts were supposed to serve students with the greatest need (as measured by past achievement) first. Prior to receiving student applications, no district knew what the ultimate cutoff would be. The ultimate cutoff was a function of the distribution of need among applicants, which was only known after all applications were received. Because districts had no reason to systematically record what the ultimate cutpoint was (districts were not conducting an RD impact study), we calculated these cutpoints ourselves for each assignment variable in each district. The cutpoint was calculated as the value that maximizes the difference in "selection" rates between students below and above the cutpoint. A student was defined as being selected for services if either (1) the district assigned him or her to a specific SES provider or (2) he or she received any tutoring services (hours > 0).

For each mini-study, Table A.4 shows the number of students and the number of unique values of the assignment variable that fall on either side of the cutpoint, both overall and within the bandwidth that was chosen for the RD analysis (see Appendix B for a description of how this bandwidth was selected). For District 4, the total numbers of students differ from the numbers shown in Table A.3 because Table A.4 excludes some data far from the cutpoint that were excluded from the analysis because they were sparse and were indicative of a relationship between the outcome and the assignment variable that was fundamentally different from the relationship nearer to the cutpoint.

E. Hours of Student SES Participation by District and Mini-Study

Table A.5 shows that the mean hours of student SES participation for the overall study was 21.2 hours. Table A.5 also shows that there was substantial variation across districts in the average number of tutoring hours provided; mean hours ranged from 16.5 in District 3 to 41.3 in District 1.

Table A.6 shows hours of student SES participation by subject area (reading versus math). For students who received tutoring in math, mean math hours were 12.5. For students who received tutoring in reading, mean reading hours were 17.2.

Table A.7 shows the hours of student SES participation by mini-study. There was substantial variation across mini-studies in the number of hours received by students. Specifically, the mean number of tutoring hours received by SES participants ranges from 14 (for the mini-studies in District 3, grades 7 and 8) to 44 (for the mini-studies in District 1, grade 3). For District 4, the

total number of participants (11,960) differs from the number shown in Table A.1 (11,961) because Table A.7 excludes some data far from the cutpoint that were excluded from the analysis.

F. RD Impact Estimates on Participation

To illustrate that there truly was a jump at the cutpoint in the probability of a student receiving SES, we estimated RD impacts on selection and participation indicators, using the same benchmark methods that were used to estimate impacts on test scores. A student was defined as being selected for services if either (1) the district assigned him or her to a specific SES provider or (2) he or she received any tutoring services (hours > 0). A student was defined as a participant if he or she received any tutoring services (hours > 0).

Table A.8 presents RD impact estimates on selection and participation by district and for the overall study; Table A.9 presents similar results by mini-study. Impact estimates are aggregated across all RD mini-studies using the variance-minimizing weight described in Chapter II. For the overall study, the RD impact estimate on the probability of being selected to receive SES was 0.55. In other words, the jump at the cutpoint in the predicted probability of being selected was 0.55. For the overall study, the RD impact estimate on the probability of participating was 0.57; this is the jump at the cutpoint in the predicted probability of participating in SES.

Figures A.1 through A.22 show a plot of the participation indicator against the assignment variable for each mini-study. In some mini-studies, such as the one portrayed in Figure A.18 (District 5, grade 3), most of the students below the cutpoint participated in SES and most of the students above the cutpoint did not. Thus, the jump in the predicted probability of participating at the cutpoint is fairly high (0.88). In other mini-studies, such as the one portrayed in Figure A.17 (District 4, grade 8), many students below the cutpoint did not receive SES. Because of this, the jump at the cutpoint in the predicted probability of participating in SES is much lower (0.06).

Table A.1. Student Sample Sizes, by District

School District	Number of Students That Applied to SES	Are in Study Grades (3–8)	Met Statutory Requirements for SES Participation and Were Assigned to Treatment and Comparison Groups Using Prior Achievement	Were Offered SES	Participated in SES
1	2,008	1,110	1,072	672	551
2	3,968	2,572	1,965	1,932	1,583
3	2,059	814	449	333	272
4	28,638	17,169	16,913	14,102	11,961
5	9,777	6,660	1,505	989	882
6	4,393	2,348	2,209	1,722	1,705
Total	**50,843**	**30,673**	**24,113**	**19,750**	**16,954**

Source: School district records.

SES = supplemental education services.

Table A.2. Treatment and Comparison Group Sample Sizes, by District

District	Sample Sizes by Cutpoint		Sample Sizes by Selection		Proportion that Attended SES	
	Treatment	Comparison	Treatment	Comparison	Treatment	Comparison
1	650	422	672	400	0.78	0.11
2	1,852	113	1,932	33	0.81	0.75
3	406	43	333	116	0.66	0.07
4	15,111	1,802	14,102	2,811	0.79	0.00
5	1,045	460	989	516	0.84	0.02
6	1,850	359	1,722	487	0.92	0.00
Total	**20,914**	**3,199**	**19,750**	**4,363**	**0.80**	**0.04**

Source: School district records.

Note: The first two columns show the number of students on either side of the cutpoint. The second two columns show the number of students selected by the district for the treatment and comparison groups. Some of the districts did not adhere strictly to the cutpoint when assigning students to receive services; that is, some students below the cutpoint were not selected for services and some students above the cutpoint were selected. The last two columns show the proportion of students on either side of the cutpoint who attended SES.

SES = supplemental educational services.

Table A.3. Treatment and Comparison Group Sample Sizes, by Mini-Study

	Sample Sizes Based on Cutpoint		Sample Sizes Based on Selection		Proportion that Attended SES	
Mini-Study Description	Treatment	Comparison	Treatment	Comparison	Treatment	Comparison
District 1 / grade 3	197	85	184	98	0.82	0.00
District 1 / grade 4	146	105	144	107	0.86	0.00
District 1 / grade 5	96	88	132	52	0.57	0.41
District 1 / grade 6	97	53	93	57	0.81	0.02
District 1 / grade 7	63	45	61	47	0.76	0.02
District 1 / grade 8	51	46	58	39	0.73	0.15
District 2 / grade 3	289	16	302	3	0.85	0.81
District 2 / grade 4	297	97	369	25	0.82	0.74
District 2 / grade 5[a]	393	0	390	3	0.79	NA
District 2 / grade 6[a]	333	0	333	0	0.85	NA
District 2 / grade 7[a]	281	0	+	+	0.78	NA
District 2 / grade 8[a]	259	0	259	0	0.75	NA
District 3 / grade 4	150	7	110	47	0.61	0.00
District 3 / grade 5	124	12	99	37	0.69	0.00
District 3 / grade 6	46	19	41	24	0.59	0.05
District 3 / grade 7[b]	43	+	41	+	0.79	0.00
District 3 / grade 8[b]	43	3	42	4	0.70	0.67
District 4 / grade 3	3,917	348	3,838	427	0.88	0.00
District 4 / grade 4	3,148	247	3,023	372	0.83	0.00
District 4 / grade 5	2,648	382	2,490	540	0.83	0.00
District 4 / grade 6	2,126	342	1,926	542	0.72	0.00
District 4 / grade 7	1,833	248	1,584	497	0.68	0.00
District 4 / grade 8	1,440	234	1,242	432	0.65	0.00
District 5 / grade 3	1,045	460	989	516	0.84	0.02
District 6 / grade 3	722	185	680	227	0.93	0.00
District 6 / grade 4	575	97	535	137	0.92	0.00
District 6 / grade 5	553	77	507	123	0.91	0.00

Table A.3 *(continued)*

Source: School district records.

Note: The first column of this table presents the total number of study-eligible SES applicants for each mini-study. An applicant is study-eligible if he or she met statutory requirements for SES participation and was assigned to the treatment or control group using a regression discontinuity design. The next two columns show the number of students on either side of the cutpoint. The next two columns show the number of students selected by the district for the treatment and comparison groups. Some of the districts did not adhere strictly to the cutpoint when assigning students to receive services; that is, some students below the cutpoint were not selected for services and some students above the cutpoint were selected. The last two columns show the proportion of students on either side of the cutpoint that attended SES.

[a]Impact estimates for these mini-studies could not be calculated because there was no control group (that is, all students were on the left side of the cutpoint).

[b]Impact estimates for these mini-studies could not be calculated because of inadequate sample sizes.

+Suppressed to protect confidentiality.

SES = supplemental educational services.

NA = not available.

Table A.4. Assignment Variable and Bandwidth Descriptions, by Mini-Study

		Number of Students, Overall		Number of Unique Values, Overall		Number of Students in Bandwidth		Number of Unique Values in Bandwidth	
Mini-Study Description	Assignment Variable	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
District 1 / grade 3 / math test scores	Test Score	166	116	15	14	158	78	11	9
District 1 / grade 3 / reading test scores	Test Score	166	116	15	14	130	67	9	6
District 1 / grade 4 / math test scores	Test Score	146	105	140	81	113	60	107	58
District 1 / grade 4 / reading test scores	Test Score	146	105	140	81	87	53	84	51
District 1 / grade 5 / math test scores	Test Score	96	88	94	63	64	45	62	44
District 1 / grade 5 / reading test scores	Test Score	96	88	94	63	70	51	68	50
District 1 / grade 6 / math test scores	Test Score	97	53	95	44	50	36	48	36
District 1 / grade 6 / reading test scores	Test Score	97	53	95	44	54	38	52	38
District 1 / grade 7 / math test scores	Test Score	63	45	62	35	45	31	44	31
District 1 / grade 7 / reading test scores	Test Score	63	45	62	35	36	28	35	28
District 1 / grade 8 / math test scores	Test Score	51	46	51	36	46	33	46	32
District 1 / grade 8 / reading test scores	Test Score	51	46	51	36	35	25	35	24
District 2 / grade 3 / math test scores	Ranking	289	16	289	16	15	15	15	15
District 2 / grade 3 / reading test scores	Ranking	289	16	289	16	16	16	16	16
District 2 / grade 4 / math test scores	Ranking	297	97	297	97	96	96	96	96
District 2 / grade 4 / reading test scores	Ranking	297	97	297	97	115	97	115	97
District 2 / grade 5 / math test scores[a]	Ranking	393	0	393	0	NA	NA	NA	NA
District 2 / grade 5 / reading test scores[a]	Ranking	393	0	393	0	NA	NA	NA	NA
District 2 / grade 6 / math test scores[a]	Ranking	333	0	333	0	NA	NA	NA	NA
District 2 / grade 6 / reading test scores[a]	Ranking	333	0	333	0	NA	NA	NA	NA
District 2 / grade 7 / math test scores[a]	Ranking	281	0	281	0	NA	NA	NA	NA
District 2 / grade 7 / reading test scores[a]	Ranking	281	0	281	0	NA	NA	NA	NA
District 2 / grade 8 / math test scores[a]	Ranking	259	0	259	0	NA	NA	NA	NA
District 2 / grade 8 / reading test scores[a]	Ranking	259	0	259	0	NA	NA	NA	NA
District 3 / grade 4 / math test scores	Scaled Test Score	150	7	106	6	69	7	46	6
District 3 / grade 4 / reading test scores	Scaled Test Score	150	7	106	6	35	5	29	4
District 3 / grade 5 / math test scores	Scaled Test Score	124	12	87	12	37	10	22	10
District 3 / grade 5 / reading test scores	Scaled Test Score	124	12	87	12	38	10	23	10
District 3 / grade 6 / math test scores	Scaled Test Score	46	19	38	17	31	18	24	16
District 3 / grade 6 / reading test scores	Scaled Test Score	46	19	38	17	26	14	21	12
District 3 / grade 7 / math test scores[b]	Scaled Test Score	43	+	38	+	NA	NA	NA	NA
District 3 / grade 7 / reading test scores[b]	Scaled Test Score	43	+	38	+	NA	NA	NA	NA
District 3 / grade 8 / math test scores[b]	Scaled Test Score	43	3	34	3	NA	NA	NA	NA
District 3 / grade 8 / reading test scores[b]	Scaled Test Score	43	3	34	3	NA	NA	NA	NA
District 4 / grade 3 / math test scores	Ranking	3,917	113	3,917	113	108	108	108	108
District 4 / grade 3 / reading test scores	Ranking	3,917	113	3,917	113	111	111	111	111
District 4 / grade 4 / math test scores	Ranking	3,148	83	3,148	83	124	83	124	83
District 4 / grade 4 / reading test scores	Ranking	3,148	83	3,148	83	87	83	87	83
District 4 / grade 5 / math test scores	Ranking	2,648	131	2,648	131	161	131	161	131
District 4 / grade 5 / reading test scores	Ranking	2,648	131	2,648	131	182	131	182	131
District 4 / grade 6 / math test scores	Ranking	2,126	144	2,126	144	342	144	342	144
District 4 / grade 6 / reading test scores	Ranking	2,126	144	2,126	144	264	144	264	144
District 4 / grade 7 / math test scores	Ranking	1,833	68	1,833	68	64	64	64	64

Table A.4 *(continued)*

Mini-Study Description	Assignment Variable	Number of Students, Overall		Number of Unique Values, Overall		Number of Students in Bandwidth		Number of Unique Values in Bandwidth	
		Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
District 4 / grade 7 / reading test scores	Ranking	1,833	68	1,833	68	186	68	186	68
District 4 / grade 8 / math test scores	Ranking	1,440	63	1,440	63	60	60	60	60
District 4 / grade 8 / reading test scores	Ranking	1,440	63	1,440	63	123	63	123	63
District 5 / grade 3 / math test scores	Scaled Test Score	1,045	460	247	263	400	328	162	161
District 5 / grade 3 / reading test scores	Scaled Test Score	1,045	460	247	263	363	303	148	146
District 6 / grade 3 / math test scores	Z-score	722	185	118	49	355	158	36	33
District 6 / grade 3 / reading test scores	Z-score	722	185	118	49	342	153	34	32
District 6 / grade 4 / math test scores	Z-score	575	97	42	8	310	94	11	6
District 6 / grade 4 / reading test scores	Z-score	575	97	42	8	280	86	10	5
District 6 / grade 5 / math test scores	Z-score	553	77	42	9	290	73	15	7
District 6 / grade 5 / reading test scores	Z-score	553	77	42	9	160	65	9	5

Source: School district records.

Note: For each mini-study, this table shows the number of students on either side of the cutpoint, both overall and within the bandwidth. It also shows the number of unique values of the assignment variable on either side of the cutpoint, both overall and within the bandwidth. For District 4, the number of students overall differs from the numbers shown in Table A.3 because this table (Table A.4) excludes some data far from the cutpoint that were excluded from the analysis. The last four columns contain "NA" when the number of study-eligible students was too small to calculate a bandwidth.

[a]Impact estimates for these mini-studies could not be calculated because there was no control group (that is, all students were on the left side of the cutpoint).

[b]Impact estimates for these mini-studies could not be calculated because of inadequate sample sizes.

+ Suppressed to protect confidentiality.

C = control; T = treatment.

NA = not available.

Table A.5. Hours of Student SES Participation, by District

District	Number of Participants	Hours of Participation		
		Minimum	Mean	Maximum
1	551	1.0	41.3	101.0
2	1,583	0.3	28.5	107.5
3	272	1.5	16.5	29.2
4	11,961	1.0	19.4	41.0
5	882	1.0	23.5	63.0
6	1,705	1.0	20.6	43.0
Total	**16,954**	**0.3**	**21.2**	**107.5**

Source: School district records.

SES = supplemental education services.

Table A.6. Hours of Student SES Participation, by Subject Area and District

District	Hours of Math Services[a]			Hours of Reading Services[b]		
	Minimum	Mean	Maximum	Minimum	Mean	Maximum
1	N/A	N/A	N/A	1.0	41.3	101.0
2	0.3	19.8	87.8	0.5	19.3	107.5
3	1.0	10.3	20.0	0.5	15.7	29.2
4	1.0	10.4	39.0	1.0	15.2	40.0
5	0.8	13.9	31.5	0.8	20.5	43.5
6	1.0	19.8	43.0	1.0	20.4	43.0
Total	**0.3**	**12.5**	**87.8**	**0.5**	**17.2**	**107.5**

Source: School district records.

Note: District 1 did not provide tutoring in math.

[a]Excluding students with zero math hours.

[b]Excluding students with zero reading hours.

N/A = not applicable. SES = supplemental education services.

Table A.7. Hours of Student SES Participation, by Mini-Study

Mini-Study Description	Number of Participants	Hours of Participation		
		Minimum	Mean	Maximum
District 1 / grade 3	162	5.2	43.7	84
District 1 / grade 4	125	3.5	42.6	64
District 1 / grade 5	91	4	41.3	101
District 1 / grade 6	80	1	40.7	74.7
District 1 / grade 7	49	2	39.3	58
District 1 / grade 8	44	1.8	31.5	53
District 2 / grade 3	260	1.8	31.8	87.8
District 2 / grade 4	317	1	30.8	107.5
District 2 / grade 5[a]	310	0.3	28.7	90.5
District 2 / grade 6[a]	282	1	26.4	76.5
District 2 / grade 7[a]	220	1	26.6	92.5
District 2 / grade 8[a]	194	0.9	25.0	69
District 3 / grade 4	92	3.5	18.0	29.2
District 3 / grade 5	86	2	16.9	29.2
District 3 / grade 6	28	2	17.1	20
District 3 / grade 7[b]	34	3	13.7	20
District 3 / grade 8[b]	32	1.5	13.7	20.5
District 4 / grade 3	3,438	1	20.1	40
District 4 / grade 4	2,614	1	19.8	40
District 4 / grade 5	2,186	1	19.9	41
District 4 / grade 6	1,533	1	18.4	41
District 4 / grade 7	1,250	1	18.0	40
District 4 / grade 8	939	1	17.8	40
District 5 / grade 3	882	1	23.5	63
District 6 / grade 3	675	1	20.8	43
District 6 / grade 4	527	1	21.0	43
District 6 / grade 5	503	1	20.0	43

Source: School district records.

Note: The total number of participants reported in this table for District 4 (11,960) is smaller than the total number of participants reported in Table A.1 for District 4 (11,961) because this table (Table A.7) excludes some data far from the cutpoint that were excluded from the analysis.

[a]Impact estimates for these mini-studies could not be calculated because there was no control group (that is, all students were on the left side of the cutpoint).

[b]Impact estimates for these mini-studies could not be calculated because of inadequate sample sizes.

SES = supplemental educational services.

Table A.8. RD Impact Estimates on Selection into and Participation in SES

District	Impact Estimate on Selection	Impact Estimate on Participation
1	0.47*	NA[a]
2	0.23*	0.04
3	0.33*	0.22
4	0.24*	0.30*
5	0.89*	0.88*
6	0.88*	0.87*
Aggregated Across All Mini-Studies Using Variance-Minimizing Weight	0.55*	0.57*

Source: School district records.

Note: This evaluation consists of 42 RD mini-studies, each of which corresponds to a separate outcome/cutoff-value/grade combination. Aggregate impact estimates are calculated as a weighted average of the impact estimates for each RD mini-study, where the weight is equal to the inverse of the variance of that mini-study's impact estimate. For each RD mini-study, the impact estimate on selection and participation is calculated using a linear functional form within an optimal bandwidth selected using the Imbens-Kalyanaraman (2009) algorithm and standard errors are adjusted to account for clustering of observations within unique values of the assignment variable as in Lee and Card (2008). For each student, the selection indicator variable equals one if either (1) the district assigned him or her to a specific SES provider, or (2) he or she received any tutoring services (hours > 0). The participation indicator variable equals one if the student received any tutoring services (hours > 0).

[a]Data on hours of service received in District 1 pertain to a point in time during summer 2009, rather than to the hours of service received during the 2008–2009 school year, so this district was excluded from the aggregate RD impact estimate on participation.

*Significantly different from zero at the 0.05 level, two-tailed test.

RD = regression discontinuity; SES = supplemental education services.

NA = not available.

Table A.9. RD Impact Estimates on Selection into SES and Participation in SES, by Mini-Study

Mini-Study Description	Impact Estimate on Selection	Impact Estimate on Participation"
District 1 / grade 3	0.09*	NA[a]
District 1 / grade 4	0.95*	NA[a]
District 1 / grade 5	0.50	NA[a]
District 1 / grade 6	0.74*	NA[a]
District 1 / grade 7	0.68*	NA[a]
District 1 / grade 8	0.72*	NA[a]
District 2 / grade 3	-0.45	-0.66
District 2 / grade 4	0.35*	0.17
District 3 / grade 4	0.14	0.11
District 3 / grade 5	0.13	0.10
District 3 / grade 6	0.61	0.45
District 4 / grade 3	0.55*	0.57*
District 4 / grade 4	0.52*	0.51*
District 4 / grade 5	0.34*	0.33*
District 4 / grade 6	0.26*	0.32*
District 4 / grade 7	0.08	0.12
District 4 / grade 8	0.00	0.06
District 5 / grade 3	0.89*	0.88*
District 6 / grade 3	0.89*	0.89*
District 6 / grade 4	0.82*	0.82*
District 6 / grade 5	0.84*	0.85*

Source: School district records.

Note: This evaluation consists of 42 RD mini-studies, each of which corresponds to a separate outcome/cutoff-value/grade combination. Twenty-one of these correspond to a math test score outcome and 21 correspond to a reading test score outcome. For each of the 21 mini-studies, the impact estimate on selection and participation is calculated using a linear functional form within an optimal bandwidth selected using the Imbens-Kalyanaraman (2009) algorithm and standard errors are adjusted to account for clustering of observations within unique values of the assignment variable as in Lee and Card (2008). For each student, the selection indicator variable equals one if either (1) the district assigned him or her to a specific SES provider or (2) he or she received any tutoring services (hours > 0). The participation indicator variable equals one if the student received any tutoring services (hours > 0).

[a]Impact estimates on participation were not calculated for District 1 because the data on hours of service received come from a point in time during summer 2009 and do not represent the hours of service received during the 2008–2009 school year.

*Significantly different from zero at the 0.05 level, two-tailed test.

RD = regression discontinuity; SES = supplemental education services.

NA = not available.

Table A.10. Comparing District 2 SES Providers With Low And High Concentrations of ELL Students

Value	Percentage of Providers with Low Concentrations of ELL Students	Percentage of Providers with High Concentrations of ELL Students
Variable 1: Hours of Training in First Year		
1	3.8	0
2	11.5	0
3	3.8	0
4	11.5	25
5	11.5	0
6	7.7	0
7	3.8	0
8	23.1	0
10	0	50
12	3.8	0
18	0	25
20	11.5	0
40	3.8	0
44	3.8	0
p-value from test of difference across low and high providers:		0.04
Variable 2: How Closely Aligned Is SES Curriculum to State Standards?		
0 (Not at all aligned with state)	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.0
5	0.0	0.0
6	0.0	0.0
7	0.0	0.0
8	3.6	50.0
9	10.7	25.0
10 (Completely aligned with state)	85.7	25.0
p-value from test of difference across low and high providers:		0.01
Variable 3: Degree to Which Parent Email Access Was Communication Challenge		
1. Not a challenge	35.7	0.0
2. Minor challenge	14.3	0.0
3. Moderate challenge	0.0	50.0
4. Serious challenge	50.0	50.0
p-value from test of difference across low and high providers:		0.01

Source: School district records and provider survey.

ELL = English language learner; SES = supplemental education services.

Table A.11. Comparing District 4 SES Providers With Low and High Concentrations of ELL Students

Value	Percentage of Providers with Low Concentrations of ELL Students	Percentage of Providers with High Concentrations of ELL Students
Variable 1: Subjects Offered to SES Students in Intermediate Elementary Grades in District—Math		
0. No	4.5	36.4
2. Yes	95.5	63.6
p-value from test of difference across low and high providers:		0.03
Variable 2: Manipulatives Used in Math SES		
0. No	3.3	36.4
1. Yes	96.7	63.6
p-value from test of difference across low and high providers:		0.01
Variable 3: How Often Parents Received Information in Person		
1. At least once per week	3.3	15.4
2. A few time per month	30.0	23.1
3. At least once per month	33.3	7.7
4. A few times per year	23.3	7.7
5. Didn't use in-person comm.	10.0	46.2
p-value from test of difference across low and high providers:		0.02
Variable 4: How Often Progress Related to SES Was Addressed with Teachers		
1. At least once per week	10.0	0.0
2. A few time per month	30.0	0.0
3. At least once per month	50.0	100.0
4. A few times per year	10.0	0.0
5. Never	0.0	0.0
p-value from test of difference across low and high providers:		0.05

Source: School district records and provider survey.

ELL = English language learner; SES = supplemental education services.

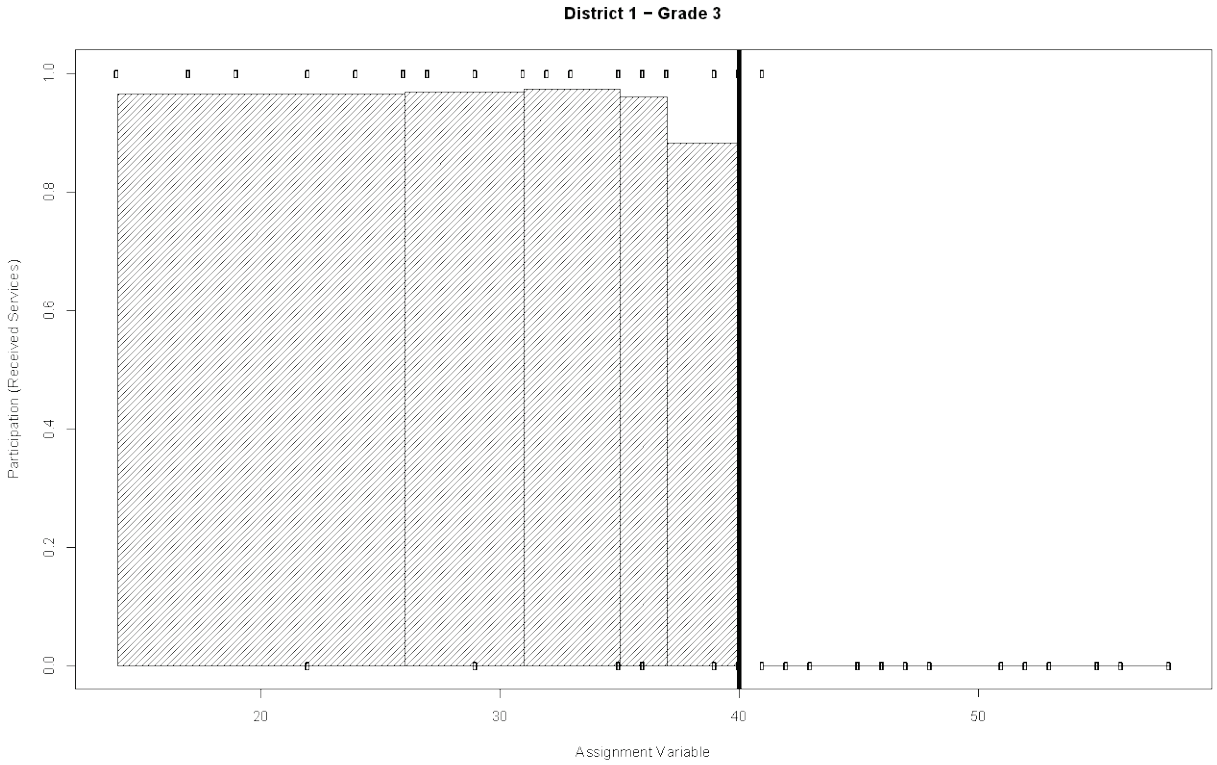
Table A.12. Comparing District 4 SES Providers With Low And High Concentrations of SPED Students

Value	Percentage of Providers with Low Concentrations of SPED Students	Percentage of Providers with High Concentrations of SPED Students
Variable 1: Percentage of Parents Who Approved Plan Without Contributing		
1. Most (more than 50%)	87.0	50.0
2. Some (25–50%)	8.7	7.1
3. Few (fewer than 25%)	0.0	21.4
4. None	4.3	21.4
p-value from test of difference across low and high providers:		0.02
Variable 2: Percentage of Parents Who Contributed to Content of Plan		
1. Most (more than 50%)	28.0	29.4
2. Some (25–50%)	4.0	41.2
3. Few (fewer than 25%)	64.0	17.6
4. None	4.0	11.8
p-value from test of difference across low and high providers:		0.00
Variable 3: How Often Did Organization Staff Communicate with Others in School About Student Progress		
1. At least once per week	4.3	20.0
2. A few time per month	26.1	33.3
3. At least once per month	43.5	20.0
4. A few times per year	21.7	0.0
5. SES staff did not communicate with other school staff	4.3	26.7
p-value from test of difference across low and high providers:		0.03
Variable 4: Number of Schools in Which SES Participants in District Were Enrolled		
1	8.7	6.2
2	13.0	0.0
3	0.0	6.2
4	0.0	6.2
5	17.4	0.0
7	8.7	6.2
8	8.7	0.0
9	0.0	6.2
10	4.3	25.0
11	0.0	6.2
12	4.3	6.2
15	0.0	12.5
18	4.3	0.0
30	4.3	0.0
38	4.3	0.0
40	8.7	0.0
47	0.0	6.2
62	4.3	0.0
72	8.7	0.0
90	0.0	6.2
167	0.0	6.2
p-value from test of difference across low and high providers:		0.02

Source: School district records and provider survey.

SES = supplemental education services; SPED = special education.

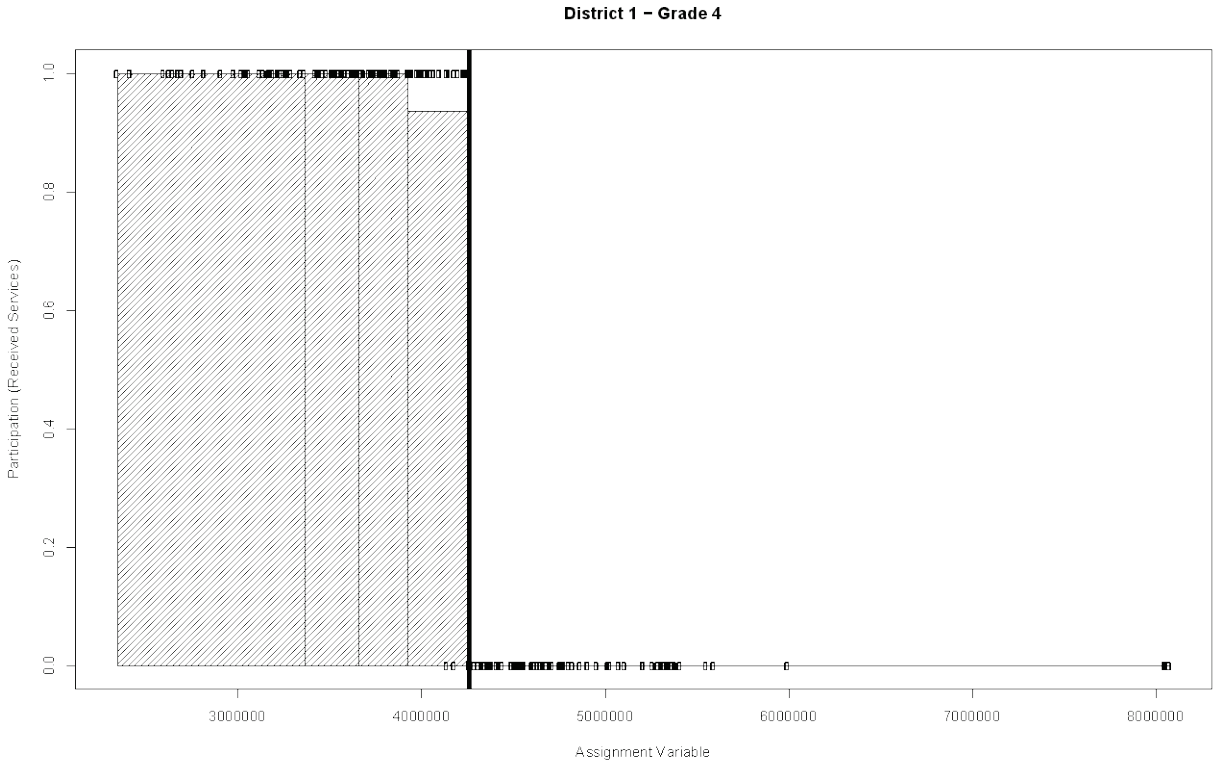
Figure A.1. Participation Indicator versus Assignment Variable for Mini-Study 1



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.3.

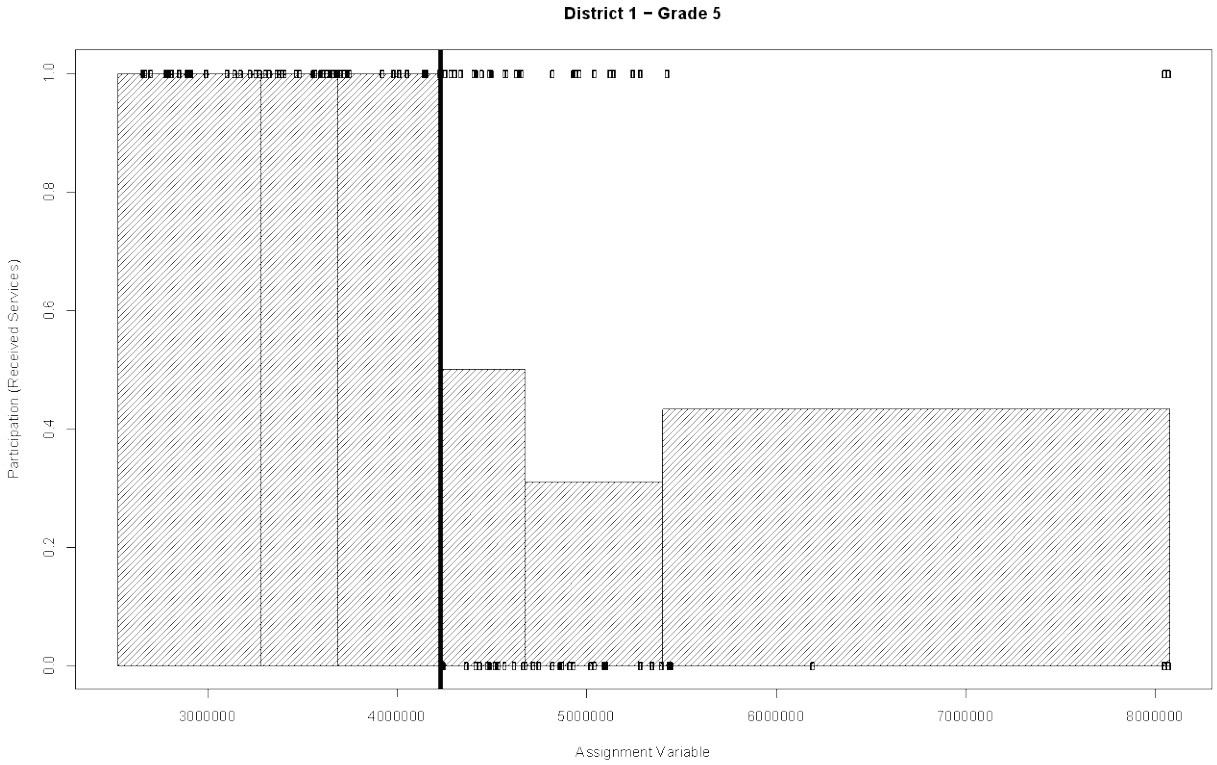
Figure A.2. Participation Indicator versus Assignment Variable for Mini-Study 2



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.3.

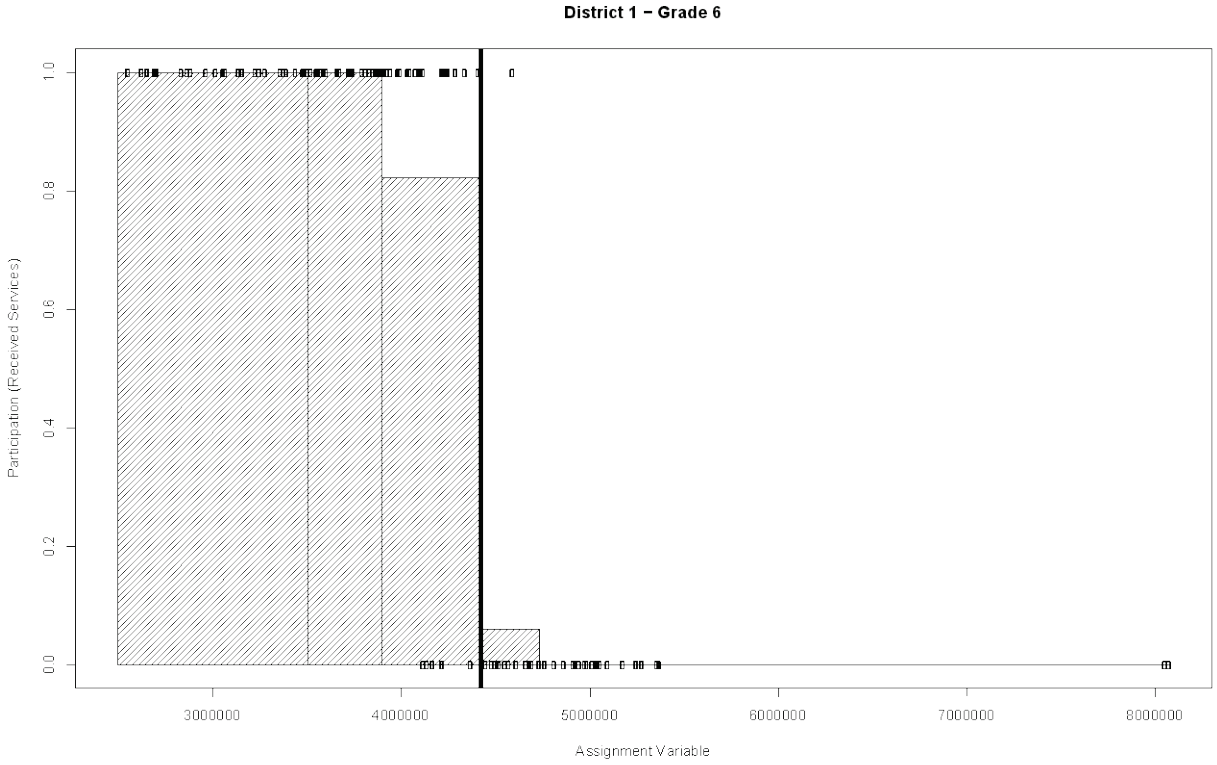
Figure A.3. Participation Indicator versus Assignment Variable for Mini-Study 3



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.3.

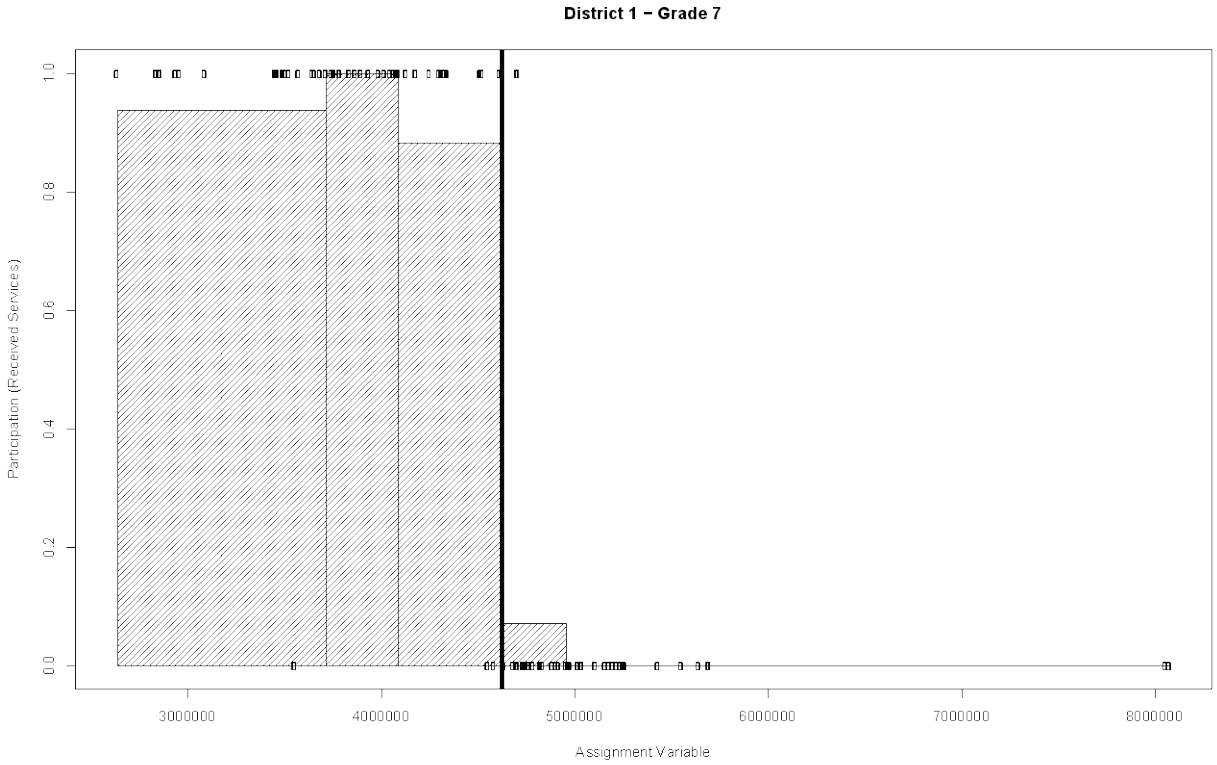
Figure A.4. Participation Indicator versus Assignment Variable for Mini-Study 4



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.3.

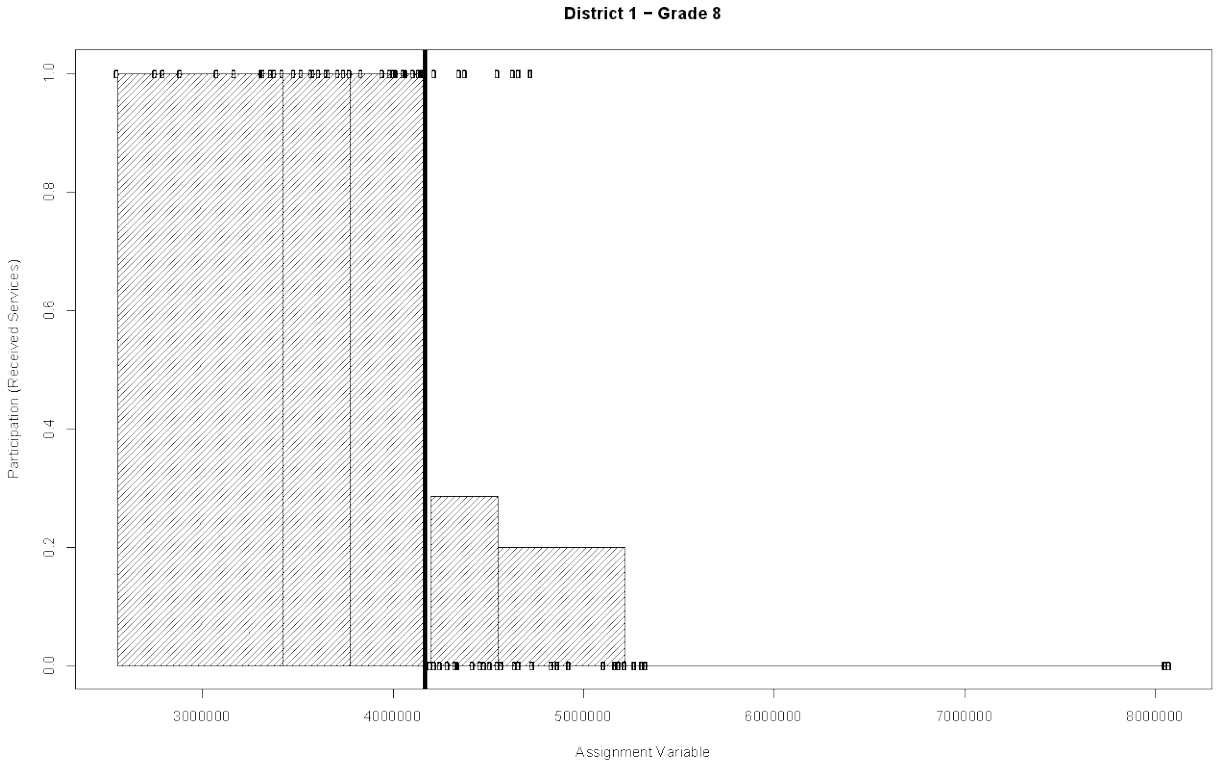
Figure A.5. Participation Indicator versus Assignment Variable for Mini-Study 5



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.3.

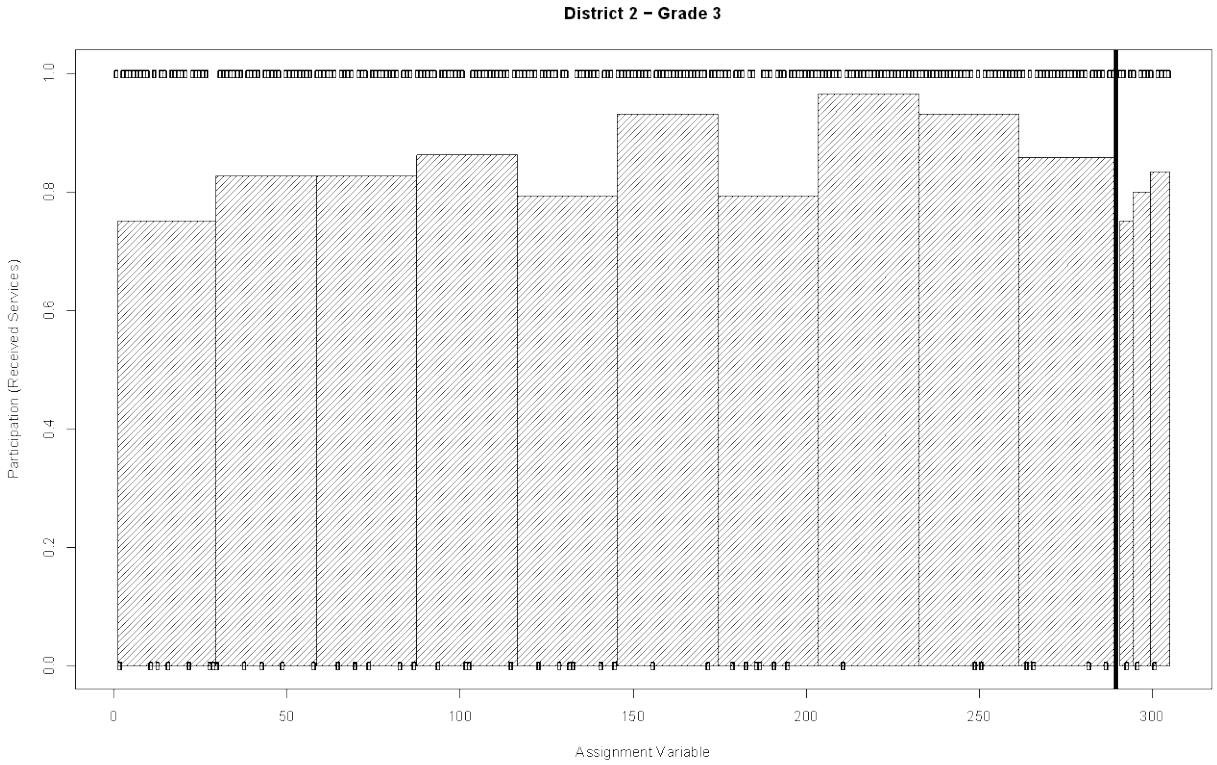
Figure A.6. Participation Indicator versus Assignment Variable for Mini-Study 6



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.3.

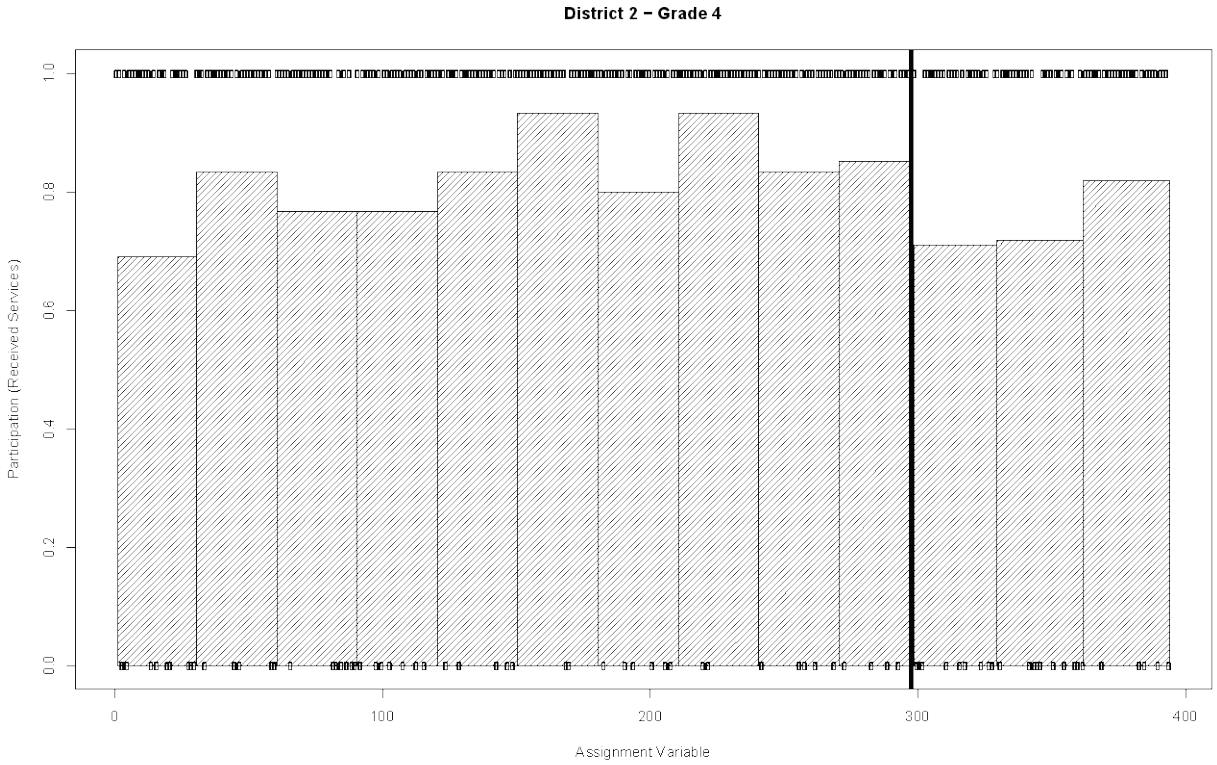
Figure A.7. Participation Indicator versus Assignment Variable for Mini-Study 7



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on pages A.3-A.4.

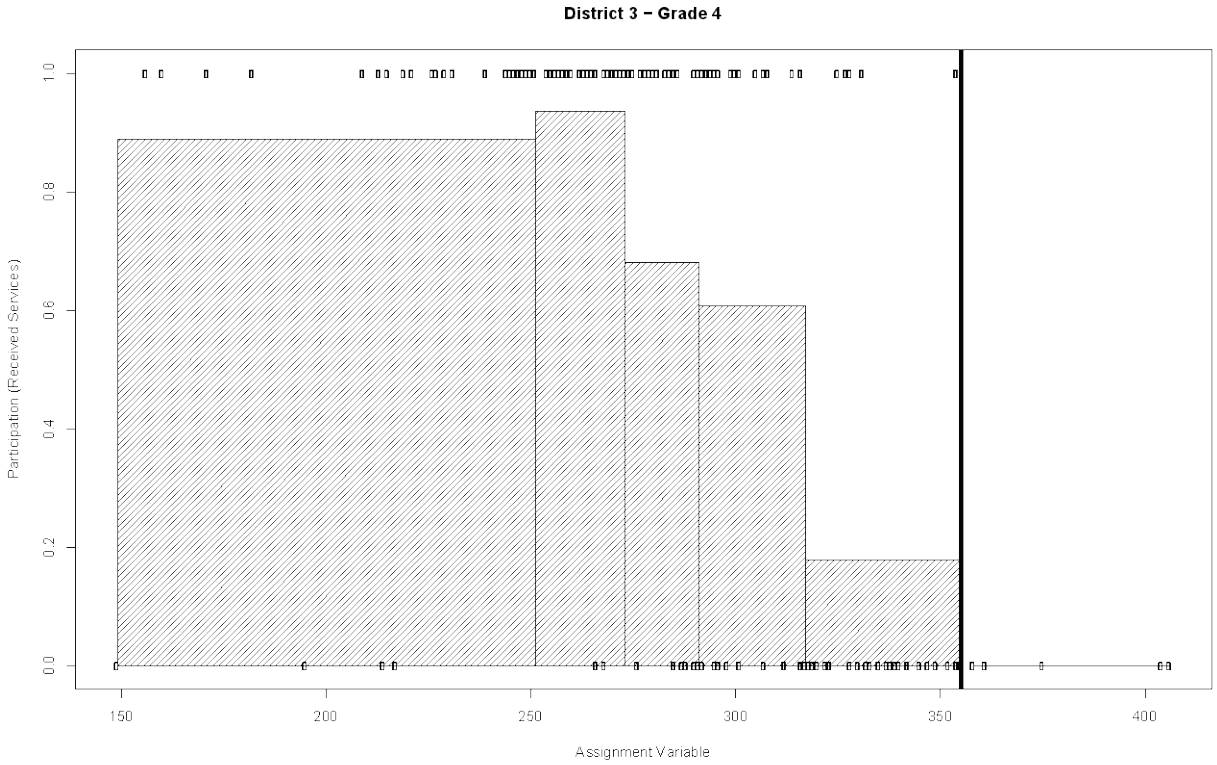
Figure A.8. Participation Indicator versus Assignment Variable for Mini-Study 8



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on pages A.3-A.4.

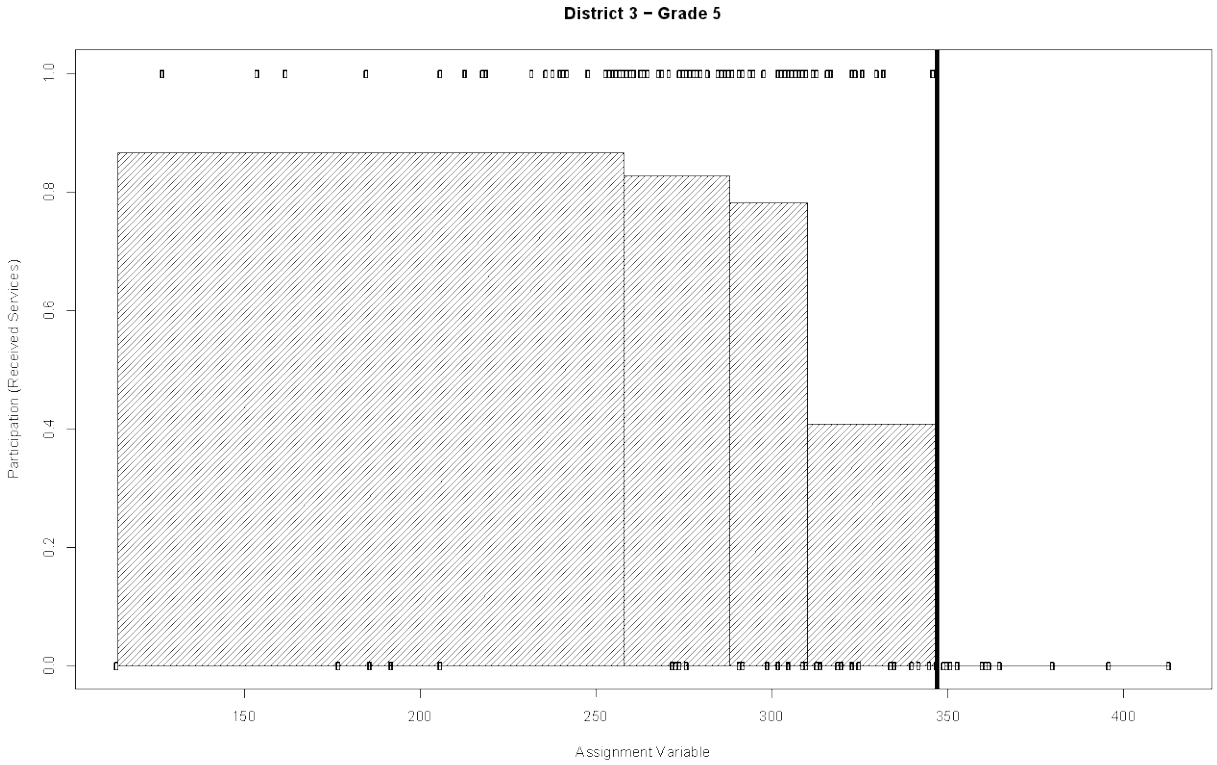
Figure A.9. Participation Indicator versus Assignment Variable for Mini-Study 9



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

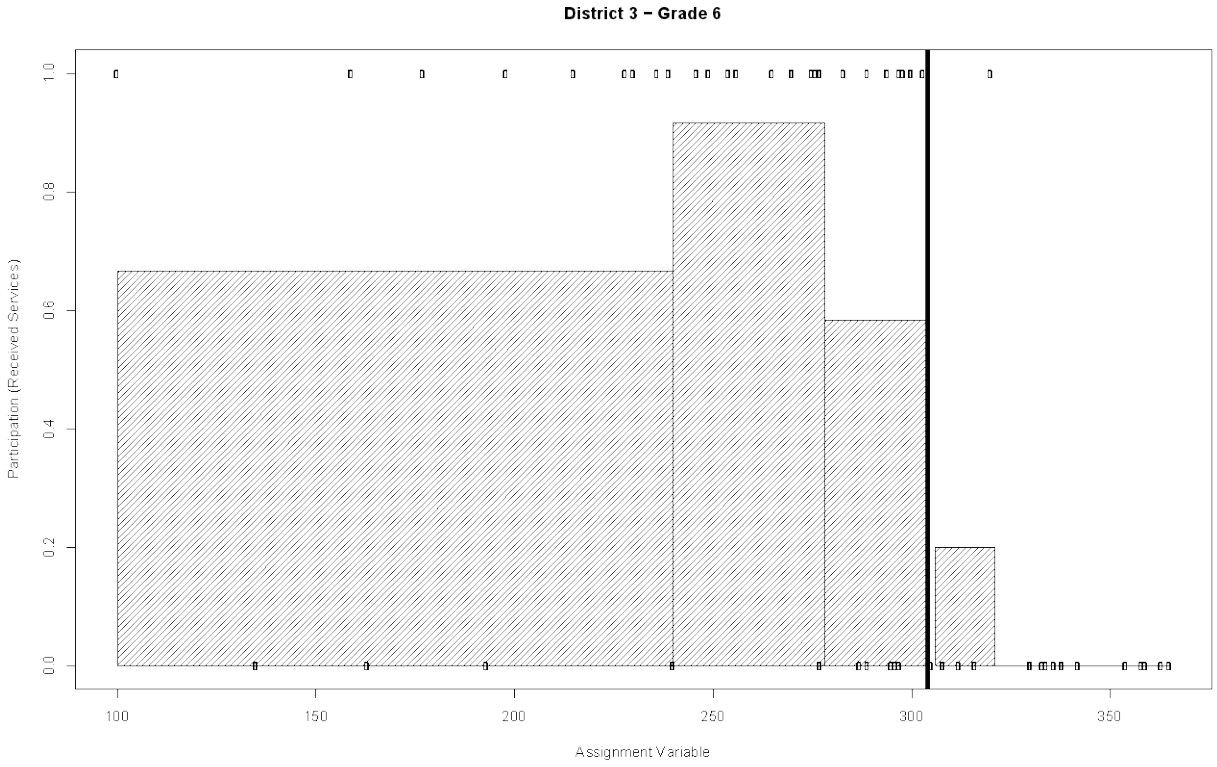
Figure A.10. Participation Indicator versus Assignment Variable for Mini-Study 10



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

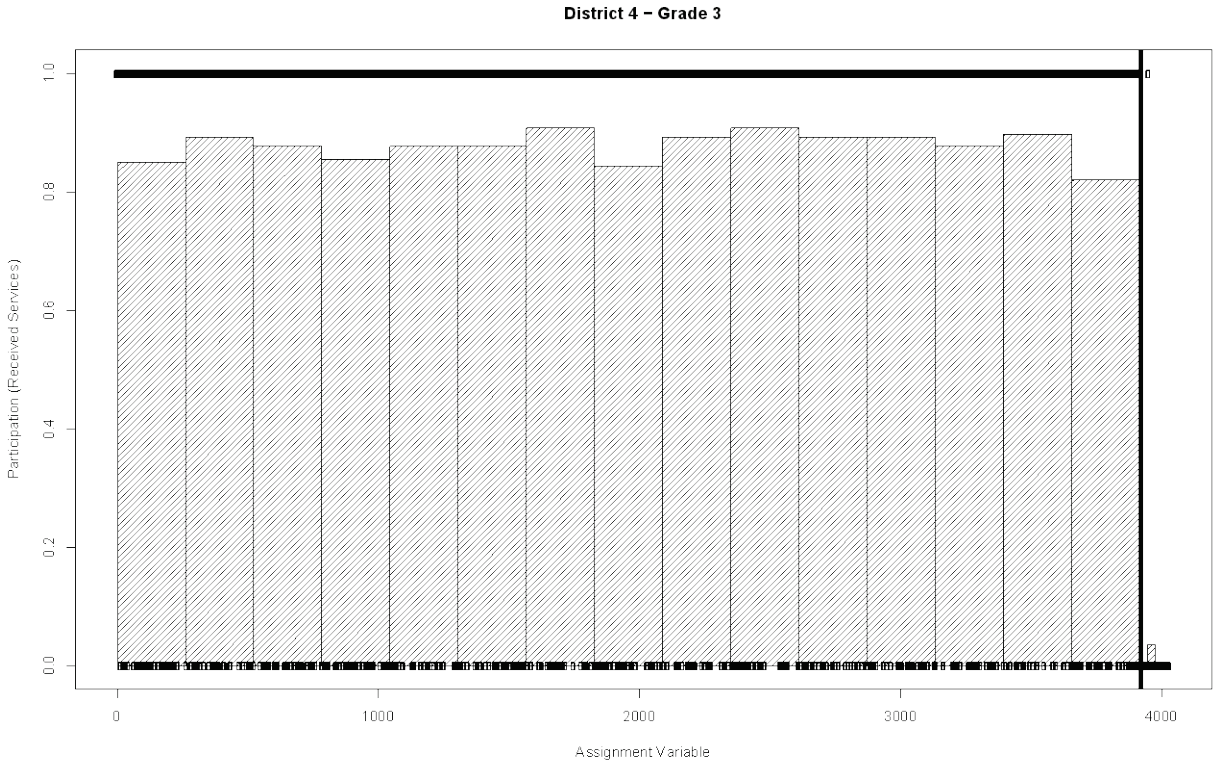
Figure A.11. Participation Indicator versus Assignment Variable for Mini-Study 11



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

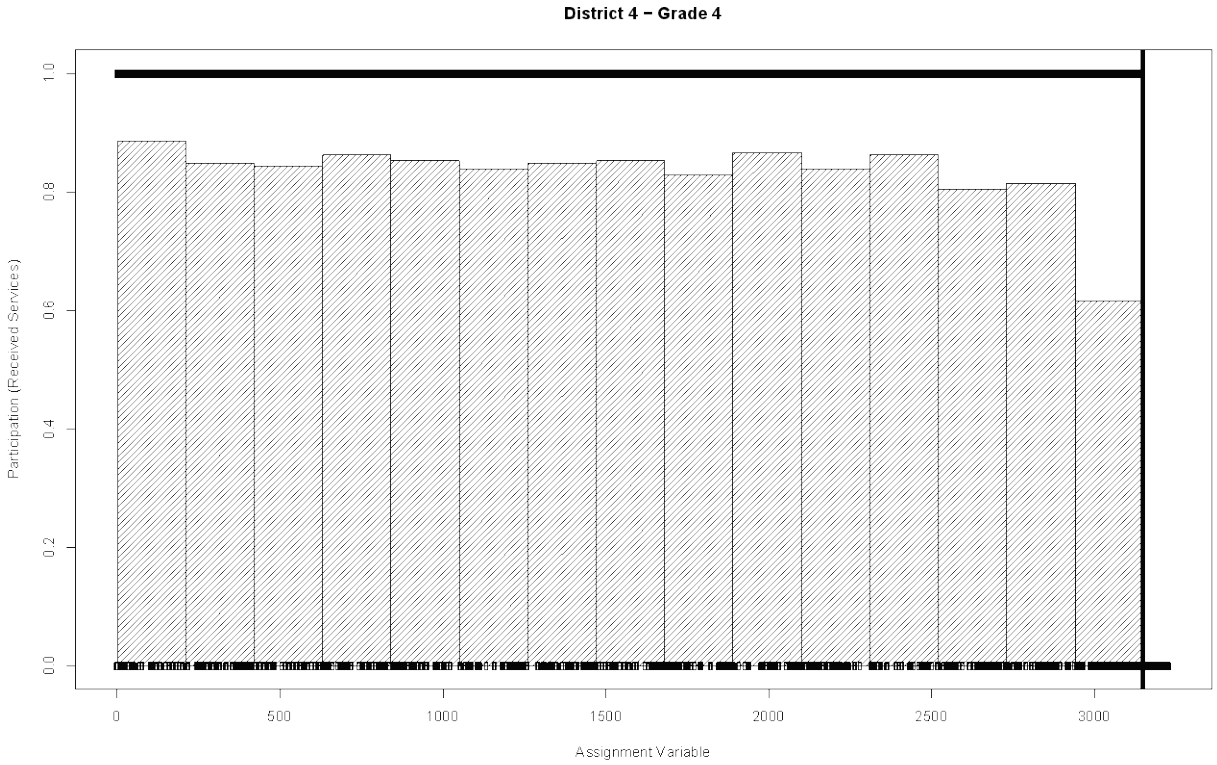
Figure A.12. Participation Indicator versus Assignment Variable for Mini-Study 12



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

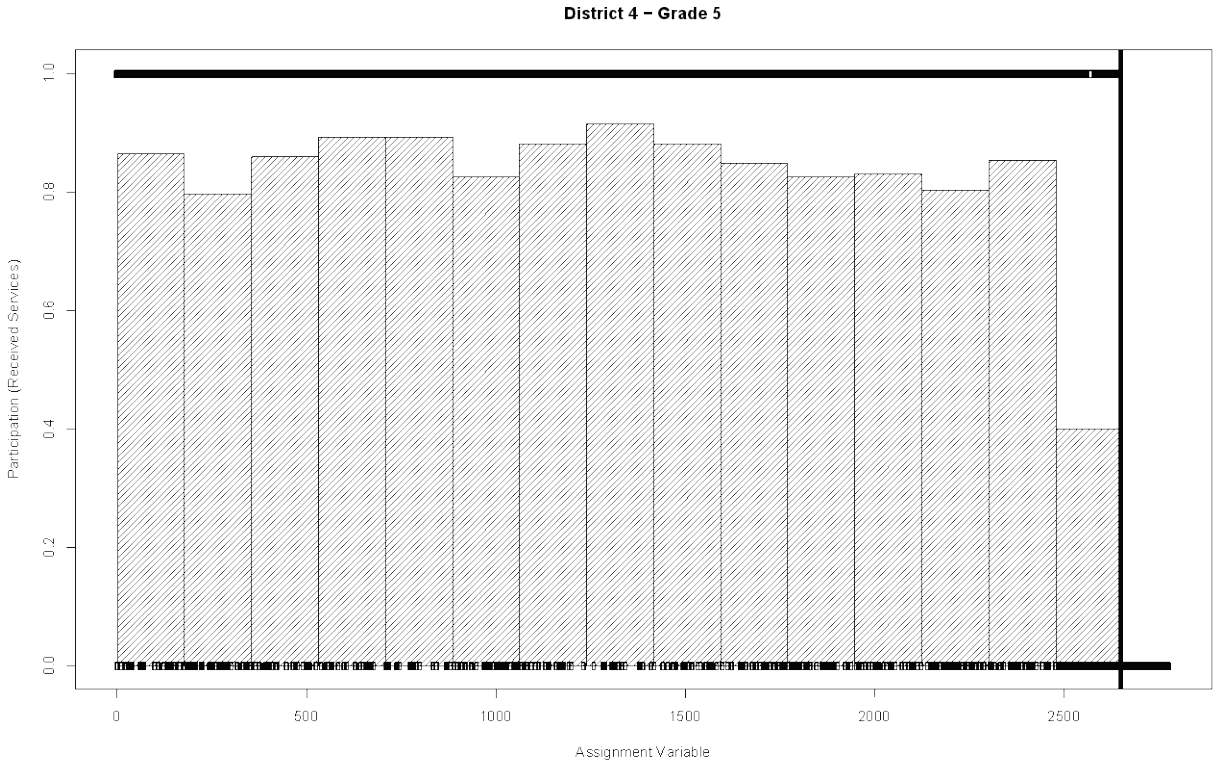
Figure A.13. Participation Indicator versus Assignment Variable for Mini-Study 13



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

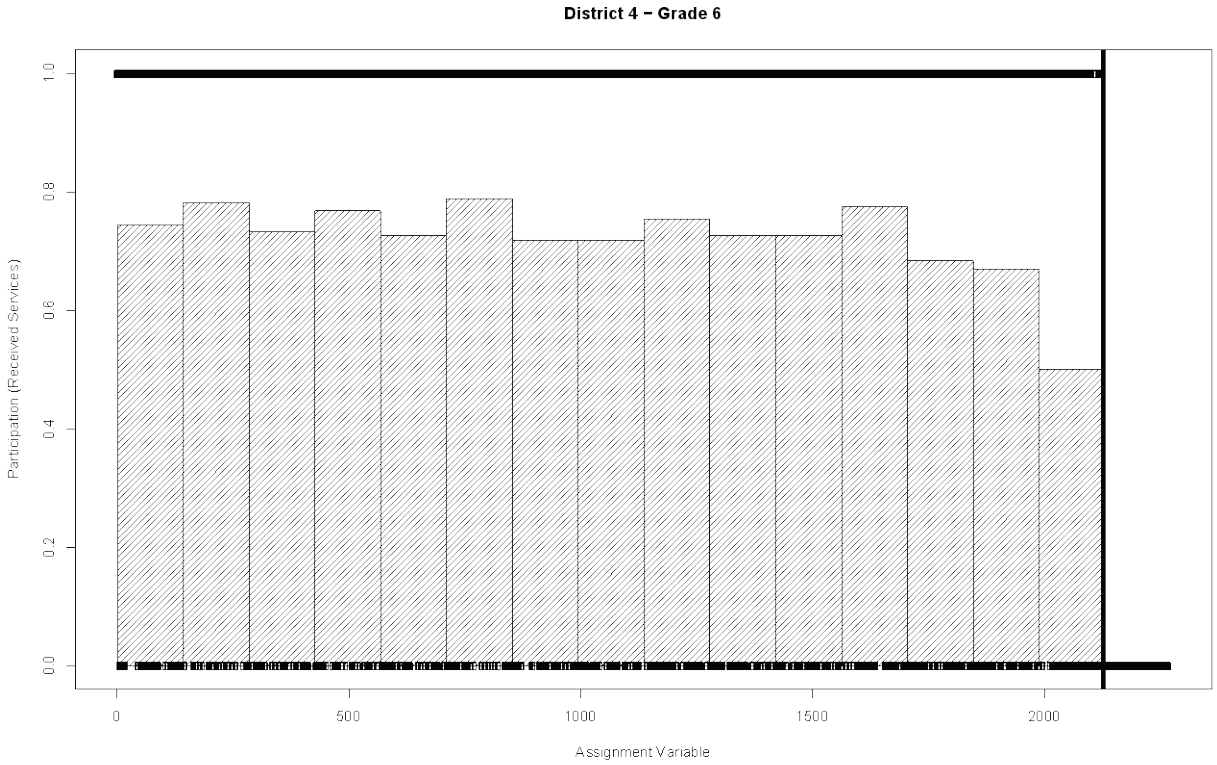
Figure A.14. Participation Indicator versus Assignment Variable for Mini-Study 14



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

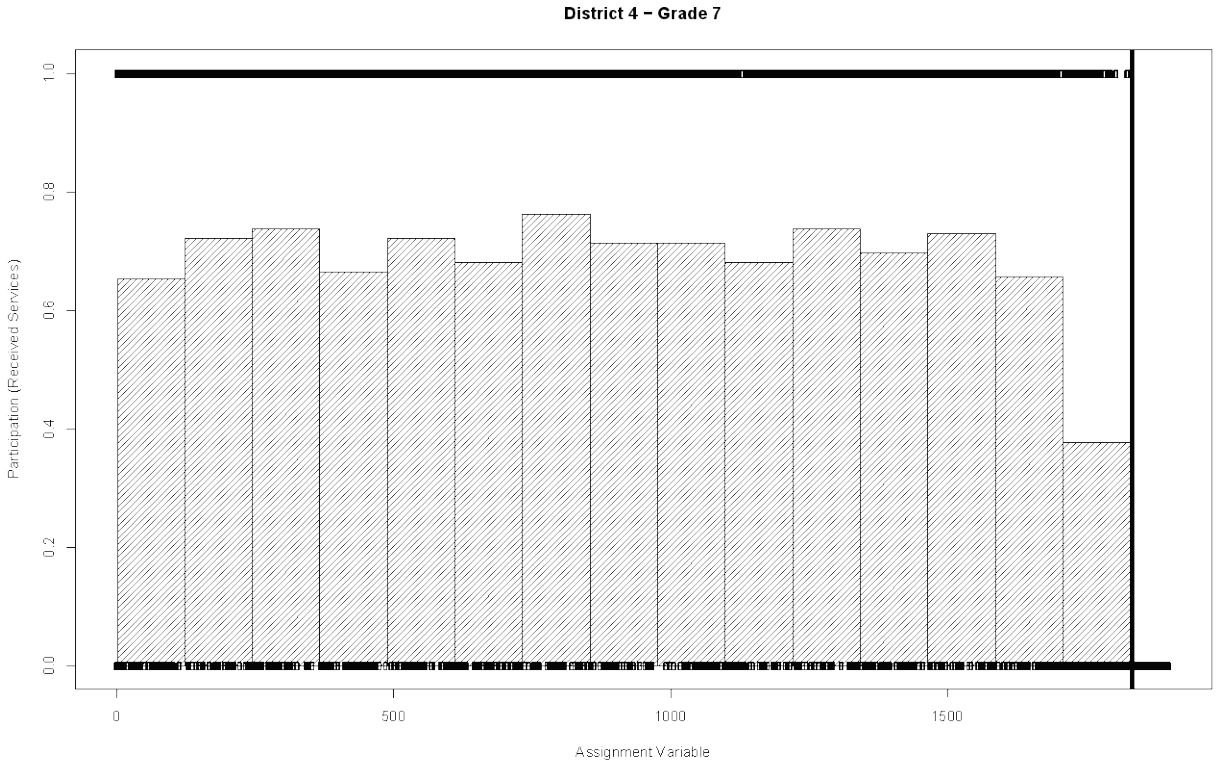
Figure A.15. Participation Indicator versus Assignment Variable for Mini-Study 15



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

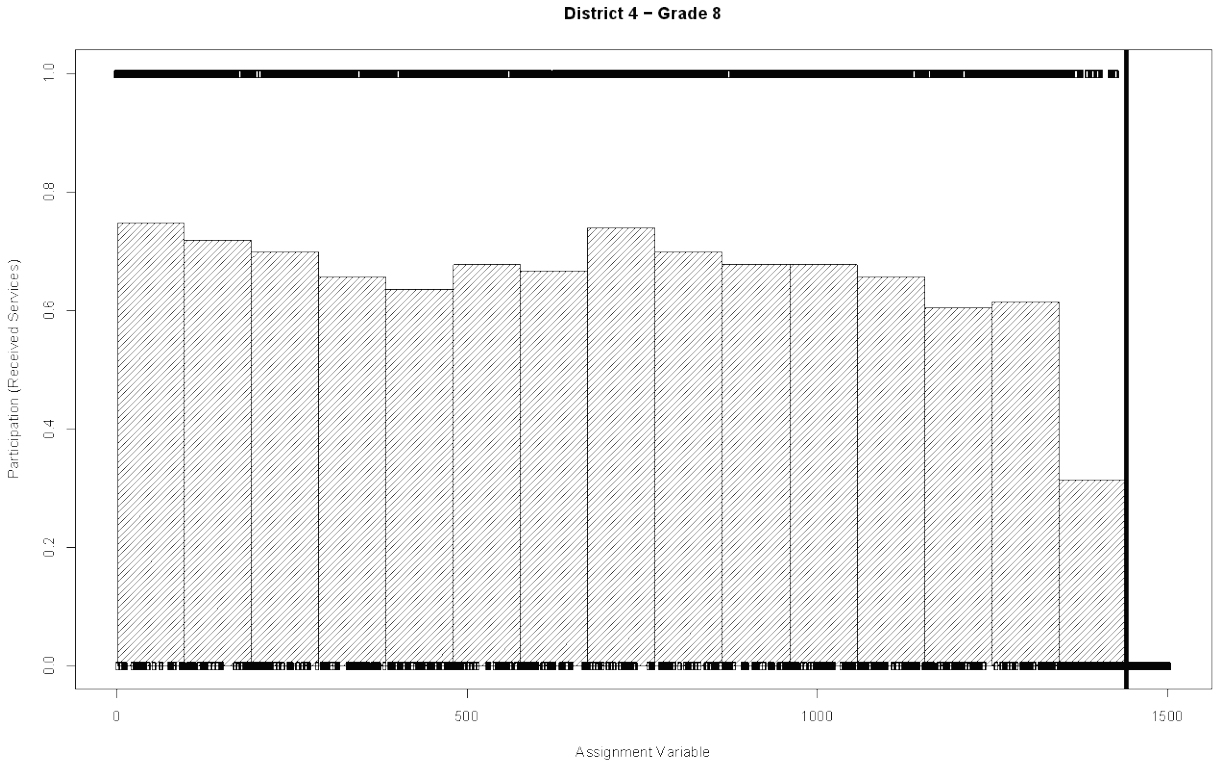
Figure A.16. Participation Indicator versus Assignment Variable for Mini-Study 16



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

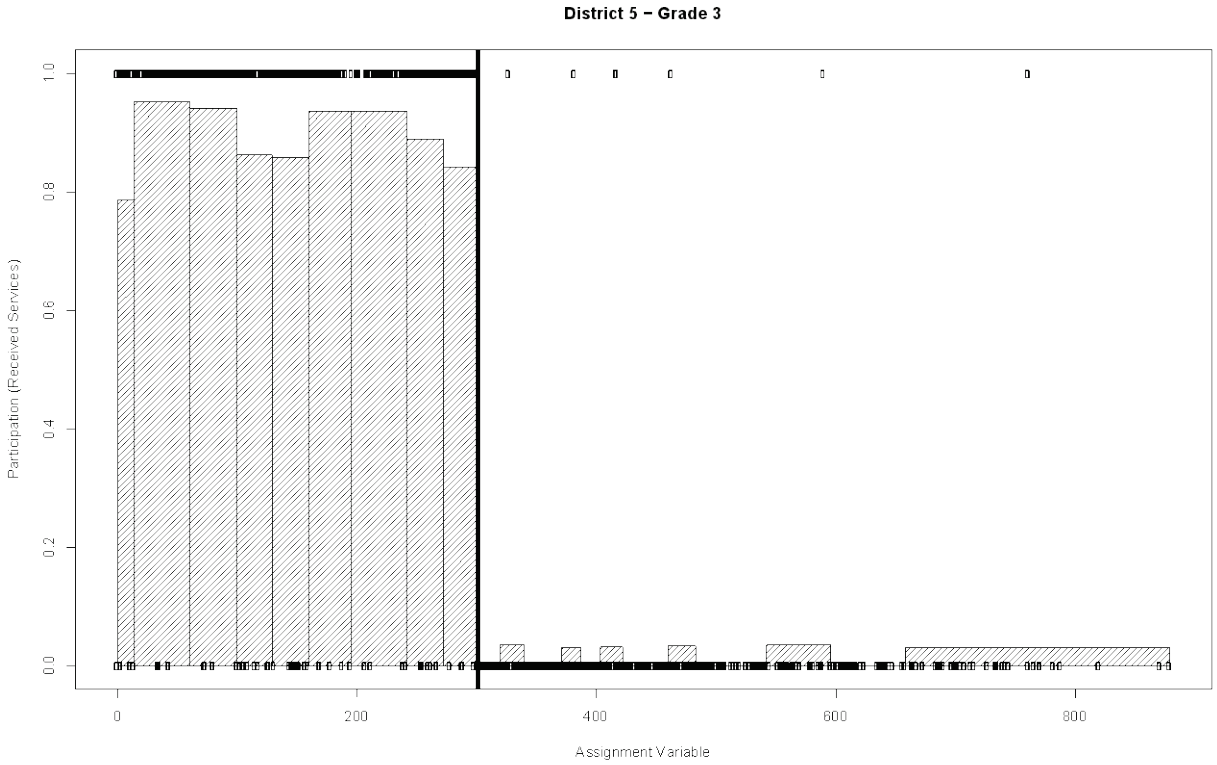
Figure A.17. Participation Indicator versus Assignment Variable for Mini-Study 17



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

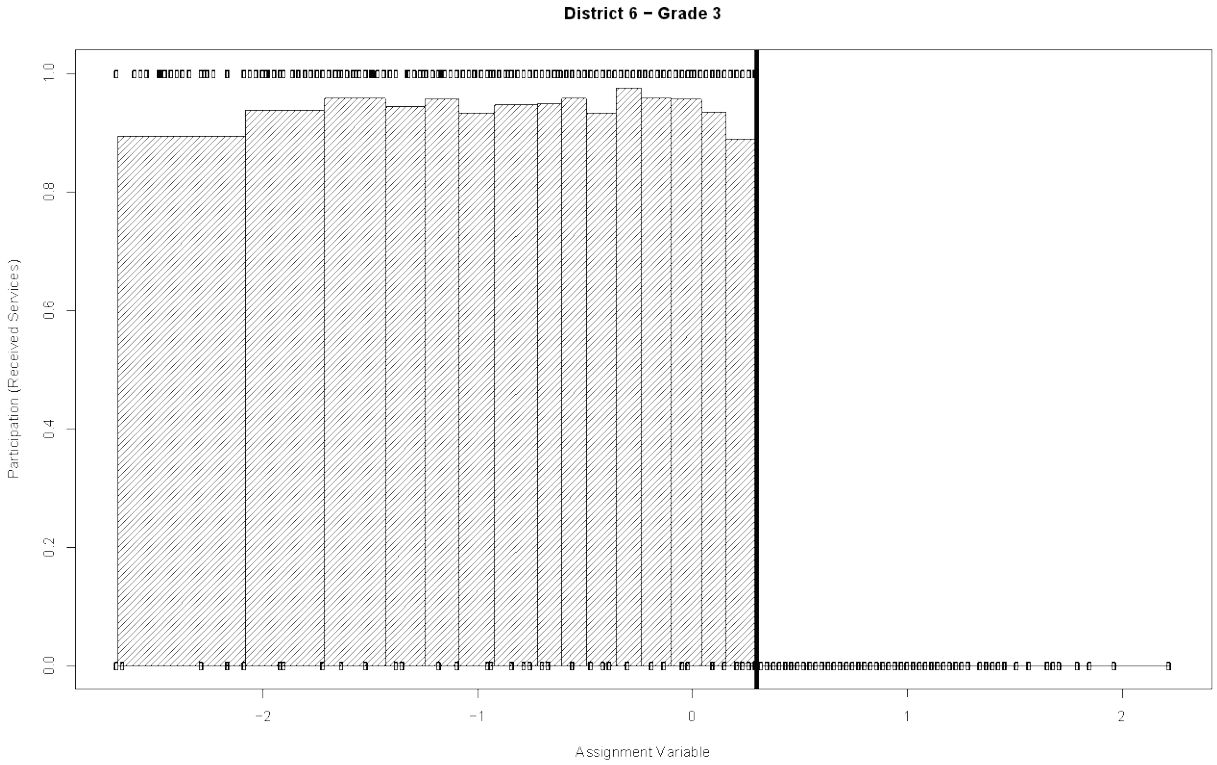
Figure A.18. Participation Indicator versus Assignment Variable for Mini-Study 18



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A5.

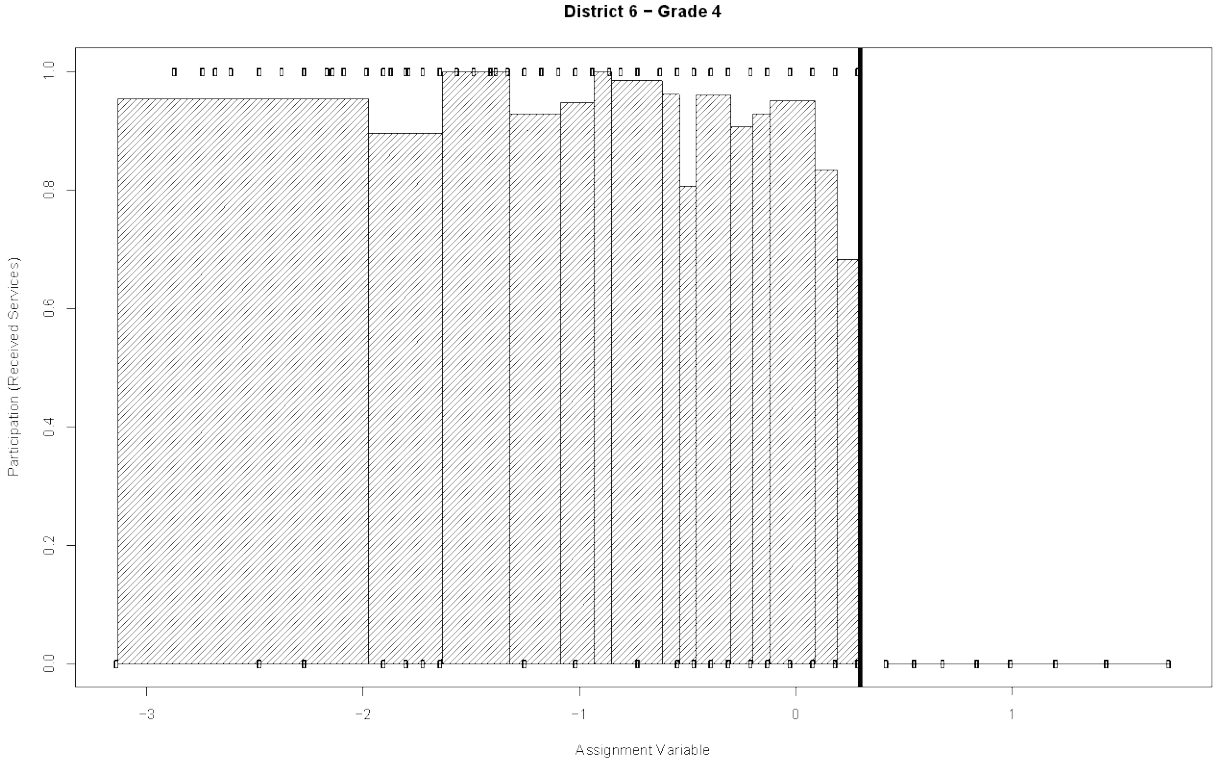
Figure A.19. Participation Indicator versus Assignment Variable for Mini-Study 19



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

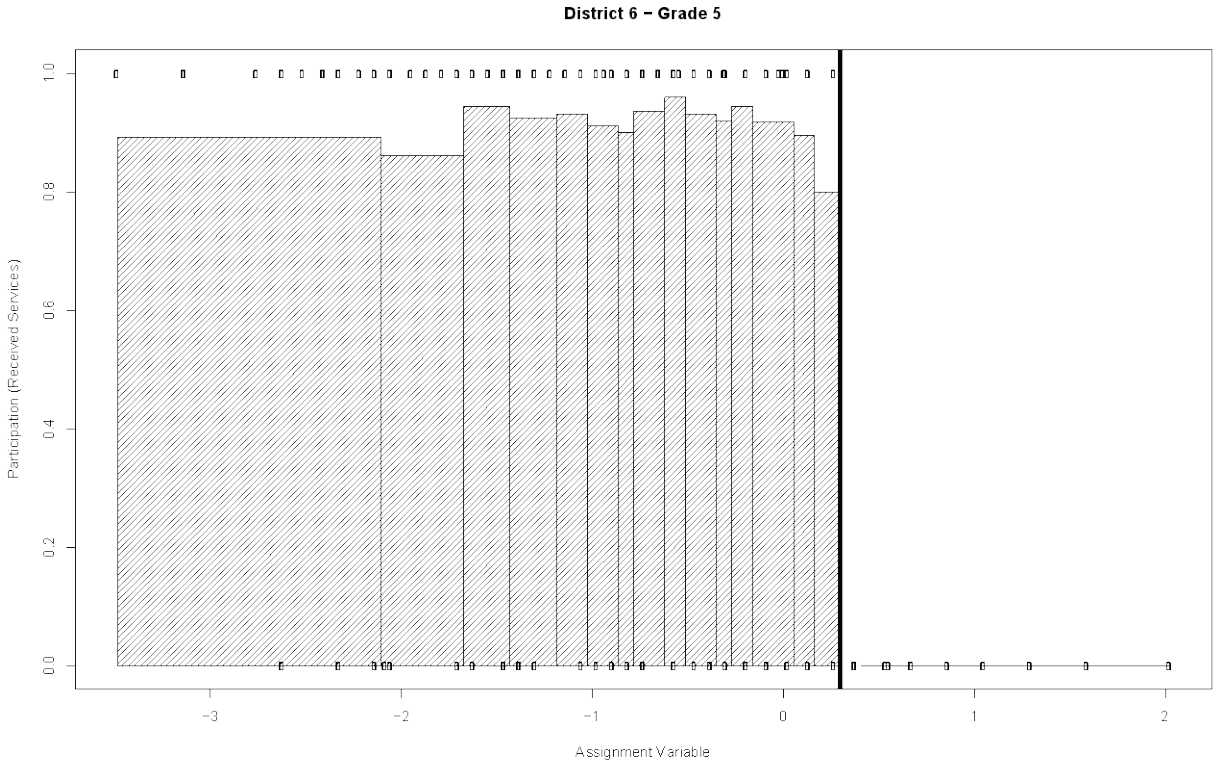
Figure A.20. Participation Indicator versus Assignment Variable for Mini-Study 20



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

Figure A.21. Participation Indicator versus Assignment Variable for Mini-Study 21



Source: School district records.

Notes: For each student, the participation indicator equals one if the student received any tutoring services (that is, hours of service were greater than 0). The shaded vertical bars represent the average participation rate for students in each range of the assignment variable defined by the width of each bar. The assignment variable is described on page A.5.

APPENDIX B

ESTIMATION METHODS

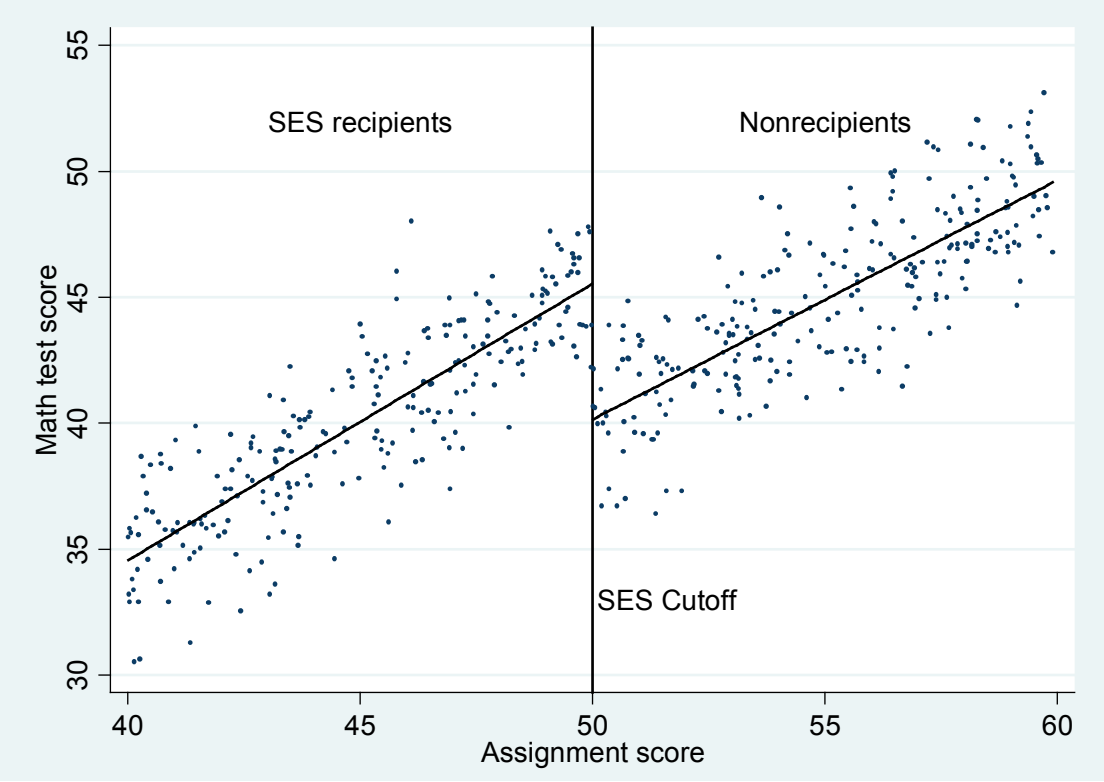
This appendix provides additional details about the estimation methods used in the study. The first section provides a basic description of the regression discontinuity (RD) design. The second section describes the method for choosing the bandwidth that was used when estimating RD impacts. The third section describes the method used to adjust for clustering. The fourth section describes the method for choosing the order of the optimal polynomial regression used in the sensitivity analyses. The fifth section describes the weights used to aggregate impact estimates across mini-studies. The final section describes the method used to test for homogeneity of impact estimates across providers.

A. Regression Discontinuity Design

Under the RD design, estimates of the impacts of SES on students at the cutoff can be obtained by comparing the outcomes (follow-up test scores) of students below and above the cutoff value, after adjusting for the score on the assignment variable. Unlike an RCT, in which the estimated average treatment effect applies to all students in the study, the RD impact applies only to students near the cutoff value of the assignment variable. This estimate does not necessarily represent the impact of SES on students far away from the cutoff value of the assignment variable. In other words, the RD analysis examines whether there is a discontinuity in the relationship between the assignment variable (prior achievement) and the outcome (subsequent achievement) at the prior achievement level that is the cutoff for assignment to services.

Figure B.1 illustrates the RD design graphically using a hypothetical example. In this example, students with an assignment score of 50 or below receive SES, and those with a score

Figure B.1. Hypothetical Example of the Regression Discontinuity Method



Source: Simulated data.

above 50 do not. This figure plots student math test scores against assignment scores. It also displays the fitted regression line for the treatment and comparison groups. The estimated impact on math test scores is the vertical distance between the two regression lines at the cutoff value of 50. An important consideration in calculating impacts using an RD design is the approach used to model the relationship between the posttest and the variable used to differentiate the two groups, in this case the cutoff score used to assign students to SES. In the example shown in Figure B.1, this relationship is clearly linear, but this often is not the case in practice.

With an RD design, the relationship between the outcome and the assignment variable must be modeled appropriately in order to obtain rigorous impact estimates. For example, if the true relationship between the outcome and the assignment variable is quadratic, then fitting a linear regression line to the data on either side of the cutoff might result in a statistically significant impact estimate even when the intervention has no effect (see Figure B.2). The methodological literature on RD designs includes two approaches to improving the accuracy of the model. The first approach fits a higher-order polynomial to the data (for example, Trochim, 1984), such as a model in which the outcome is regressed on the baseline test score and its square. The second approach selects a bandwidth that restricts the analysis to the range of data such that the relationship between the outcome and the assignment variable is approximately linear on both sides of the cutoff (for example, Imbens & Kalyanaraman, 2009).

Figure B.2. Example of Misspecification Bias

Source: Simulated data.

Note: The graph on the left shows a correctly modeled relationship between the outcome and the assignment variable. The graph on the right shows an incorrectly modeled relationship between the outcome and the assignment variable, leading to a biased impact estimate.

B. Choosing the IK Bandwidth

As discussed in Chapter II, the benchmark approach used to estimate intent-to-treat (ITT) impacts was a linear regression (estimated separately on either side of the cutoff) within a bandwidth selected by the optimal bandwidth selection algorithm described in Imbens and Kalyanaraman (2009). This section describes the method for choosing the Imbens and Kalyanaraman (IK) bandwidth.

The IK bandwidth is a data-driven estimate of the "optimal" bandwidth that minimizes the mean squared error (MSE) of the impact estimate.[1] This optimal bandwidth balances a tradeoff between the impact estimate having greater bias versus greater variance. To be specific, let τ_{RD} denote the average effect of the treatment for students with assignment variable values equal to the cutoff value, and let $\hat{\tau}_{RD}(h)$ denote the estimate of τ_{RD} using a bandwidth of h. From Imbens and Kalyanaraman (2009), the MSE of the impact estimate is:

$$(1) \quad MSE(h) = E[(\hat{\tau}_{RD}(h) - \tau_{RD})^2].$$

Let h^* be the optimal bandwidth that minimizes this criterion:

$$(2) \quad h^* = \arg\min MSE(h).$$

Now define the asymptotic mean squared error (AMSE) as a function of the bandwidth:

$$(3) \quad AMSE(h) = C_1 * h^4 * (m_+^{(2)}(c) - m_-^{(2)}(c))^2 + \frac{C_2}{N * h} * \left(\frac{\sigma_+^2(c)}{f_+(c)} + \frac{\sigma_-^2(c)}{f_-(c)} \right),$$

where C_1 and C_2 are constants, c is the cutoff value, N is the sample size of the original data set, $m_+^{(2)}(c)$ and $m_-^{(2)}(c)$ are the right and left limits of the second derivative of the relationship between the outcome and the assignment variable at the cutoff, $\sigma_+^2(c)$ and $\sigma_-^2(c)$ are the right and left limits of the conditional variance of the outcome variable given the assignment variable at the cutoff, and $f_+(c)$ and $f_-(c)$ are the right and left limits of the density of the assignment variable at the cutoff. The first term in Equation (3) corresponds to the bias of the $\hat{\tau}_{RD}(h)$ estimator, and the second term corresponds to the variance. The bias of the $\hat{\tau}_{RD}(h)$ estimator is affected by the curvature (that is, the second derivative) of the relationship between the outcome and the assignment variable at the cutoff because the more curvature that exists in the data, the more bias will exist in a linear regression estimate at the cutoff (see the example in Figure B.2).

[1] Ludwig and Miller (2005) also suggest a bandwidth selection procedure, but it is not focused on minimizing the MSE of the impact estimate.

The IK bandwidth, $\hat{h}_{opt}$, is an estimate of h^*, and equals:

$$(4)\quad h_{opt} = C_K * \left(\frac{2 * \hat{\sigma}^2(c) / \hat{f}(c)}{\left(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c)\right)^2 + (\hat{r}_+ + \hat{r}_-)} \right)^{1/5} * N^{-1/5},$$

where C_K is a constant that depends on the kernel used (we follow IK (2009) and use C_K = 3.4375, which corresponds to the edge kernel),[2] $\hat{\sigma}^2(c)$ is an estimate at the cutoff value of the conditional variance of the outcome variable given the assignment variable, $\hat{f}(c)$ is an estimate of the density of the assignment variable at the cutoff, $\hat{m}_+^{(2)}(c)$ and $\hat{m}_-^{(2)}(c)$ are estimates of the limits of the second derivatives at the cutoff value from the right and left, respectively (that is, $\hat{m}_+^{(2)}(c)$ and $\hat{m}_-^{(2)}(c)$ estimate the curvature of the data at the cutoff value), and $(\hat{r}_+ + \hat{r}_-)$ is a "regularization" term that is a function of the previous four components (all of these estimates are calculated as in IK [2009]). The regularization term addresses the problem that the curvature of the data could be spuriously underestimated when the sample size is low. It does this by imposing the conservative assumption that at least some curvature exists in the data. The size of the regularization term decreases with sample size, and nearly vanishes when the sample size is large.

In practice, we modify the above analysis by using the conditional variance of the outcome variable given all available covariates. Specifically, before choosing the bandwidth, we regress the outcome variable on all available covariates (excluding the assignment variable). The residuals from this regression were then used as the outcome variable when determining the bandwidth.

After selecting the optimal bandwidth around the cutoff, we estimate two separate regression equations using data within the bandwidth, one for observations to the right of the RD cutoff value and one for observations to the left of the cutoff value. The right and left regression equations are:

$$(5)\quad Y_i^R = \beta_0^R + \beta_1^R X_i^R + \beta_2^R Z_i^R + u^R + \varepsilon_i^R,$$

and

$$(6)\quad Y_i^L = \beta_0^L + \beta_1^L X_i^L + \beta_2^L Z_i^L + u^L + \varepsilon_i^L,$$

where Y_i is the outcome for student *i*, X_i is the assignment variable centered at the cutoff value, Z_i is a set of mean-centered baseline covariates, u is a random effect to adjust for clustering in the assignment variable as described in Lee and Card (2008) (see Appendix B for details), ε_i is a student-level error term, and the superscripts *R* and *L* denote right and left of the RD cutoff

[2] As in IK, we also calculate impacts using weights from the edge kernel.

value. The interpretation of the constant term in a regression is "the expected mean outcome when all covariates equal zero." Thus, the assignment variable is centered at the RD cutoff value so that the intercept terms in equations (5) and (6) represent the predicted value of the outcome variable at the cutoff value. Similarly, the covariates (Z_i) are mean-centered at the mean calculated using only observations within the bandwidth. The RD impact of SES on the outcome is estimated by the difference in intercept terms: $\delta^{RD} = \beta_0^L - \beta_0^R$ (recall that students to the left of the RD cutoff value are offered SES). The baseline covariates (Z_i) are included in this model to increase precision and vary by district and grade level depending on data availability.

C. Adjustment for Clustering

Students were assigned to SES based on the value of the assignment variable. In cases in which there are fewer unique values of the assignment variable than there are students, the unit of assignment becomes unique values of the assignment variable, not individual students (that is, students are clustered within unique values of the assignment variable, and it is those clusters that are assigned to treatment or comparison groups). Thus, the benchmark approach adjusts standard errors to account for clustering of students within unique values of the assignment variable as suggested by Lee and Card (2008). This section describes the adjustment for clustering in more detail.

The RD design is predicated on assignment to treatment and comparison status using a *continuous* variable, not a binary or categorical variable. For example, assigning students to treatment and comparison groups using letter grades (A–F) would not be viewed as an RD design because the assignment variable is not continuous. In practice many variables are not truly continuous. For example, there are a finite number of unique values of a test score because there are a finite number of questions on a test.

Lee and Card (2008) observe that a lack of continuity in an assignment variable can lead to random misspecification error and they suggest using clustered standard errors to protect against false inferences that might arise from this error. In this case, students are "clustered" within unique values of the assignment variable. Intuitively this is analogous to a clustered lottery in which schools rather than students are assigned to treatment and comparison groups. Here, an analogous clustering is taking place—groups of students are being assigned to treatment and comparison status, but instead of being randomly assigned they are being purposefully assigned, and instead of the group being defined by a school it is defined by everyone with the same score on a pretest.

We addressed this issue in our analysis in two ways: (1) by reporting the number of unique values of the assignment variable on either side of the cutoff in order to provide a descriptive view of how continuous or discrete the variable is; and (2) by adjusting our standard errors for the clustering of observations within unique values of the assignment variable, using Equation (11) from Lee and Card (2008). This equation starts with the classic "sandwich" estimator of the impact variance, in which the clusters are defined by the unique values of the assignment variable. It then adds another term—Equation (12) in Lee and Card—to account for the possibility that the random specification error above the cutoff is independent of the random specification error below the cutoff. Our analysis of data from past studies revealed that this expanded formula performed better than the classic sandwich estimator. We also found that this formula performs better than bootstrapping, both in simulations and using data from past studies.

D. Choosing the Order of the Optimal Polynomial

To test the sensitivity of the impact findings to our choice of benchmark analytic methods, we estimated impacts with an optimal polynomial regression using all of the data. This section describes the method for choosing the order of the optimal polynomial.

Our optimal polynomial approach is loosely based on the guidance from Trochim (1984), who suggests modeling the relationship between the outcome and the assignment variable using higher-order polynomials of the assignment variable (for example, quadratic and cubic terms). We selected the optimal polynomial degree separately on either side of the cutoff using a stepwise algorithm. The stepwise algorithm adds or drops higher-order polynomials sequentially to the model and computes a "goodness-of-fit" measure for each specification considered. Specifically, we used the Akaike Information Criterion (AIC) as our goodness-of-fit measure (Akaike, 1974). The stepwise algorithm then chooses the regression specification with the smallest AIC value.

Our experience in selecting an optimal polynomial (both in using real data and in simulations) has been that higher-order terms are sometimes included spuriously, resulting in extreme impact estimates that lack face validity. In order to protect against this, we included an additional stopping rule in our polynomial selection algorithm. If selecting a higher-order polynomial resulted in an impact estimate of two standard deviations or higher, we did not choose the higher order polynomial. For example, if including a linear and quadratic term yielded an impact estimate of fewer than two standard deviations but including a cubic term resulted in an impact estimate greater than two standard deviations, we did not include the cubic term, even if doing so appeared better in terms of the AIC. This does not prevent us from detecting large impacts, it merely stops us from detecting large impacts that arise only from the inclusion of a higher-order polynomial term.

E. Weights Used To Aggregate Impact Estimates

The overall impact estimate of SES is calculated as a weighted average of mini-study impact estimates. This approach to calculating impact estimates is essentially a meta-analysis. Our benchmark approach to aggregating impact estimates follows the guidance in Cooper, Hedges, and Valentine (2009) to weight each impact estimate by the inverse of its variance. In our context, this approach to weighting can be noticeably different from a sample size weight, because the variance of each RD impact estimate is a function of more than just the sample size (for example, clustering, degree of correlation between treatment status and the assignment variable, and degree of correlation between the assignment variable and the outcome can all affect the variance of an RD impact estimate).

As a sensitivity analysis we also calculated impact estimates using sample size weights and giving equal weight to each mini-study. If impact estimates vary systematically with respect to impact variance or sample size, then these alternative approaches could yield a substantively different impact estimate. However, whether these approaches yield a substantively different impact estimate would be difficult to discern in many cases due to the much larger standard errors associated with these alternative weighting approaches. For example, the standard error of the overall impact estimate on reading test scores using the inverse variance weights is 0.05, the standard error using the sample size weights is 0.11, and the standard error using equal weighting

of mini-studies is 0.41. As shown in Appendix E, the differences in overall impact estimates using these different weighting approaches are small relative to those large standard errors.

F. Exploring Variation in Provider-Specific Impact Estimates

On the SES application form, parents identified their first-choice SES provider. Among the students who were offered SES in these six districts, 98 percent were offered the opportunity to attend their first-choice provider. This suggests that it is highly likely that students in the comparison group would have been offered their first-choice provider had they been assigned to the treatment group.

We take advantage of this information to calculate provider-specific impact estimates. Specifically, for each mini-study, we identify the subgroup of students (in both the treatment and comparison groups) who named a specific first-choice provider (for example, Provider A). We then calculate the estimated impact of SES on this subgroup of students using our benchmark approach to calculating impacts (see Chapter II). If Provider A served students in multiple mini-studies, we calculate the overall estimated impact for Provider A by aggregating provider-specific impact estimates across those mini-studies (using the same approach described in Chapter II). This approach allows us to calculate unbiased impact estimates for the subgroup of students who named Provider A as their first-choice provider. We can then relate those impact estimates to provider-specific characteristics and provider-level average hours of attendance.

As discussed in Appendix D, there is apparent variation across providers in the impact estimates that they have on math and reading test scores. We investigated whether this variation was real or whether it was simply due to random chance by conducting a test for the homogeneity of impact estimates. Specifically, a simulation method was used to understand how much variation in provider-specific impact estimates one would see if the only reason for that variation was the sampling variance associated with each impact estimate. If the amount of variation that is actually observed is larger than would be expected from normal random fluctuations, then one might conclude that providers differ in terms of their impact estimates on math and reading test scores.

To conduct the simulation, we randomly generated a set of statistically independent impact estimates, each of which had a t-distribution with mean zero, a different variance (the variances were equal to the variances observed in the actual impact estimates), and a different number of degrees of freedom (based on the number of degrees of freedom from each provider-specific impact estimate). For each simulation replication we then calculated the variance of the simulated impact estimates. This process was repeated 10,000 times. We then checked whether the observed variation across provider impact estimates fell within the upper 5 percent of those 10,000 simulated variances.

This approach differs from the standard chi-square test used to test for homogeneity of impact estimates (see Equation [14.6] in Cooper, Hedges, and Valentine, 2009), which requires an assumption of normality of the individual tests. Because the t-distribution with a low number of degrees of freedom has fatter tails than the normal distribution, we would expect to see greater impact estimate variance across providers than the normality assumption would imply. The simulation-based *p*-values described earlier take this issue into account.

APPENDIX C

DIAGNOSTIC ANALYSES

We conducted diagnostic analyses for every mini-study to assess potential threats to validity. For each mini-study we conducted analyses focused on three issues: (1) integrity of the assignment variable, (2) continuity of the relationship between the assignment variable and the outcome (which is similar to a baseline equivalence analysis for an RCT), and (3) students with missing outcome data.

A. Integrity of the Assignment Variable

A key condition for an RD design to produce unbiased estimates of effects of an intervention is that there was no systematic manipulation of the assignment variable. This situation is analogous to the nonrandom manipulation of treatment and comparison group assignments under an RCT. In an RD design, manipulation means that scores for some units were systematically changed from their true values to influence treatment assignments. With nonrandom manipulation, the true relationship between the outcome and assignment variable can no longer be identified, which could lead to biased impact estimates.

In this study, manipulation is nearly impossible, because the assignment variables are standardized test scores that come from existing district databases, and because the cutoff values were determined *after* the tests were administered and scores were recorded. In other words, scorers did not have an opportunity to change a true score in order to influence treatment assignment because the scoring process happened before the cutoff value of the assignment variable was set. In addition, scorers did not have an incentive to change a true score because the scorers were completely independent and had no vested interest in the outcome of this evaluation (that is, tests were scored not by district SES staff or teachers, but by standardized testing agencies).

Although manipulation should not be possible, we did test for discontinuities in the density of the assignment variable at the cutoff value. A discontinuity in the density of the assignment variable is a potential symptom of manipulation (if the density right around the cutoff is not continuous, it could be due to scorers manipulating scores in order to make some students just eligible for the treatment group) but it could also be a symptom of other unexpected issues with the data. McCrary (2008) suggests a test of whether the density of the assignment variable is continuous over the range of values covered by the sample. The McCrary test is based on an estimator for the discontinuity in the density function of the assignment variable at the cutoff. The null hypothesis is that the discontinuity is zero.

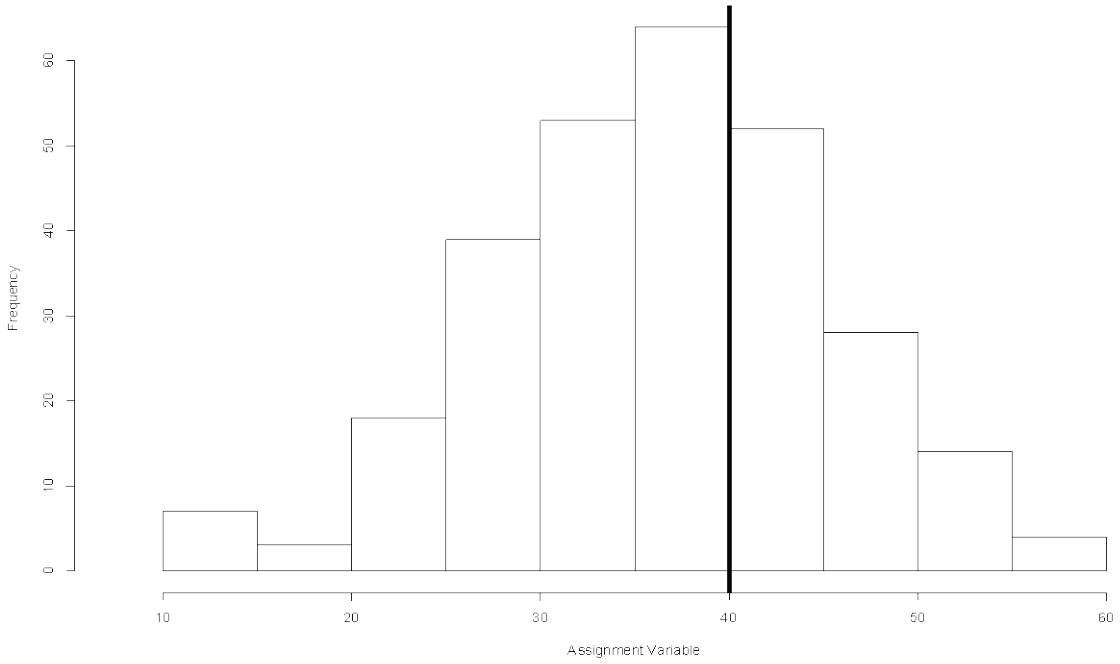
Of the 21 RD cutoff values, a McCrary statistic could be calculated for 13 (in some cases the sample sizes were too small to support calculating the McCrary statistic). Of those cases in which the statistics could be calculated, 3 were statistically significant. The 3 statistically significant McCrary statistics came from grade 3 in District 1 and grades 3 and 4 in District 6. Histograms of the assignment variable for these three cases are shown in Figure C.1. A dark vertical line indicates the cutoff value of the assignment variable. To provide the clearest view of any difference in density on either side of the cutoff value, the histograms were constructed so that the cutoff would always form the boundary between two bins. The difference in the vertical height of the bars on either side of the cutoff provides evidence regarding the existence of a discontinuity in the density at the cutoff value. Because these bars correspond to discrete bins, there will inevitably be discrete differences in the heights of these bars at the boundary of each bin. The question, then, is whether the difference in the height of the bars on either side of the cutoff is noticeably greater than at any other boundary between bins, and greater than what we

would expect given the location of the cutoff in the distribution of the assignment variable: Because the assignment variables appear to be approximately normally distributed, we might expect a larger height difference if the cutoff happens to be closer to an inflection point in the probability density function). In the first case the difference in height on either side of the cutoff is small relative to other bin boundaries. In the second and third cases the differences in height are relatively large, but roughly consistent with what we might expect, given that the cutoff values in these cases are closer to what appear to be inflection points in the probability density functions.

We hypothesize that the McCrary test might be statistically significant in these cases due to a general lack of continuity in the assignment variable at all points (as opposed to a discontinuity in the density of the assignment variable that is peculiar to the cutoff value). In District 1, grade 3, there are 6.2 times as many students as there are unique values of the assignment variable. In District 6, grade 3, there are 5.4 times as many students as there are unique values of the assignment variable. In District 6, grade 4 there are 13.4 times as many students as there are unique values of the assignment variable. We account for this type of general lack of continuity in the assignment variable by adjusting standard errors for clustering as in Lee and Card (2008).

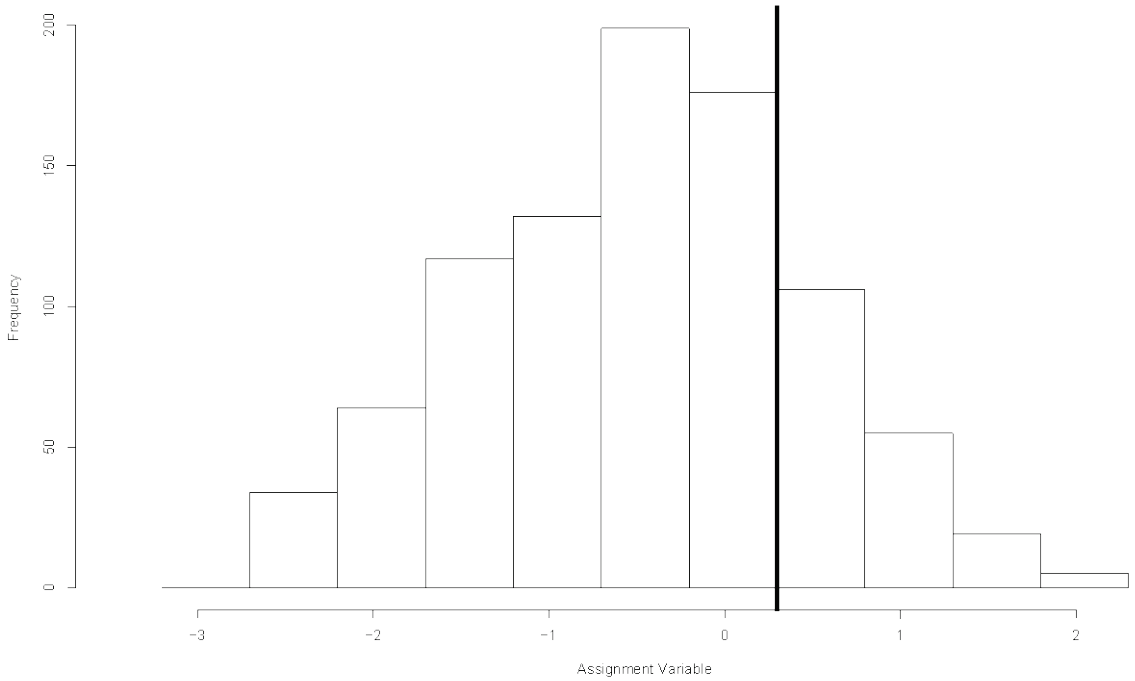
In mini-studies in which it was not possible to calculate the McCrary statistic due to small sample sizes, we examined plots similar to those in Figure C.1. None of those plots showed obvious discontinuities in the density of the assignment variable. The issue of a general lack of continuity in the assignment variable described previously also exists in other mini-studies, but in all cases in which it exists, standard errors are adjusted for clustering as in Lee and Card (2008).

Figure C.1. Kernel Density Plots Of Assignment Variables

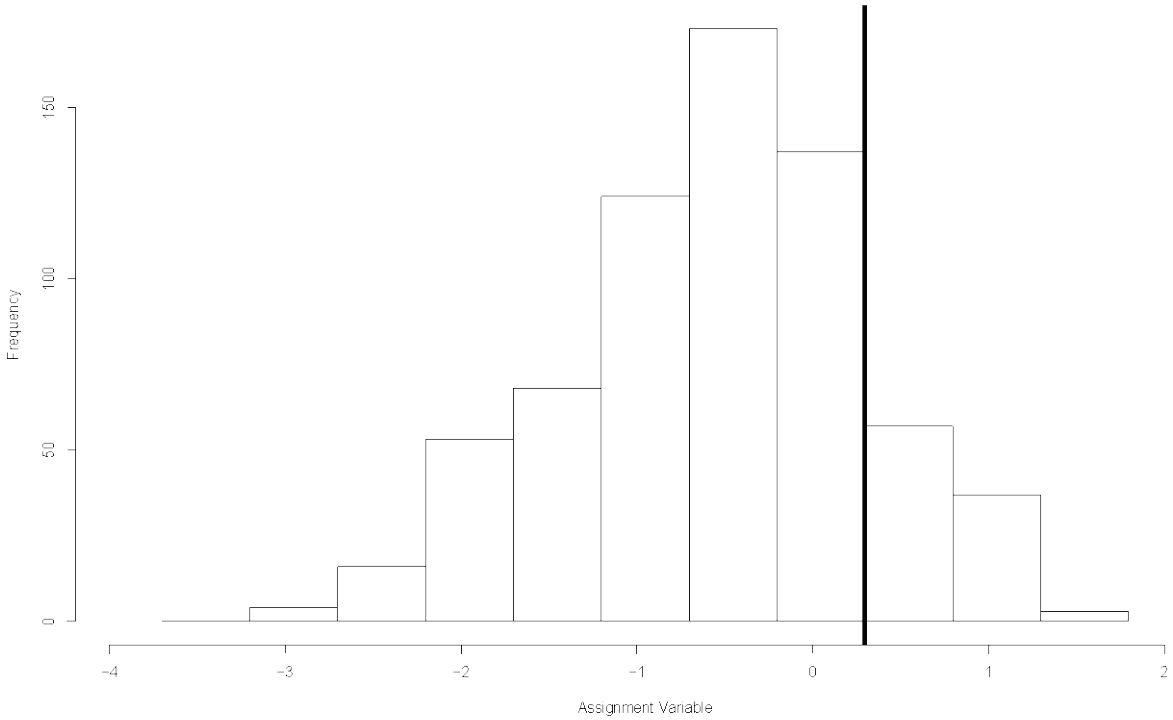


District 1, grade 3

Figure C.1 *(continued)*



District 6, grade 3



District 6, grade 4

Figure C.1 *(continued)*

Source: School district records.

Notes: These figures show the density of the assignment variable for three selected mini-studies. The assignment variable is plotted on the horizontal axis (see Appendix A for details about the assignment variables used in each district).

B. Continuity of the Relationship Between the Assignment Variable and the Outcome

To obtain a rigorous impact estimate under an RD design, there must be evidence that, in the absence of the intervention, there would be a smooth relationship between the outcome and the forcing variable at the cutoff score. This condition is needed to ensure that any observed discontinuity in the outcomes of treatment and comparison group units at the cutoff can be attributable to the intervention. This smoothness condition cannot be checked directly, although we can indirectly assess the smoothness condition by calculating impact estimates on pre-intervention characteristics of students (including past test scores that were not used as the assignment variable). This is similar to a baseline equivalence analysis for an RCT.

Across all RD mini-studies, we conducted 213 tests for impacts on pre-intervention student demographic characteristics and 76 tests for impacts on pretest scores. Of the tests for impacts on demographic characteristics, 15 were statistically significant (by chance, we would expect to find only 11 statistically significant differences). Of the tests for impacts on pretest scores, 6 were statistically significant (only 4 would be expected by chance). Table C.1 shows, by school district, the number of tests conducted, the number that are statistically significant, and the number of pretest differences less than -0.25 standard deviations, between -0.25 and 0.25 standard deviations, and the number greater than 0.25 standard deviations.

Although there were some statistically significant impact estimates on pretest scores within particular mini-studies, the aggregate impact estimates on the 2008 math and reading test scores across all mini-studies showed that the treatment and comparison groups were equivalent at baseline. Specifically, the aggregate impact estimate on 2008 math test scores was 0.02 and was not statistically significant (p-value = 0.80), and the aggregate impact estimate on 2008 reading test scores was 0.02 and was not statistically significant (p-value = 0.78). These aggregated results were calculated as a weighted average of the impact estimates on pretest scores for each mini-study, where the weight was equal to the inverse of the variance of that mini-study's impact estimate on the pretest score.

C. Students with Missing Outcome Data

As in an RCT, differences between the treatment and comparison groups in terms of the types of students who leave the study (meaning they do not have a posttest score) could bias impact estimates. Unlike in an RCT, the concern with sample attrition (i.e., missing outcome data) in an RD study applies primarily at the cutoff, rather than across the whole data set. We assessed attrition at the cutoff in two ways. First, we calculated attrition rates for the treatment

Table C.1. Differences in Baseline Covariates Between the Treatment and Comparison Groups, by District

	Pretest Covariates Treatment-Comparison Group Differences			Demographic Covariates Treatment-Comparison Group Differences	
District	Number that Are Statistically Significant	Number that Are Less than -0.25 or Greater than 0.25 Standard Deviations	Number that Are Between -0.25 and 0.25 Standard Deviations	Number Calculated	Number that Are Statistically Significant
1	0	7	15	44	0
2	3	3	+	16	+
3	0	9	8	33	0
4	0	+	16	88	12
5	0	+	5	8	0
6	3	4	4	24	+

Source: School district records.

Note: Treatment-comparison group differences are calculated using the benchmark approach to calculating RD impact estimates, except that no additional covariates are included in the model. Differences are examined separately for each variable and grade.

+ Suppressed to respect respondent confidentiality.

and comparison groups within the IK bandwidth; those numbers are shown in Table C.2 for each mini-study. For the study as a whole, the attrition rates for math test scores were four percent in the treatment group and two percent in the comparison group. Attrition rates for reading test scores were three percent in the treatment group and two percent in the comparison group.

Second, we calculated RD impact estimates on attrition in each mini-study using the IK algorithm to select an optimal bandwidth. We aggregated attrition impact estimates across mini-studies using the sample size within the IK bandwidth as the weight. Impact estimates on nonresponse by mini-study and test score outcome (reading and math) are shown in Table C.3. The aggregate impact estimate on nonresponse for reading test scores is 0.05 and is statistically

Table C.2. Attrition Rates

	Initial Sample Sizes within Bandwidth		Analysis Sample Sizes within Bandwidth		Nonresponse Rates within Bandwidth	
Mini-Study Description	Treatment	Comparison	Treatment	Comparison	Treatment	Comparison
District 1 / grade 3 / math	158	78	148	75	0.06	0.04
District 1 / grade 3 / reading	130	67	124	64	0.05	0.04
District 1 / grade 4 / math	113	60	100	58	0.12	0.03
District 1 / grade 4 / reading	87	53	76	50	0.13	0.06
District 1 / grade 5 / math	64	45	57	44	0.11	0.02
District 1 / grade 5 / reading	70	51	61	49	0.13	0.04
District 1 / grade 6 / math	50	36	45	34	0.10	0.06
District 1 / grade 6 / reading	54	38	51	36	0.06	0.05
District 1 / grade 7 / math	45	31	41	30	0.09	0.03
District 1 / grade 7 / reading	36	28	33	27	0.08	0.04
District 1 / grade 8 / math	46	33	41	33	0.11	0.00
District 1 / grade 8 / reading	35	25	31	25	0.11	0.00
District 2 / grade 3 / math	15	15	9	14	0.40	0.07
District 2 / grade 3 / reading	16	16	10	15	0.38	0.06
District 2 / grade 4 / math	96	96	82	96	0.15	0.00
District 2 / grade 4 / reading	115	97	100	97	0.13	0.00
District 3 / grade 4 / math	69	7	69	7	0.00	0.00
District 3 / grade 4 / reading	35	5	35	5	0.00	0.00
District 3 / grade 5 / math	37	10	37	10	0.00	0.00
District 3 / grade 5 / reading	57	11	57	11	0.00	0.00
District 3 / grade 6 / math	31	18	30	18	0.03	0.00
District 3 / grade 6 / reading	25	14	24	14	0.04	0.00
District 4 / grade 3 / math	108	108	108	108	0.00	0.00
District 4 / grade 3 / reading	111	111	111	111	0.00	0.00
District 4 / grade 4 / math	124	83	123	81	0.01	0.02
District 4 / grade 4 / reading	87	83	86	81	0.01	0.02
District 4 / grade 5 / math	161	131	161	127	0.00	0.03
District 4 / grade 5 / reading	182	131	182	127	0.00	0.03
District 4 / grade 6 / math	342	144	339	141	0.01	0.02
District 4 / grade 6 / reading	264	144	260	141	0.02	0.02
District 4 / grade 7 / math	64	64	63	63	0.02	0.02
District 4 / grade 7 / reading	186	68	184	67	0.01	0.01
District 4 / grade 8 / math	60	60	60	58	0.00	0.03
District 4 / grade 8 / reading	123	63	121	60	0.02	0.05
District 5 / grade 3 / math	400	328	391	324	0.02	0.01
District 5 / grade 3 / reading	363	303	354	299	0.02	0.01
District 6 / grade 3 / math	355	158	345	154	0.03	0.03
District 6 / grade 3 / reading	342	153	333	149	0.03	0.03
District 6 / grade 4 / math	310	94	302	92	0.03	0.02
District 6 / grade 4 / reading	280	86	272	84	0.03	0.02
District 6 / grade 5 / math	290	73	280	70	0.03	0.04
District 6 / grade 5 / reading	160	65	157	64	0.02	0.02
Overall Study, Reading	2,758	1,612	2,662	1,576	0.03	0.02
Overall Study, Math	2,938	1,672	2,831	1,637	0.04	0.02

Source: School district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cutpoint/grade combination. For each mini-study, the intent-to-treat impact estimate is calculated using a linear functional form within an optimal bandwidth selected using the Imbens & Kalyanaraman (2009) algorithm.

significant (p-value = 0.03). Findings for math test score nonresponse are identical. As Table C.3 indicates, there are very large impact estimates on attrition in District 2 mini-studies.

Excluding District 2, the aggregate impact estimate on attrition for reading test scores is 0.02 (p-value = 0.42) and the aggregate impact estimate on attrition for math test scores is 0.02 (p-value = 0.36). In Appendix E we conduct two sensitivity analyses: (1) we report overall impact estimates of SES on math and reading test scores excluding District 2 and (2) we report overall impact estimates of SES on math and reading test scores excluding all mini-studies in which the absolute value of the impact estimate on attrition is greater than 0.10.

Table C.3. Intent-to-Treat Impact Estimates on Reading and Math Test Nonresponse

Mini-Study Description	Adjusted Mean Nonresponse Rate		ITT Impact Estimate (Effect Size Units)	p-Value
	Treatment	Comparison		
District 1 / grade 3 / math	0.00	0.01	-0.02	0.75
District 1 / grade 3 / reading	0.05	0.01	0.04	0.48
District 1 / grade 4 / math	0.09	0.01	0.08	0.83
District 1 / grade 4 / reading	0.08	0.00	0.07	0.86
District 1 / grade 5 / math	0.27	-0.02	0.29	0.50
District 1 / grade 5 / reading	0.23	-0.02	0.25	0.55
District 1 / grade 6 / math	0.02	0.04	-0.03	0.95
District 1 / grade 6 / reading	0.02	0.04	-0.03	0.93
District 1 / grade 7 / math	0.12	-0.03	0.15	0.67
District 1 / grade 7 / reading	0.12	-0.03	0.15	0.67
District 1 / grade 8 / math	0.19	0.00	0.19	0.61
District 1 / grade 8 / reading	0.07	0.00	0.07	0.86
District 2 / grade 3 / math	0.98	0.10	0.89*	0.00
District 2 / grade 3 / reading	0.98	0.10	0.89*	0.00
District 2 / grade 4 / math	0.58	0.00	0.58*	0.00
District 2 / grade 4 / reading	0.58	0.00	0.58*	0.00
District 3 / grade 4 / math	0.00	0.00	0.00	0.99
District 3 / grade 4 / reading	0.00	0.00	0.00	0.99
District 3 / grade 5 / math	0.00	0.00	0.00	0.99
District 3 / grade 5 / reading	0.00	0.00	0.00	0.99
District 3 / grade 6 / math	-0.03	0.00	-0.03	0.89
District 3 / grade 6 / reading	-0.03	0.00	-0.03	0.89
District 4 / grade 3 / math	0.00	0.00	0.00	0.31
District 4 / grade 3 / reading	0.00	0.00	0.00	0.96
District 4 / grade 4 / math	0.05	0.02	0.03	0.58
District 4 / grade 4 / reading	0.05	0.02	0.03	0.58
District 4 / grade 5 / math	0.00	-0.02	0.02	0.32
District 4 / grade 5 / reading	0.00	-0.02	0.02	0.32
District 4 / grade 6 / math	0.03	-0.01	0.04	0.17
District 4 / grade 6 / reading	0.03	-0.01	0.04	0.18
District 4 / grade 7 / math	0.01	0.04	-0.03	0.43
District 4 / grade 7 / reading	0.02	0.04	-0.02	0.73
District 4 / grade 8 / math	0.00	0.05	-0.05	0.27
District 4 / grade 8 / reading	0.00	0.03	-0.03	0.52
District 5 / grade 3 / math	0.02	0.03	-0.01	0.79
District 5 / grade 3 / reading	0.02	0.03	-0.01	0.83
District 6 / grade 3 / math	0.00	0.01	-0.01	0.87
District 6 / grade 3 / reading	0.00	0.00	0.00	0.93
District 6 / grade 4 / math	0.04	0.01	0.02	0.51
District 6 / grade 4 / reading	0.04	0.01	0.02	0.52
District 6 / grade 5 / math	0.03	-0.02	0.05	0.11
District 6 / grade 5 / reading	0.03	-0.02	0.05	0.16

Table C.3 *(continued)*

Source: Math and reading test scores from school district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cutoff-value/grade combination. For each mini-study, the intent-to-treat (ITT) impact estimate is calculated using a linear functional form within an optimal bandwidth selected using the Imbens-Kalyanaraman (2009) algorithm and standard errors are adjusted to account for clustering of observations within unique values of the assignment variable as in Lee and Card (2008).

*Significantly different from zero at the 0.05 level, two-tailed test.

APPENDIX D

EXPLORATORY ANALYSES

In this appendix, we provide detailed information on several of the secondary and exploratory analyses that we described in summary form in the latter part of Chapter IV. These include:

- Florida-specific impact estimates
- Impact estimates for policy-relevant subgroups of students
- Estimates of the effect of participating in SES (rather than being offered a slot), a.k.a. estimates of the effect of treatment on the treated (TOT)
- Analysis of the relationship between hours of service and impacts
- Analyses of relationships between provider-specific impacts and provider characteristics and practices

A. Impacts in Florida Districts

Recall from Chapter II that Florida districts had a stronger incentive than districts in other states to ensure that as many students as possible were participating in SES as a result of statewide rules regarding unspent SES funds. They also may be particularly relevant to predicting current and future impacts nationwide, because some aspects of Florida's regulatory regime have since been enacted nationally through new regulatory guidance from the U.S. Department of Education. Given the potential importance of Florida, we conducted a supplemental analysis that included only the four participating districts in Florida.

The ITT impact estimate of offering SES on math test scores in the Florida districts was 0.05 standard deviations and was not statistically significant (p-value = 0.20). The ITT impact estimate on reading test scores was -0.05 and not statistically significant (p-value = 0.27).

B. Impacts on Policy-Relevant Subgroups

As discussed in Chapter IV, the effects of SES on at-risk subgroups of students are of interest because NCLB specifically requires Title I schools to reach AYP proficiency standards for subgroups. Tables D.1 and D.2 show impact estimates on reading and math test scores (respectively) by student subgroup. We conducted ITT analyses for a variety of different subgroups of students, including racial/ethnic minorities, English-language learners, and students with disabilities, and students (for Florida districts only) who were members of any group for which the student's own school had fallen short of AYP (which often included multiple designated subgroups). The tables also show the regression-adjusted mean outcomes (that is, the predicted outcomes) for the treatment and comparison groups, as well as the difference in impact estimates across categories within a subgroup. The p-values associated with these differences come from a test of whether the impact estimates across categories are statistically different from each other. Out of 12 comparisons, one was statistically significant: the difference in math impact estimates for special education vs. non-special education students (-0.25 for special education students, 0.05 for non-special education students, p-value of 0.01 on the difference).

Table D.1. Estimated Impacts of SES on Reading Test Scores of Students Near the Cutoff for Services, by Student Subgroups

Student Subgroup	Adjusted Mean Outcomes		ITT Impact Estimate (Effect Size Units)	*p*-Value
	Treatment	Comparison		
Race/Ethnicity				
White	-0.01	-0.07	0.06	0.64
Black	0.02	0.08	-0.06	0.38
Hispanic	0.20	0.31	-0.12	0.08
White-black difference			0.12	0.42
White-Hispanic difference			0.18	0.24
Black-Hispanic difference			0.06	0.56
English Language Learner (ELL) status				
ELL students	-0.52	-0.30	-0.22	0.16
Non-ELL students	0.18	0.23	-0.05	0.30
Difference			-0.17	0.32
Disability/Special Education Status				
Disabled/special education students	0.37	0.46	-0.09	0.26
Non-disabled/non-special education students	-0.01	0.04	-0.05	0.32
Difference			-0.04	0.68
AYP Status (Florida districts only)				
Meeting AYP	0.30	0.30	-0.01	0.94
Not meeting AYP	0.25	0.44	-0.20*	0.03
Difference			0.19	0.08

Source: Test scores from school district records.

Note: A subgroup impact estimate was calculated for every mini-study by restricting the analysis sample for that mini-study to the students in the subgroup of interest. Impact estimates were aggregated using the same approach as for overall impact estimates (see note to Table IV.1). All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect size units. Adjusted mean outcomes for the treatment and comparison groups are equal to the estimated intercept terms from the regressions of the outcome (student test scores) on the assignment variable and other covariates that were estimated separately on either side of the cutoff value. The outcomes were standardized using state-grade-level means and standard deviations.

*Significantly different from zero at the 0.05 level, two-tailed test.

AYP = adequate yearly progress

C. Impacts of Treatment on the Treated

The process we used to infer the cutoff value was constructed to minimize the extent to which children below the cutoff value do not participate in SES, but this kind of "fuzziness" does exist in the data. In the context of an RCT this would be called "noncompliance". In both RCTs and RDDs it is possible to calculate the impact of "treatment on the treated" (TOT), sometimes called the complier average causal effect (CACE), which adjusts for noncompliance/fuzziness.

For a TOT impact from a fuzzy RDD to be internally valid, we need only one additional requirement beyond those for the ITT to be valid. The additional requirement is that there must be a sufficiently large discontinuity in the probability of receiving treatment at a cutoff value of an assignment variable. The implication of picking the wrong cutoff value would be that this requirement might not hold—that is, that there would not be a discontinuity at the cutoff in the proportion of students receiving treatment (or that the discontinuity might be very small).

Table D.2. Estimated Impacts of SES on Math Test Scores of Students Near the Cutoff for Services, by Student Subgroups

Student Subgroup	Adjusted Mean Outcomes		ITT Impact Estimate (Effect Size Units)	*p*-Value
	Treatment	Comparison		
Race/Ethnicity				
White	-0.44	-0.29	-0.15	0.21
Black	-0.16	-0.06	-0.09	0.24
Hispanic	0.28	0.28	0.00	0.98
White-black difference			-0.06	0.68
White-Hispanic difference			-0.16	0.29
Black-Hispanic difference			-0.09	0.40
English Language Learner (ELL) Status				
ELL students	-0.42	-0.56	0.14	0.42
Non-ELL students	0.08	0.08	0.00	0.96
Difference			0.14	0.45
Disability/Special Education Status				
Disabled/special education students	0.37	0.61	-0.25*	0.01
Non-disabled/non-special education students	0.03	-0.02	0.05	0.33
Difference			-0.30*	0.01
AYP Status (Florida districts only)				
Meeting AYP	0.36	0.34	0.01	0.87
Not meeting AYP	0.33	0.26	0.06	0.46
Difference			-0.05	0.68

Source: Test scores from school district records.

Note: A subgroup impact estimate was calculated for every mini-study by restricting the analysis sample for that mini-study to the students in the subgroup of interest. Impact estimates were aggregated using the same approach as for overall impact estimates (see note to Table IV.1). All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect size units. Adjusted mean outcomes for the treatment and comparison groups are equal to the estimated intercept terms from the regressions of the outcome (student test scores) on the assignment variable and other covariates that were estimated separately on either side of the cutoff value. The outcomes were standardized using state-grade-level means and standard deviations.

*Significantly different from zero at the 0.05 level, two-tailed test.

AYP = adequate yearly progress

We find that there is a substantial difference in participation rates at the cutoff value. Overall, 86 percent of applicants who were below the cutoff on the RD assignment variable actually did participate in SES. In order to calculate the impact of participating in SES, we conducted a TOT impact analysis—also called a "fuzzy" RD analysis—for each mini-study (Imbens & Lemieux, 2008). In a TOT analysis, the ITT impact of SES is rescaled by dividing the impact estimate by the difference in predicted participation rates between students just below the cutoff and students just above the cutoff. Stated differently, we divided the ITT impact estimate by the impact estimate on the SES participation rate in each mini-study. This is directly analogous to the instrumental variables approach used to calculate TOT impact estimates in experimental studies. We estimated the impact on the participation rate in each mini-study using the same techniques described in Chapter II for estimating impacts on test scores. We then calculated a weighted average of the mini-study specific TOT impact estimates using a weight equal to the inverse of the variance of each TOT impact estimate (similar to our approach to aggregating the ITT impact estimates). Note that because the weight used to aggregate TOT impact estimates is not the same as the weight used to aggregate ITT impact estimates, the aggregate TOT impact estimate is not a simple rescaling of the aggregate ITT impact estimate.

We defined participation as receiving a positive number of tutoring hours, regardless of subject area: that is, a participant is defined as anyone with total service hours > 0. This reduces the size of the treatment group (as compared to the ITT impact estimate in Chapter IV) because some students who were assigned to services did not actually receive them. Table D.3 shows the overall TOT impact estimates on math and reading test scores. The impact estimate on reading test scores is -0.10 and not statistically significant (p-value = 0.25) and the impact estimate on math test scores is 0.11 and not statistically significant (p-value = 0.21). TOT impact estimates were also not statistically significant when using alternative definitions of SES participation (see Appendix E).

Table D.3. Overall Treatment-on-Treated (TOT) Impact Estimates of SES on Reading and Math Test Scores of Students Near the Cutoff for Services

Outcome	TOT Impact Estimate (Effect Size Units)	*p*-Value
Reading Test Scores	-0.10	0.25
Math Test Scores	0.11	0.21

Source: Math and reading test scores from school district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cutoff-value/grade combination. For each mini-study, the TOT impact estimate is calculated as the ratio of the ITT impact estimate to the impact estimate on participation. Aggregate impact estimates are calculated as a weighted average of the impact estimates for each mini-study, where the weight is equal to the inverse of the variance of that mini-study's TOT impact estimate. All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect size units.

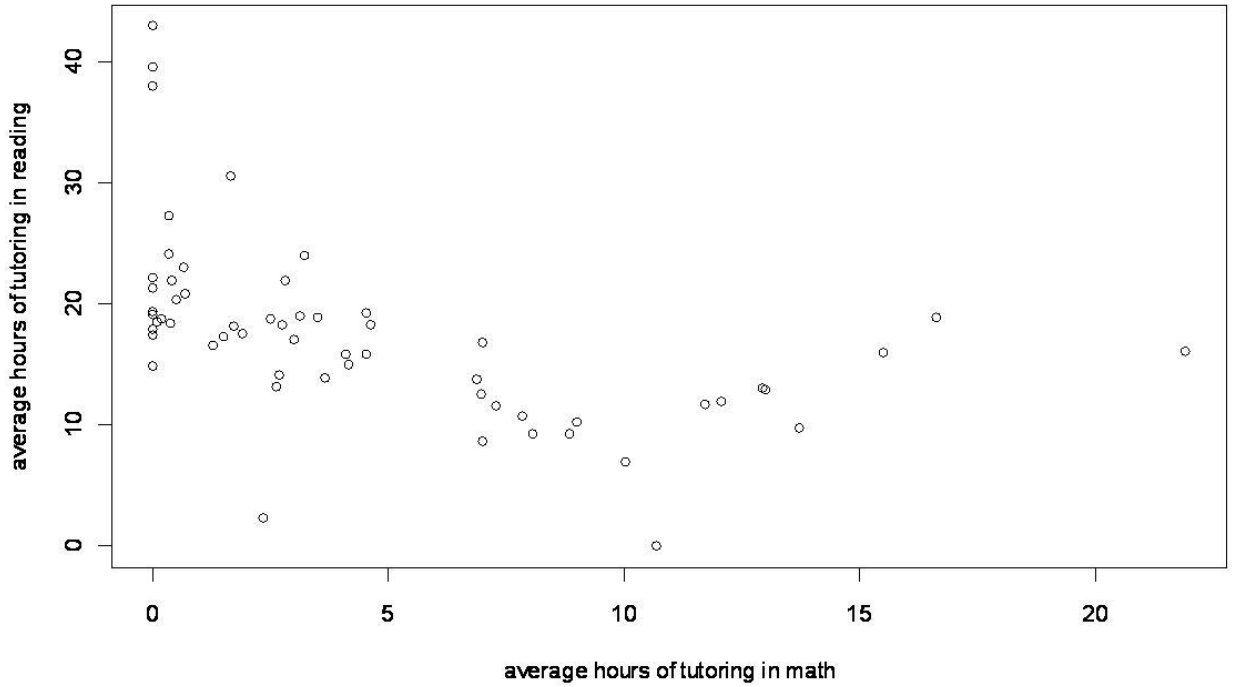
*Significantly different from zero at the 0.05 level, two-tailed test.

D. Impact Estimates for Students Who Receive More Hours of Tutoring

Although the overall ITT impact estimates of SES on math and reading test scores are not statistically significantly different from zero, it could be the case that particular providers had impacts that were more positive (or larger) or more negative in a particular subject than other providers as a result of providing more tutoring hours in that subject. To investigate this research question, we first examined whether there was variation across providers in the number of math and reading service hours provided to students. We then looked for a relationship between provider-specific impact estimates and the average number of math or reading service hours students received.

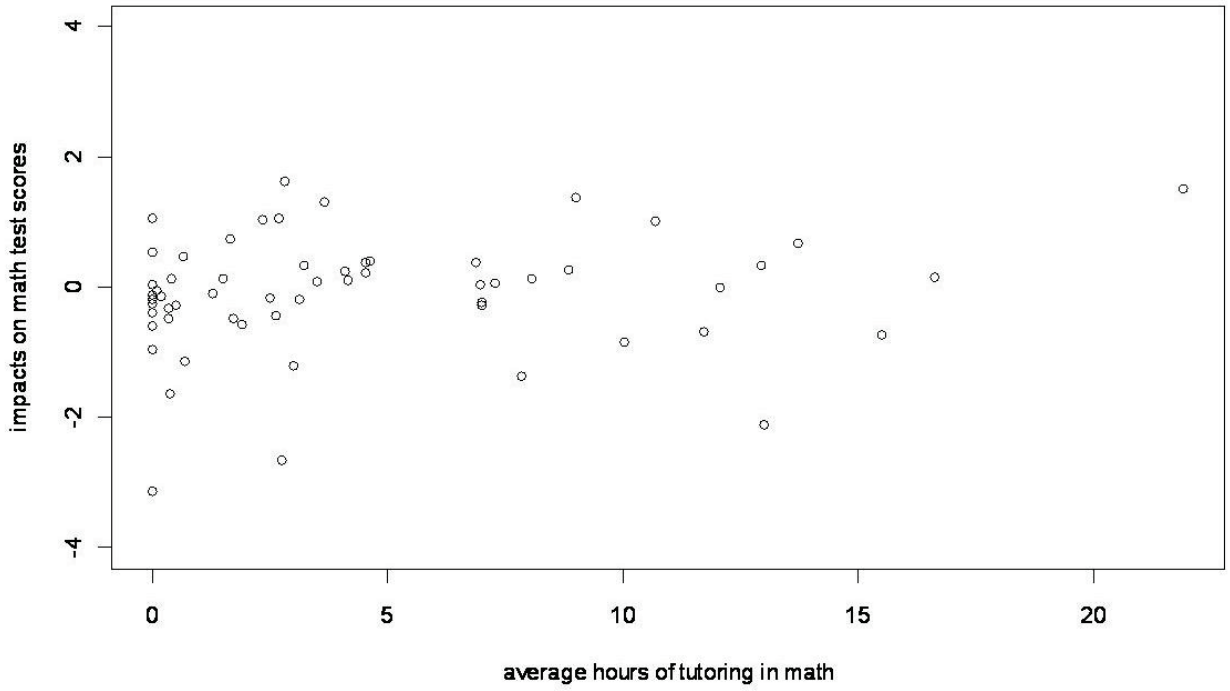
Variation in the provider-specific average of math and reading hours attended by SES participants is shown in Figure D.1. Provider-specific average math or reading hours are calculated as the average number of math or reading hours among students attending each provider. Average math hours ranged from 0 to 27 across providers and average reading hours ranged from 0 to 43. The relationship between provider-specific impact estimates on math test scores and hours of math receipt is shown in Figure D.2 and the relationship between impact estimates on reading test scores and reading hours is shown in Figure D.3. Figures D.1, D.2, and D.3 show findings for 59 providers. Other providers were excluded from these analyses for three reasons. First, we excluded 42 providers that did not complete the provider survey. We also excluded 52 providers that did not deliver services to students in our sample (meaning providers for whom there were no students with positive math or reading hours). Third, we excluded 85 providers for whom we could not estimate provider-specific impacts because of inadequate student sample sizes.

Figure D.1. Provider-Level Average Hours of Tutoring in Math and Reading



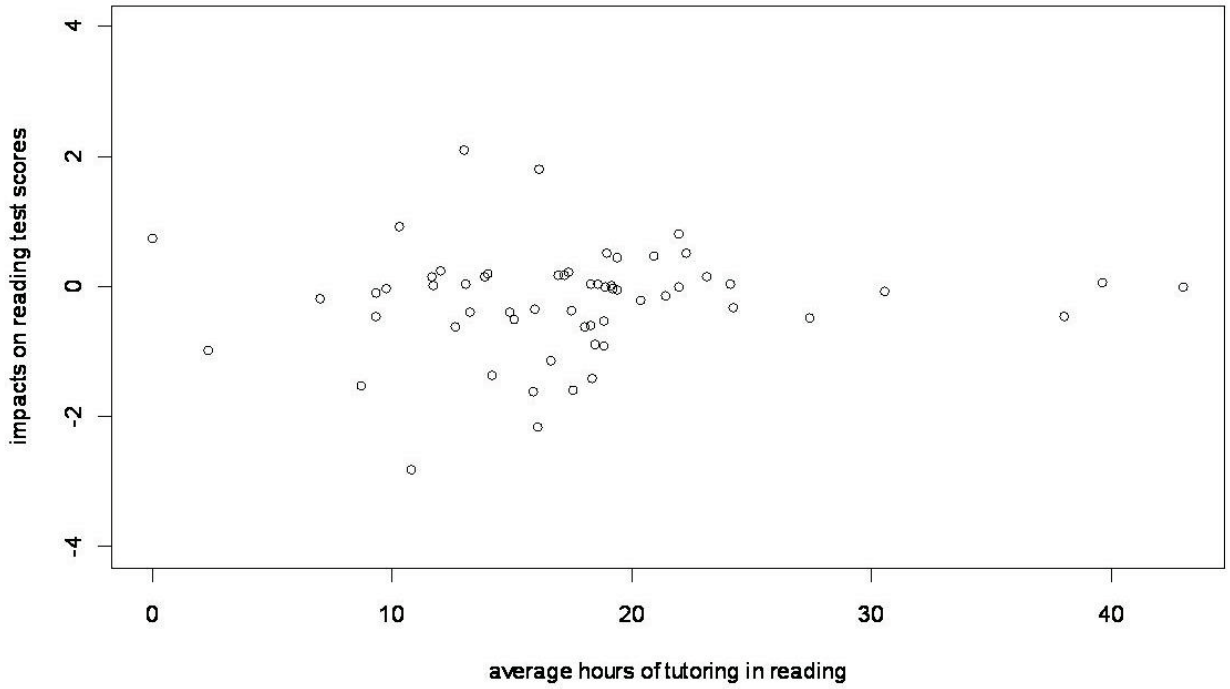
Source: School district records and SES provider survey.

Figure D.2. Provider-Specific Impact Estimates on Math Test Scores Versus Provider-Level Average Hours of Math Services



Source: School district records and SES provider survey.

Figure D.3. Provider-Specific Impact Estimates on Reading Test Scores Versus Provider-Level Average Hours of Reading Services



Source: School district records and SES provider survey.

In order to draw inferences from the patterns shown in these figures, we conducted regression analyses to assess whether there is a tradeoff between reading and math hours (as Figure D.1 seems to suggest) and whether there is any relationship between impact estimates and hours (which is difficult to discern from the figures themselves). Regarding the first question, provider-specific average reading and math hours are negatively correlated ($r = -0.50$; p-value < 0.01), indicating a tradeoff between math and reading hours across providers. We next examined whether this variation across providers in the number of math and reading hours provided to students was related to provider-specific impact estimates, using a two-step procedure. First, we estimated provider-specific impact estimates on math (or reading) test scores and provider-specific impact estimates on math (or reading) hours. Second, we regressed the estimated impacts on test scores on the estimated impacts on hours.[1] Neither relationship was statistically significant; the regression coefficient for math was 0.002 (p-value = 0.83) and the coefficient for reading was 0.0005 (p-value = 0.91). Thus, we conclude that there is no

[1] A less rigorous approach to testing for a relationship between hours of SES received and SES impacts would be to regress student-level test scores on hours of service receipt at the student level. However, that approach would lead to biased estimates because students who choose to participate for more hours may differ in unobserved ways from students who choose to participate for fewer hours. The approach we used controls for student-level bias by taking treatment status into account (in technical terms, we used treatment status as an "instrument" for provider-level service hours). However, our approach does not control for provider-level biases (for example, providers with larger impacts on hours of participation might differ in unobserved ways from providers with lower impacts on hours of participation).

statistically significant positive (or negative) relationship between hours of services received and estimated impacts on test scores.

E. Variation in Impact Estimates across Providers

Although the overall estimated impact of SES on math and reading test scores is zero, we were interested in whether some providers had larger or more positive impact estimates than others (or more negative impact estimates than others), because we know from the provider survey that providers varied in terms of the tutoring strategies used and the level of instructor experience and qualifications. We investigated this research question in two ways. First, we tested whether the variation in impact estimates across providers was greater than one would expect by chance. Second, we looked for relationships between provider impact estimates and provider characteristics.

We observe no more variation in impact estimates across providers than what we would expect to see by chance. We conducted a test for the homogeneity of impact estimates and failed to reject the null hypothesis of impact homogeneity (math impact p-value = 0.94; reading impact p-value = 0.88).[2]

The inability to confirm that any of the providers produced statistically significant achievement impacts does not bode well for an exercise of correlating estimated impacts with provider characteristics, particularly given the high p-values on those tests. Nonetheless, the failure to reject the test of impact homogeneity across providers does not tell us that there is no true variation in impacts—only that we could not confirm the existence of variation. Prior research has found that the homogeneity test can have low power to detect true heterogeneity in impacts (Huedo-Medina et al. 2006)[3]. Also, we have verified through simulations that it is possible to find a statistically significant correlation between impacts and provider characteristics even when we fail to reject the null hypothesis of impact homogeneity so long as variation in impacts is very highly correlated with a provider characteristic. We therefore proceeded to examine whether characteristics of providers or services were related to impacts with the understanding that such relationships could only be found if they were very strong.

The first characteristic we examined was the intensity of services, as measured both by the number of hours in a complete course of services and by the average number of hours actually attended in reading or math. Students enrolled in providers with longer courses of services did not experience achievement impacts that were significantly different from students enrolled with providers offering shorter courses. Similarly, students attending for a greater number of hours did not experience impacts that were significantly different from students attending for fewer hours.

[2] We also conducted a test for whether the variation in impact estimates across mini-studies is larger than one would expect by chance. We failed to reject the null hypothesis of impact homogeneity (math impact p-value = 0.92; reading impact p-value = 0.998).

[3] The homogeneity test that we used likely has even less power than the Q-statistic test because the Q-statistic test assumes normality of the individual impacts while our test accounted for the small sample sizes of many of the provider impacts and treated the impacts as following the *t*-distribution with appropriate degrees of freedom for each impact.

We examined a variety of other characteristics of services/providers, as described in detail in Chapter III, that could be used to select providers and monitor their services, including provider size, instructional staff qualifications, and tutoring practices (e.g., student group size of service, frequency in the use of diagnostic tests to assess student progress). Recognizing the likelihood of false positives associated with multiple comparisons, we examined all characteristics available from the SES provider survey and found that, out of 344 statistical tests conducted, 12 were statistically significant, which is fewer than the 17 that one would expect by random chance.

Finally, we compared the characteristics of providers with positive impact estimates to those of providers with negative impact estimates (regardless of whether or not these impact estimates were statistically significant). The provider type and staff qualifications of providers with positive and negative impact estimates are shown in Table D.4. None of the differences in the table are statistically significant.

Table D.4. Characteristics of SES Providers with Positive and Negative Estimated Impacts on Test Scores of Students Near the Cutoff for Services

	Providers with Estimated Impacts on Reading Test Scores that Are		Providers with Estimated Impacts on Math Test Scores that Are	
Characteristic	Positive	Negative	Positive	Negative
Organization Type				
Non-profit	8.3	4.2	4.2	8.3
Faith-based	0.0	4.2	4.2	0.0
Community-based	16.7	4.2	12.5	8.3
For-profit	66.7	83.3	75.0	75.0
Public school	4.2	0.0	4.2	0.0
College or university	0.0	4.2	0.0	4.2
Other	4.2	0.0	0.0	4.2
p-value	0.38		0.99	
Staff Required to Be Certified Teachers	29.2	45.8	33.3	41.7
p-value	0.38		0.78	
Minimum Educational Qualification for Staff Who Are Not Required to Be Certified Teachers				
Experience teaching or tutoring	5.9	15.4	6.2	14.3
Minimum score on assessment	5.9	0.0	6.2	0.0
Four-year degree related to content	11.8	23.1	12.5	21.4
Any four-year degree	11.8	15.4	0.0	28.6
Some college coursework	11.8	7.7	12.5	7.1
Two-year college degree (or equivalent)	35.3	23.1	37.5	21.4
Other	17.6	15.4	25.0	7.1
p-value	0.97		0.19	

Source: School district records and provider survey.

*Significantly different from zero at the .05 level, two-tailed test.

APPENDIX E

ASSESSING ROBUSTNESS OF IMPACT ESTIMATES

The benchmark approach to estimating ITT impacts involved a variety of methodological choices that could potentially influence the study's findings. In addition, diagnostic analyses (described in Appendix C) identified potential threats to validity in some mini-studies. To assess the robustness of the study's ITT impact findings to methodological choices and to the exclusion of mini-studies with potential threats to validity, we conducted several sensitivity analyses. To facilitate comparisons between the impact estimates from these different sensitivity tests, we report the results of these analyses in Table E.1, showing impact estimates from the benchmark model alongside impact estimates from each of the sensitivity analyses. In Sections A through F, we describe each ITT sensitivity analysis and present findings from each analysis.

As described in Appendix D, we conducted a treatment-on-treated (TOT) impact analysis that involved rescaling the ITT impact estimate of SES. Specifically, we divided the ITT impact estimate by the impact estimate on the SES participation rate in each mini-study. As a sensitivity check on the study's TOT impact findings, we calculated TOT impact estimates using several different definitions of SES participation. We present the results of these sensitivity analyses in Section G.

A. Alternative Weighting Approaches

The benchmark approach to aggregating impact estimates across mini-studies is to calculate a weighted average impact estimate for which the weight is the inverse of the variance of each impact estimate. If impact estimates vary systematically across mini-studies, then alternative weighting approaches could yield different aggregate impact estimates. In Table E.1 we report impact estimates for two alternative weighting methods: (1) "equal weight," which means each mini-study receives a weight of one; and (2) "sample size weight," which means that each mini-study is weighted by the number of students included in the impact analysis (this sample size is equal to the number of students within the IK bandwidth). Impact estimates on math and reading test scores remain statistically insignificant for each of these alternative weighting approaches.

B. Alternative RD Estimation Approaches

We examine two alternatives to our benchmark approach of estimating RD impacts using the IK bandwidth selection algorithm. The first approach is to estimate a linear relationship between the outcome and the assignment variable using all available data. The second approach is to fit an optimal polynomial using all available data. Before discussing the results of the sensitivity analyses, we first state our reasons for selecting the IK bandwidth as our benchmark approach. (Appendix B describes the IK bandwidth and the procedure we used to select the optimal functional form.)

We chose the IK bandwidth as our benchmark approach to estimating RD impacts both for its theoretical appeal and because it performed better than alternative approaches in analyses of data from earlier studies. From a theoretical perspective the IK bandwidth is appealing because it is designed specifically to minimize the mean squared error (MSE) of the RD impact estimate. To our knowledge, no other estimation approach minimizes the MSE.

Table E.1. Sensitivity Analyses: Overall Intent to Treat Impact Estimates of SES on Reading and Math Test Scores on Students Near the Cutoff for Services

		Weighting Methods		RD Analytic Methods					
Outcome	Benchmark	Equal Weight	Sample Size Weight	Linear	Optimal Polynomial	Impute Missing Outcome Values	No Covariates	Exclude District 2	Exclude Mini-Studies with Differential Attrition > 0.10
Reading Test Scores	-0.03	0.02	-0.04	0.02	-0.03	-0.03	-0.05	-0.04	-0.03
Math Test Scores	0.05	0.02	0.02	0.00	0.09	0.06	0.00	0.04	0.04

Source: Math and reading test scores from school district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cutoff-value/grade combination. Aggregate impact estimates for the mini-studies are calculated as a weighted average of the impact estimates for each mini-study, in which the weights are either equal to the inverse of the variance of each mini-study's impact estimate (variance-minimizing weight), one (equal weight), or the number of students included in the RD bandwidth in each mini-study (sample size weight). For each mini-study, the intent-to-treat (ITT) impact estimate is calculated using either a linear functional form within an optimal bandwidth selected using the Imbens-Kalyanaraman (2009) algorithm (benchmark), a linear functional form using all the data, or an optimal polynomial using all the data. Standard errors for each mini-study are adjusted to account for clustering of observations within unique values of the assignment variable as in Lee and Card (2008).

*Significantly different from zero at the .05 level, two-tailed test.

RD = regression discontinuity.

To assess the performance of the IK algorithm using actual data, we conducted an analysis of data from several past education studies involving baseline and follow-up test scores,[1] but in which treatment was *not* assigned using the baseline test score (meaning that there should be no RD impacts). We treated the baseline test score in each study as the RD assignment variable and the follow-up test score as the outcome, and we calculated RD impact estimates for multiple artificial cutoffs. Specifically, for each of 27 combinations of outcomes and forcing variables we calculated RD impact estimates at every value of the artificial assignment variable for which an RD impact could be estimated (a total of 181 impact estimates). Because these data came from studies in which students were not being assigned to participate in an intervention using the baseline test score, we know that there were no RD impacts of the forcing variable at any of these cutoffs. Therefore, all of the RD impact estimates should be close to zero and only 5 percent of them should be statistically significant when conducting two-tailed tests with a significance level of 0.05. Results of this analysis are shown in Table E.2. Of the three methods examined, the IK method has the error rate (0.04) closest to 0.05.

Table E.2. Error Rates Associated with Alternative RD Estimation Approaches

RD Method	Error Rate
Linear: Use All Observations	0.27
Linear: Use Imbens-Kalyanaraman Bandwidth	0.04
Optimal Polynomial	0.08

Source: Test score data from past education studies.

RD = regression discontinuity.

We also estimated RD impacts with an optimal polynomial regression (instead of a linear regression) using all of the data. In this case, the right and left regression equations are:

$$(3) \quad Y_i^R = \beta_0^R + \sum_{j=1}^{M} \beta_1^R (X_i^R)^j + \beta_{M+1}^R Z_i^R + u^R + \varepsilon_i^R ,$$

and

$$(4) \quad Y_i^L = \beta_0^L + \sum_{j=1}^{M} \beta_1^L (X_i^L)^j + \beta_{M+1}^L Z_i^L + u^L + \varepsilon_i^L ,$$

where variables and parameters are defined as before, but now we add higher-order terms of the assignment variable (X_i).

[1] The data come from five studies conducted by Mathematica for IES. One of the studies involves several distinct substudies, yielding a total of eight studies/substudies. The tests from these studies were all math and reading tests, mostly of students in elementary schools. In most cases the tests were administered by the study, but in some instances the test scores were taken from school district records. Among the studies that administered their own tests, the tests were generally standardized tests of the sort that might be used by states and school districts (for example, the Stanford Achievement Test 9 or Stanford Achievement Test 10).

Table E.1 shows how RD impact estimates vary when using methods other than the IK bandwidth selection method. The impact estimates on math and reading test scores remain statistically insignificant using either alternative method.

C. Impute Missing Outcome Values

Our benchmark approach is to drop from the analysis students who are missing the outcome variable. As a sensitivity analysis we calculated impact estimates when imputing missing values using multiple imputations. Impact estimates on both math and reading test scores remain statistically insignificant when missing outcomes are imputed (see Table E.1).

D. Covariate Adjustment

Our benchmark impact regressions include additional covariates (such as pretests and demographic characteristics) to improve statistical precision. Dropping these covariates from the impact regressions does not change our findings—impact estimates on math and reading test scores remain statistically insignificant (see Table E.1).

E. Exclude Districts with Diagnostic Issues

Excluding districts where diagnostic analyses (see Appendix C) identified potential threats to validity results in little change in the impact estimates on math and reading test scores. Both impact estimates remain statistically insignificant with this exclusion (see Table E.1).

F. Exclude Mini-Studies with High Differential Attrition

In Appendix C we found differential attrition between the treatment and comparison groups greater than 0.10 in some mini-studies. We examine the sensitivity of findings to excluding mini-studies with differential attrition greater than 0.10 and to excluding the district with the highest differential attrition rate (Appendix Table C.3). A large difference in attrition rates could lead to biased impacts due to unobserved differences in the treatment and comparison groups, although removing districts from the study as the result of an impact finding could also lead to bias. We followed WWC standards, which do not allow studies to meet the attrition standard by dropping school districts (or other strata) after the fact.[2] We calculated attrition rates on the basis of the original sample and conducted tests to examine the sensitivity of the estimates to attrition. We see in Table E.1 that excluding mini-studies and districts with high differential attrition does not change the sign or significance of impacts, suggesting that the attrition issues are not affecting our findings.

G. Alternative Definitions of SES Participation

Table E.3 shows TOT impact estimates calculated using a variety of definitions of SES participation. First, we defined participation as being assigned to an SES provider; this differs from the ITT impact estimate in Chapter IV because assignment processes did not always perfectly follow the rank order of students based on prior achievement levels. Second, we

[2] http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf

defined participation as students receiving a positive number of tutoring hours, regardless of subject area: that is, a participant is defined as anyone with total service hours > 0. This reduces the treatment group further because some students who were assigned to services did not actually show up to receive them. Third, we defined participation by subject area. Specifically, when estimating impacts on math test scores, we defined a participant as anyone receiving a positive number of math hours. Similarly, when estimating impacts on reading test scores, we defined a participant as anyone receiving a positive number of reading hours. The TOT impact estimates that use subject-specific definitions of participation (receiving any math hours, receiving any reading hours) require the additional assumption that math hours received have no effect on reading test scores and that reading hours received have no effect on math test scores. Finally, we defined participation as the number of hours received (either total hours, math hours, or reading hours), rather than as a binary variable indicating whether hours were greater than zero. The TOT impact estimates on math and reading test scores remain statistically insignificant regardless of which definition is used.

Table E.3. Sensitivity Analyses: Overall Treatment-on-Treated Impact Estimates of SES on Reading and Math Test Scores on Students Near the Cutoff for Services

Definition of Participation	TOT Impact Estimate (Effect Size Units)	*p*-value
Reading Test Scores		
Assigned to a Provider	-0.09	0.27
Participated in SES (hours > 0) (preferred TOT estimate)	-0.10	0.25
Total Hours Received	-0.00	0.45
Math Hours Received	-0.01	0.59
Reading Hours Received	-0.01	0.36
Participated in Math Services (math hours > 0)	-0.25	0.32
Participated in Reading Services (reading hours > 0)	-0.12	0.26
Math Test Scores		
Assigned to a Provider	0.10	0.21
Participated in SES (hours > 0) (preferred TOT estimate)	0.11	0.21
Total Hours Received	0.01	0.13
Math Hours Received	0.02	0.15
Reading Hours Received	0.01	0.18
Participated in Math Services (math hours > 0)	0.21	0.37
Participated in Reading Services (reading hours > 0)	0.13	0.25

Source: Math and reading test scores from school district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cutoff-value/grade combination. For each mini-study, the TOT impact estimate is calculated as the ratio of the ITT impact estimate to the impact estimate on participation. Aggregate impact estimates are calculated as a weighted average of the impact estimates for each mini-study, in which the weight is equal to the inverse of the variance of that mini-study's TOT impact estimate.

*Significantly different from zero at the 0.05 level, two-tailed test.

ITT = intent to treat; SES = supplemental education services; TOT = treatment on treated.

APPENDIX F

IMPACT ESTIMATES AND GRAPHICAL ANALYSES BY MINI-STUDY

This appendix presents impact estimates and graphical analyses by mini-study. Impact estimates are presented in Table F.1. Scatter plots of outcomes versus assignment variables are shown in Figures F.1a–F.21b. Each figure includes a scatter plot, a smoothed curve, a solid vertical line indicating the cutoff value of the assignment variable, and two dashed vertical lines indicating the IK bandwidth. The smoothed curves are estimated using smoothing splines estimated in R with the package MGCV, which used generalized cross-validation to select smoothing parameters (Wood 2004). Two smoothed curves are estimated independently in each figure—one for points above the cutoff value, and one for points below the cutoff value. Note that actual impacts are estimated using a linear functional form within the IK bandwidth, not the smoothed curves shown in these figures. Because the IK bandwidth is intended to include only observations for which a linear functional form is a reasonable choice, we expect the smoothed curve within the vertical dashed lines to be approximately linear.

Table F.1. Intent to Treat (ITT) Impact Estimates of SES on Reading and Math Test Scores on Students Near the Cutoff for Services, by Mini-Study

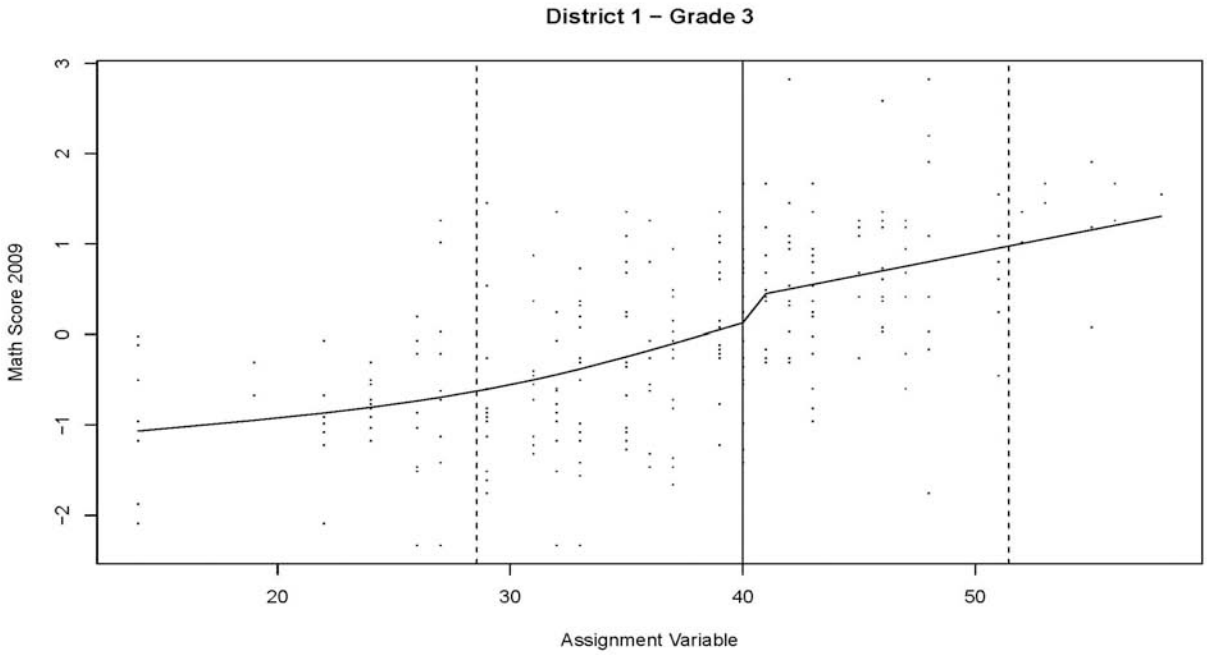
Mini-Study Description	Adjusted Mean Outcomes		ITT Impact Estimate (Effect Size Units)	*p*-Value
	Treatment	Comparison		
District 1 / grade 3 / math test scores	-0.92	-0.77	-0.15	0.43
District 1 / grade 3 / reading test scores	-0.63	-0.79	0.16	0.35
District 1 / grade 4 / math test scores	-0.69	-0.51	-0.18	0.81
District 1 / grade 4 / reading test scores	-0.64	-0.71	0.07	0.91
District 1 / grade 5 / math test scores	-0.82	-0.75	-0.07	0.93
District 1 / grade 5 / reading test scores	-0.98	-0.90	-0.07	0.90
District 1 / grade 6 / math test scores	-0.51	-0.52	0.01	0.99
District 1 / grade 6 / reading test scores	-0.78	-0.63	-0.15	0.78
District 1 / grade 7 / math test scores	-0.53	-0.72	0.19	0.72
District 1 / grade 7 / reading test scores	-0.70	-0.35	-0.36	0.50
District 1 / grade 8 / math test scores	-0.97	-0.94	-0.03	0.95
District 1 / grade 8 / reading test scores	-0.92	-0.95	0.03	0.93
District 2 / grade 3 / math test scores	-2.59	-1.89	-0.71	0.53
District 2 / grade 3 / reading test scores	-1.04	-1.73	0.69	0.92
District 2 / grade 4 / math test scores	-0.57	-0.98	0.41	0.24
District 2 / grade 4 / reading test scores	-0.72	-0.99	0.27	0.34
District 3 / grade 4 / math test scores	0.41	0.33	0.08	0.91
District 3 / grade 4 / reading test scores	0.86	0.52	0.34	0.51
District 3 / grade 5 / math test scores	0.20	0.19	0.01	0.94
District 3 / grade 5 / reading test scores	0.30	0.13	0.18	0.85
District 3 / grade 6 / math test scores	0.15	-0.77	0.91	0.10
District 3 / grade 6 / reading test scores	0.05	-0.03	0.08	0.89
District 4 / grade 3 / math test scores	0.45	0.48	-0.03	0.89
District 4 / grade 3 / reading test scores	0.48	0.50	-0.02	0.90
District 4 / grade 4 / math test scores	0.62	0.77	-0.15	0.36
District 4 / grade 4 / reading test scores	0.57	0.91	-0.34*	0.04
District 4 / grade 5 / math test scores	0.56	0.55	0.01	0.88
District 4 / grade 5 / reading test scores	0.57	0.59	-0.03	0.82
District 4 / grade 6 / math test scores	0.38	0.30	0.08	0.35
District 4 / grade 6 / reading test scores	0.36	0.40	-0.04	0.67
District 4 / grade 7 / math test scores	0.18	0.47	-0.29	0.08
District 4 / grade 7 / reading test scores	0.36	0.50	-0.14	0.31
District 4 / grade 8 / math test scores	0.57	0.31	0.25*	0.05
District 4 / grade 8 / reading test scores	0.50	0.36	0.14	0.33
District 5 / grade 3 / math test scores	-0.33	-0.17	-0.17	0.61
District 5 / grade 3 / reading test scores	-0.45	-0.28	-0.17	0.65
District 6 / grade 3 / math test scores	-0.22	-0.50	0.27	0.08
District 6 / grade 3 / reading test scores	-0.15	-0.17	0.02	0.83
District 6 / grade 4 / math test scores	-0.14	-0.13	-0.01	0.96
District 6 / grade 4 / reading test scores	-0.00	0.01	-0.01	0.97
District 6 / grade 5 / math test scores	-0.08	-0.17	0.09	0.47
District 6 / grade 5 / reading test scores	-0.06	0.23	-0.30	0.20

Source: Math and reading test scores from school district records.

Note: This evaluation consists of 42 mini-studies, each of which corresponds to a separate outcome/cutoff-value/grade combination. For each mini-study, the intent to treat (ITT) impact estimate is calculated using a linear functional form within an optimal bandwidth selected using the Imbens & Kalyanaraman (2009) algorithm and standard errors are adjusted to account for clustering of observations within unique values of the assignment variable as in Lee & Card (2008). All outcomes are standardized to have a standard deviation of 1, so impact estimates are reported in effect size units. Adjusted mean outcomes for the treatment and comparison groups are equal to the estimated intercept terms from the regressions of the outcome (student test scores) on the assignment variable and other covariates that were estimated separately on either side of the cutoff value. The outcomes were standardized using state-grade-level means and standard deviations.

*Significantly different from zero at the 0.05 level, two-tailed test.

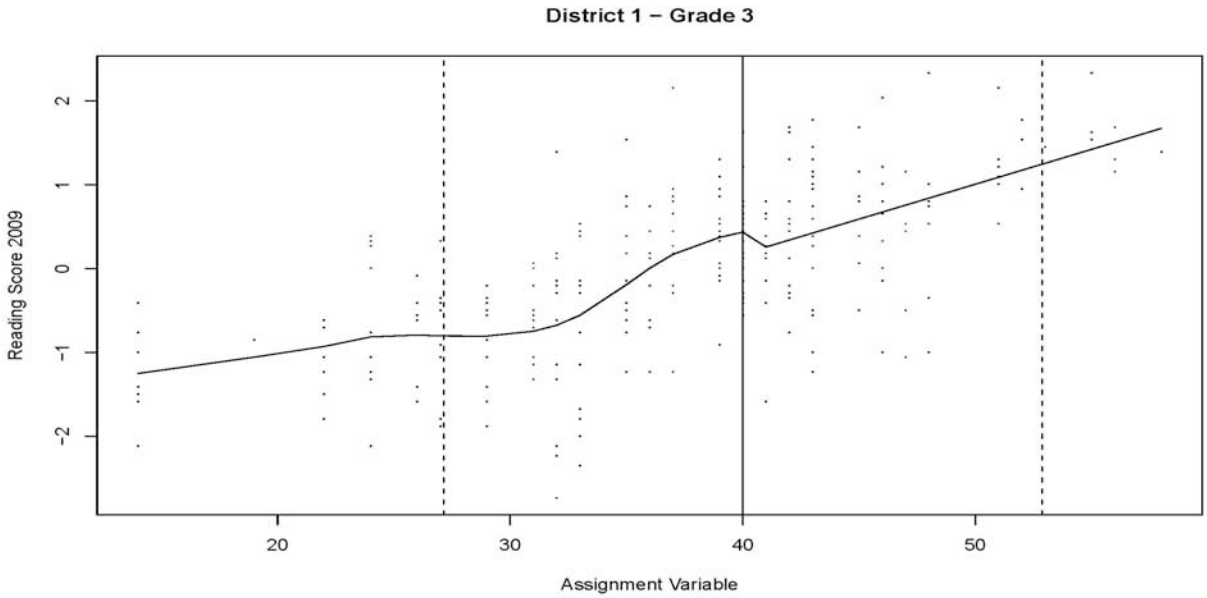
Figure F.1.a. Math Score versus Assignment Variable for Mini-Study 1



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

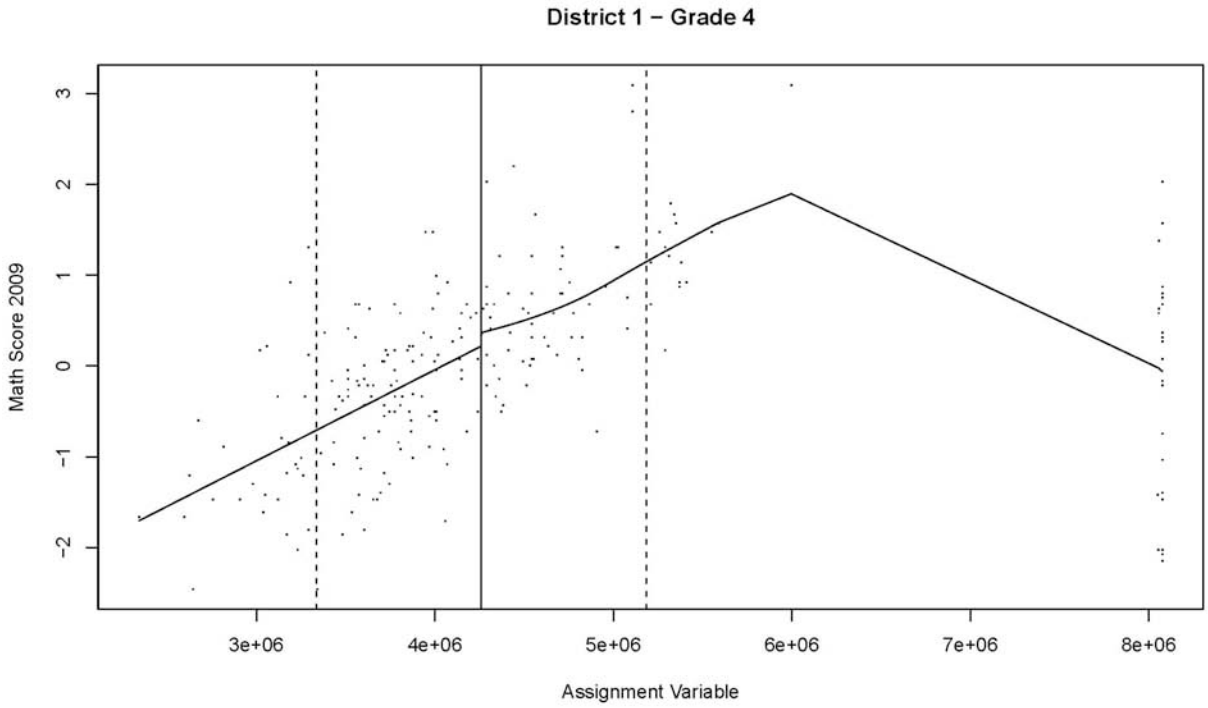
Figure F.1.b. Reading Score versus Assignment Variable for Mini-Study 1



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

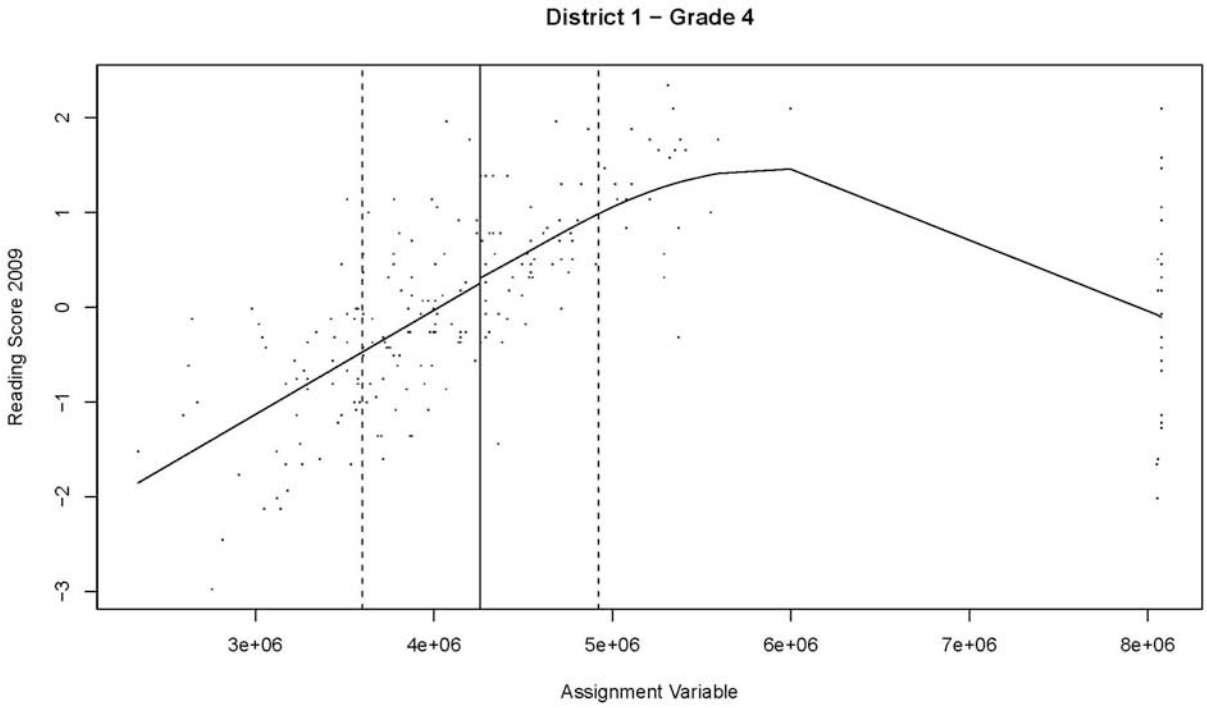
Figure F.2a. Math Score versus Assignment Variable for Mini-Study 2



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

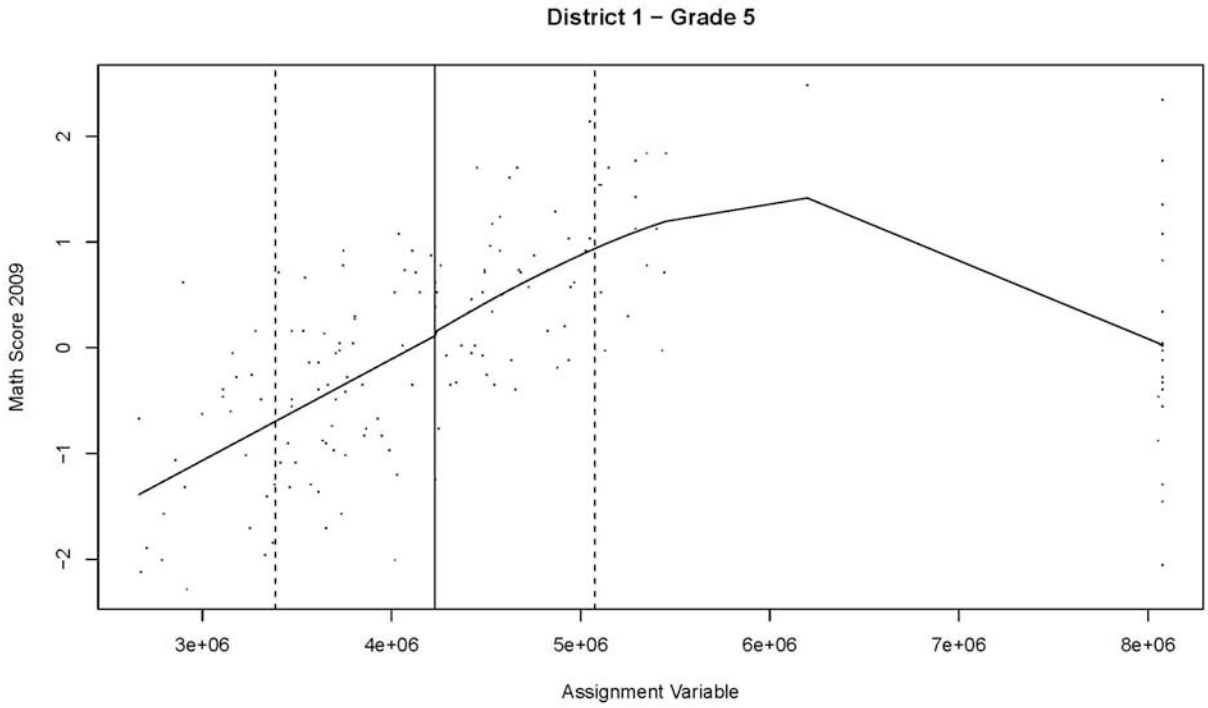
Figure F.2b. Reading Score versus Assignment Variable for Mini-Study 2



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

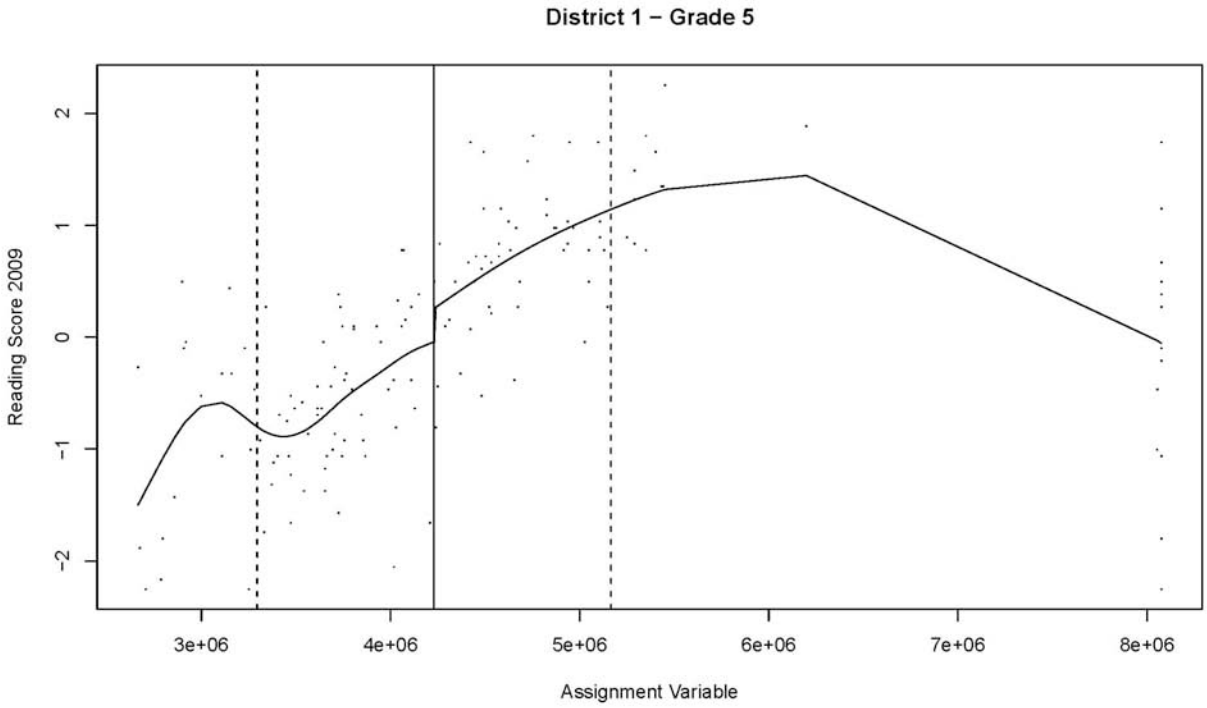
Figure F.3a. Math Score versus Assignment Variable for Mini-Study 3



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

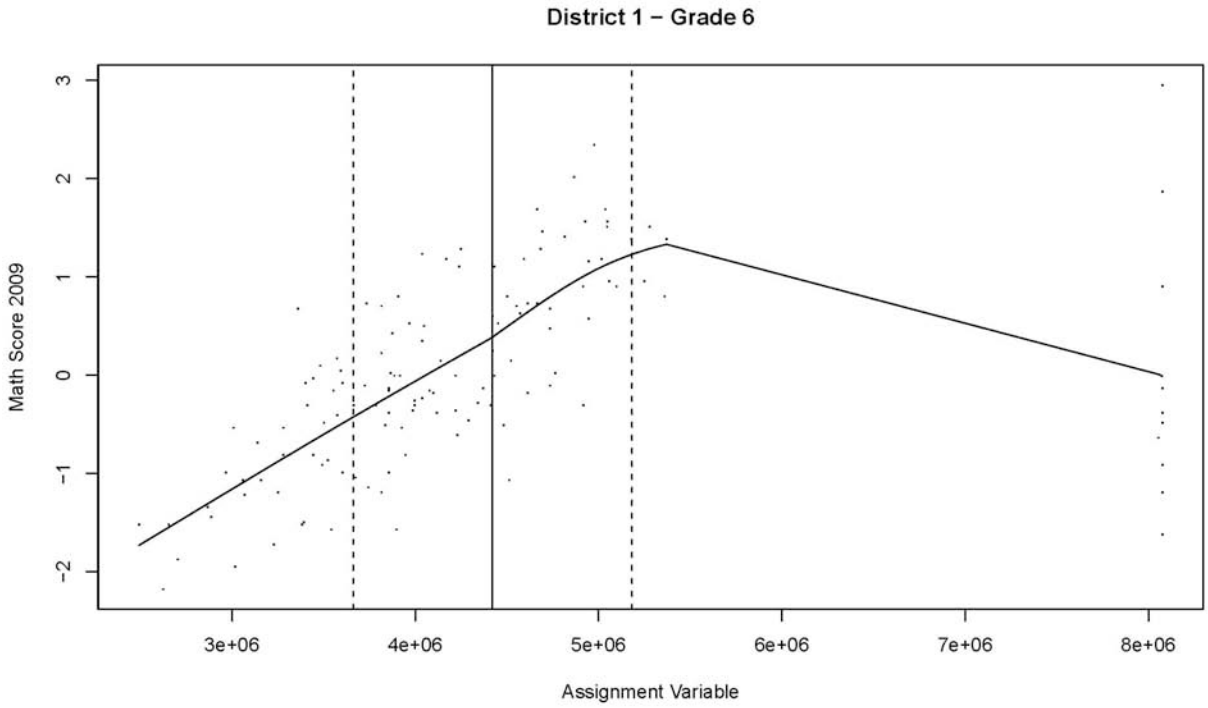
Figure F.3b. Reading Score versus Assignment Variable for Mini-Study 3



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

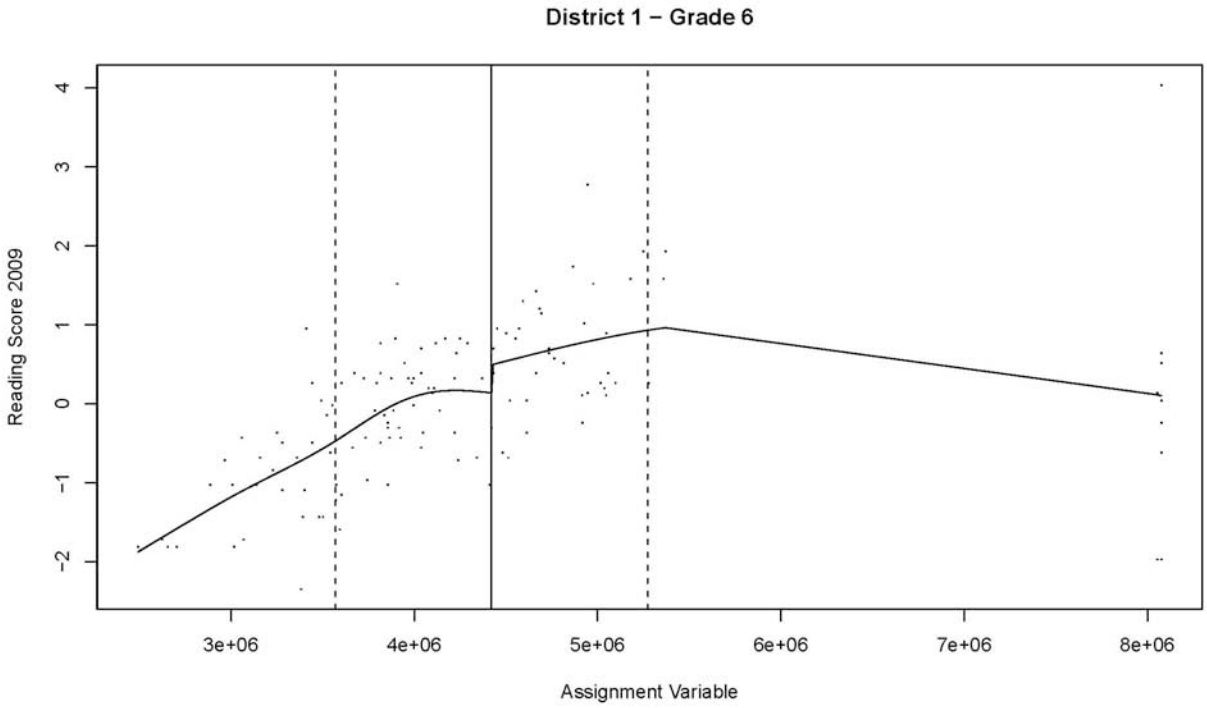
Figure F.4a. Math Score versus Assignment Variable for Mini-Study 4



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

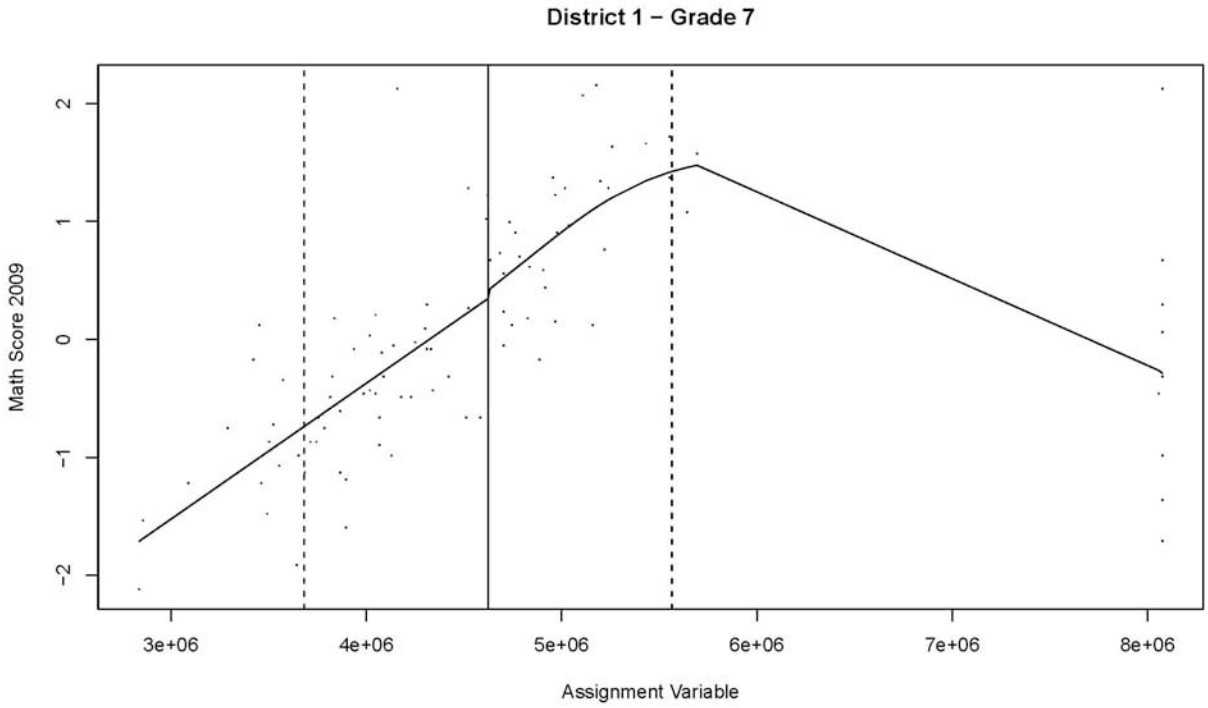
Figure F.4b. Reading Score versus Assignment Variable for Mini-Study 4



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

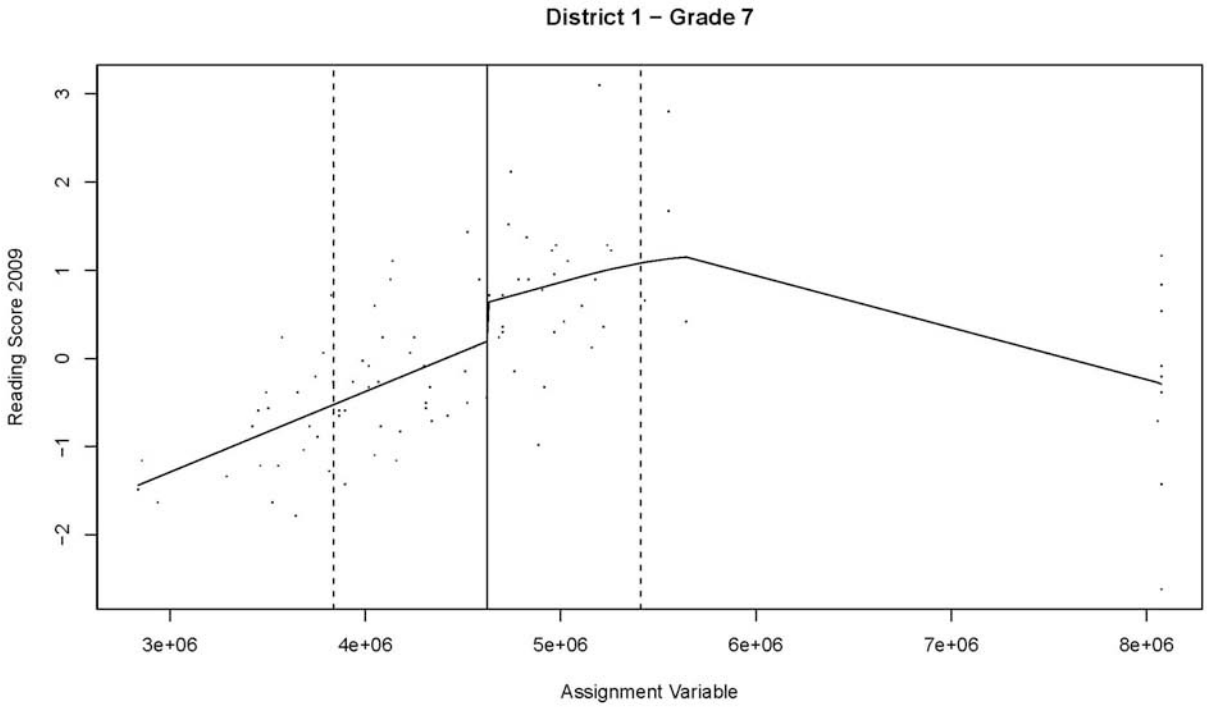
Figure F.5a. Math Score versus Assignment Variable for Mini-Study 5



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

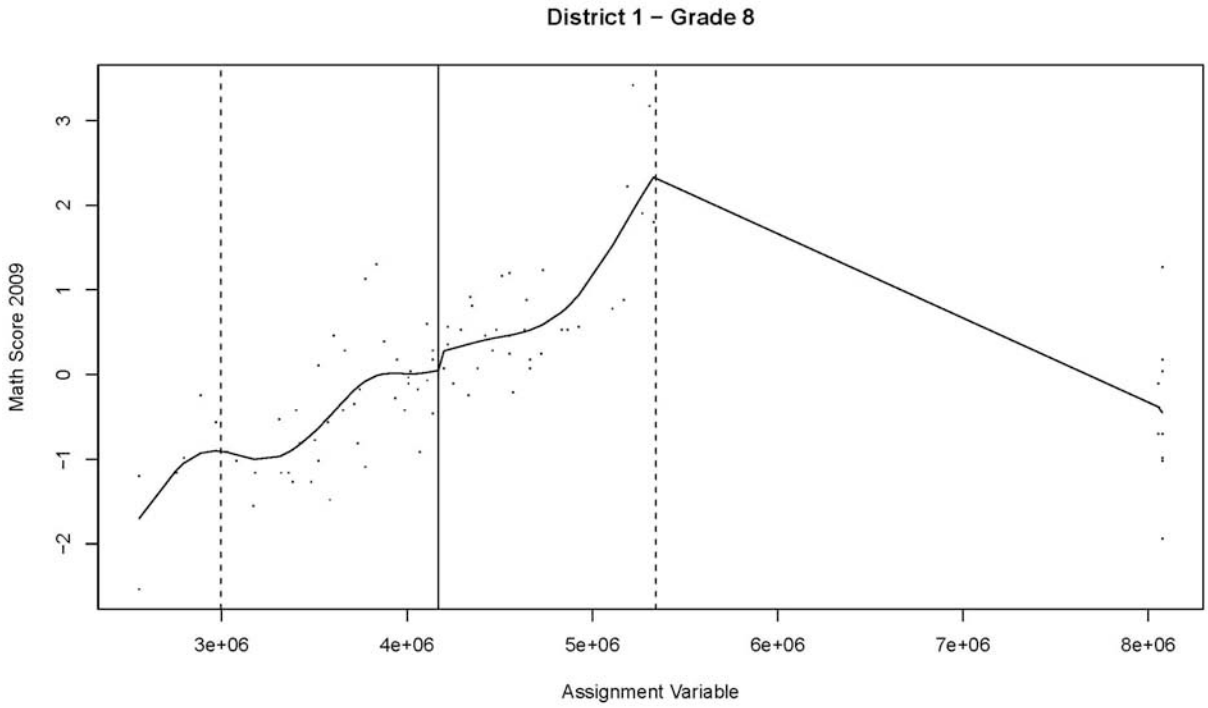
Figure F.5b. Reading Score versus Assignment Variable for Mini-Study 5



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

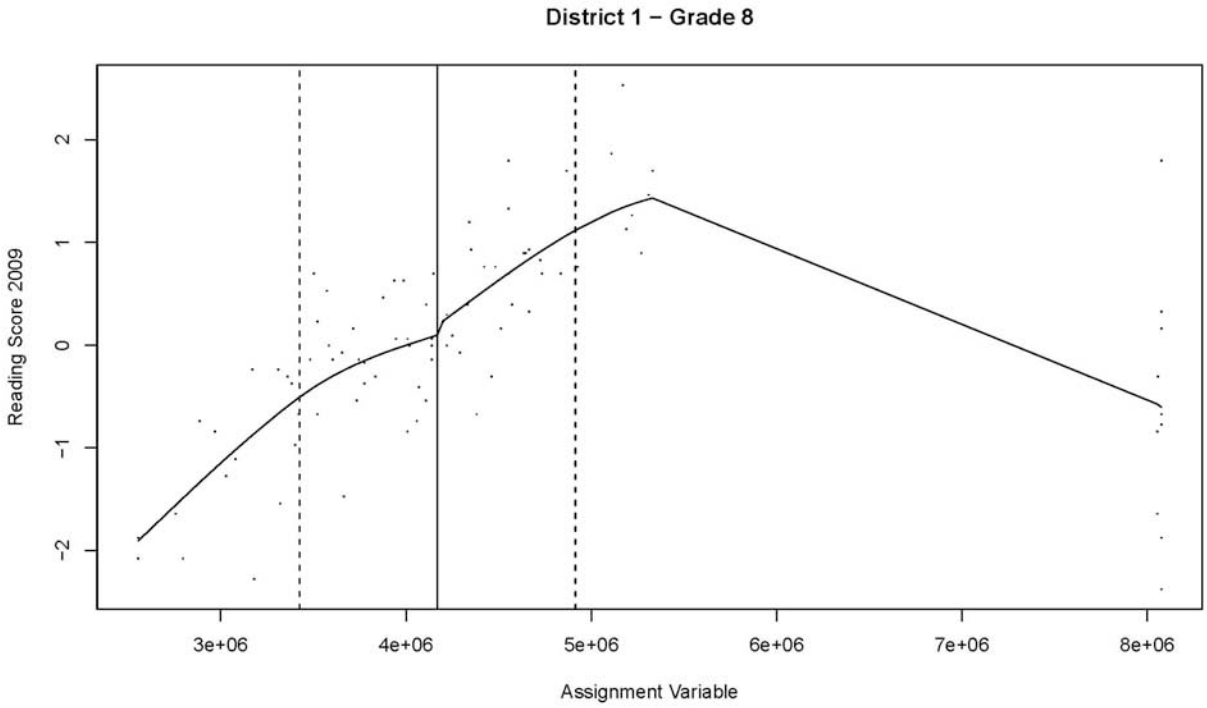
Figure F.6a. Math Score versus Assignment Variable for Mini-Study 6



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

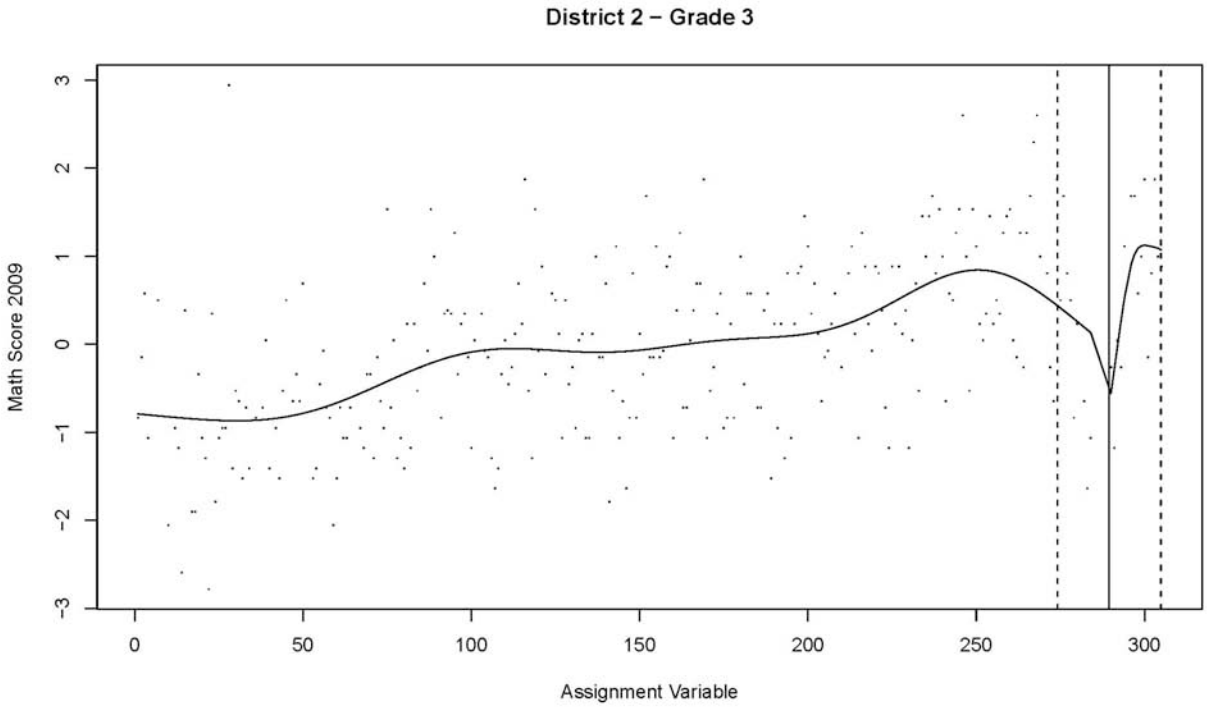
Figure F.6b. Reading Score versus Assignment Variable for Mini-Study 6



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.3.

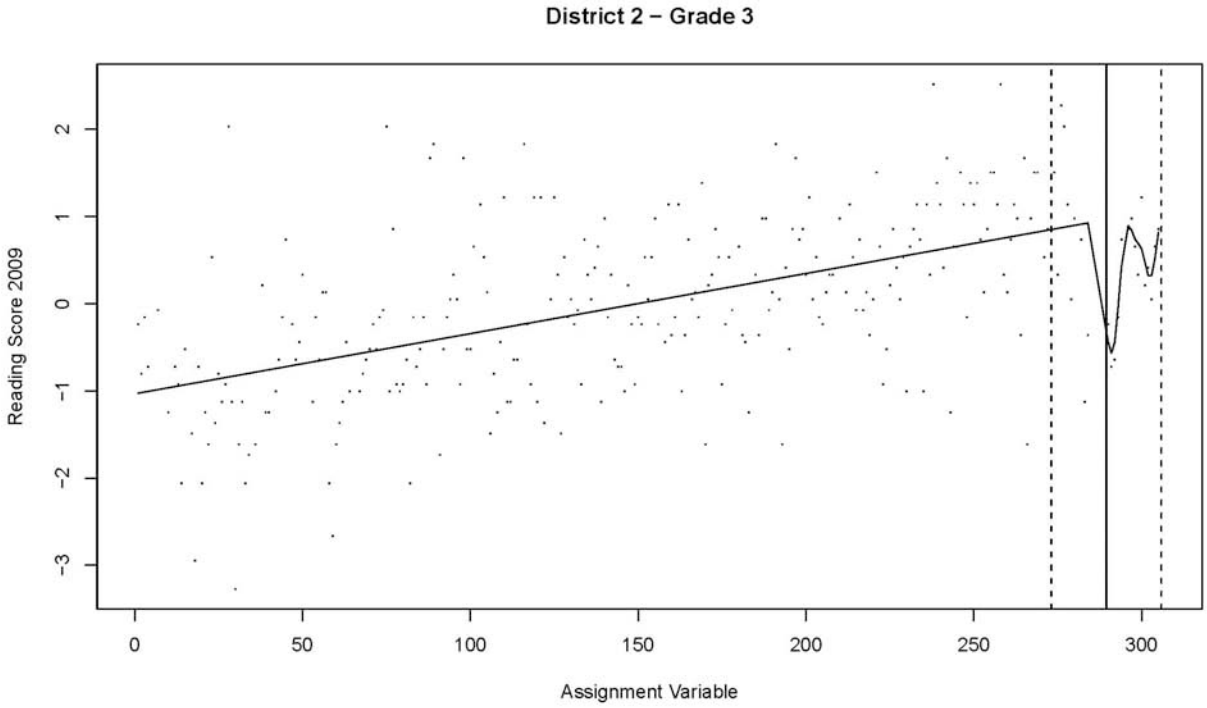
Figure F.7a. Math Score versus Assignment Variable for Mini-Study 7



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on pages A.3-A.4.

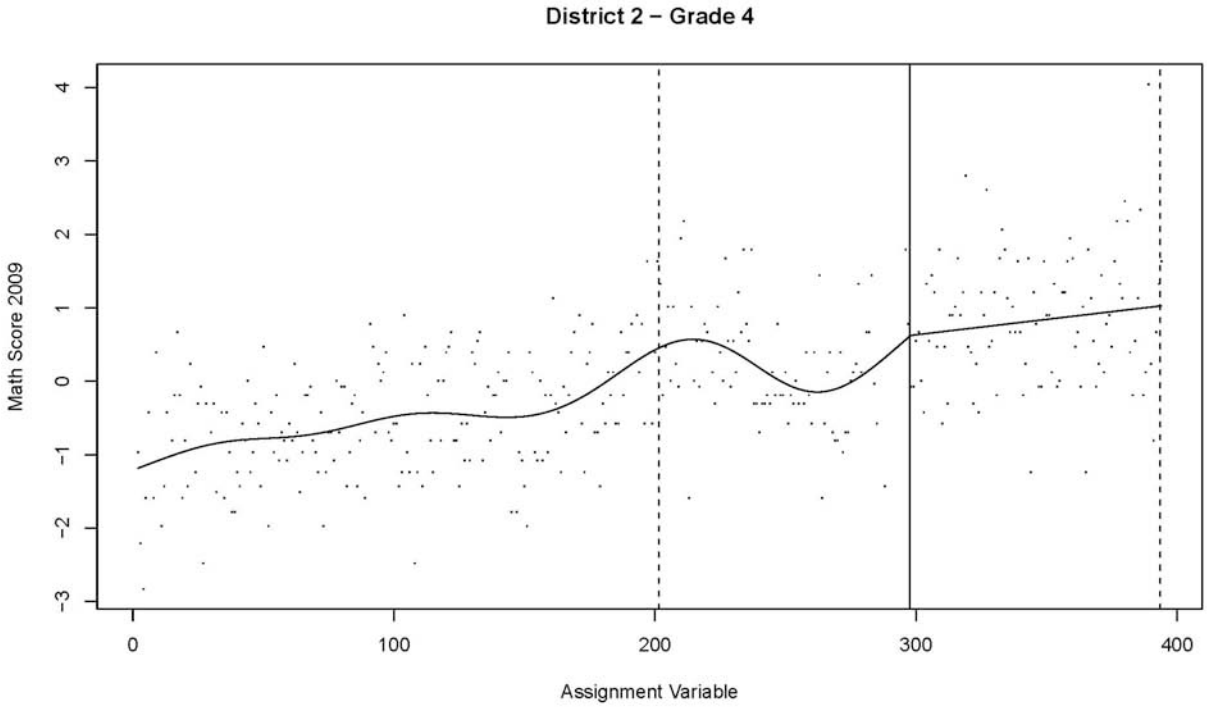
Figure F.7b. Reading Score versus Assignment Variable for Mini-Study 7



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on pages A.3-A.4.

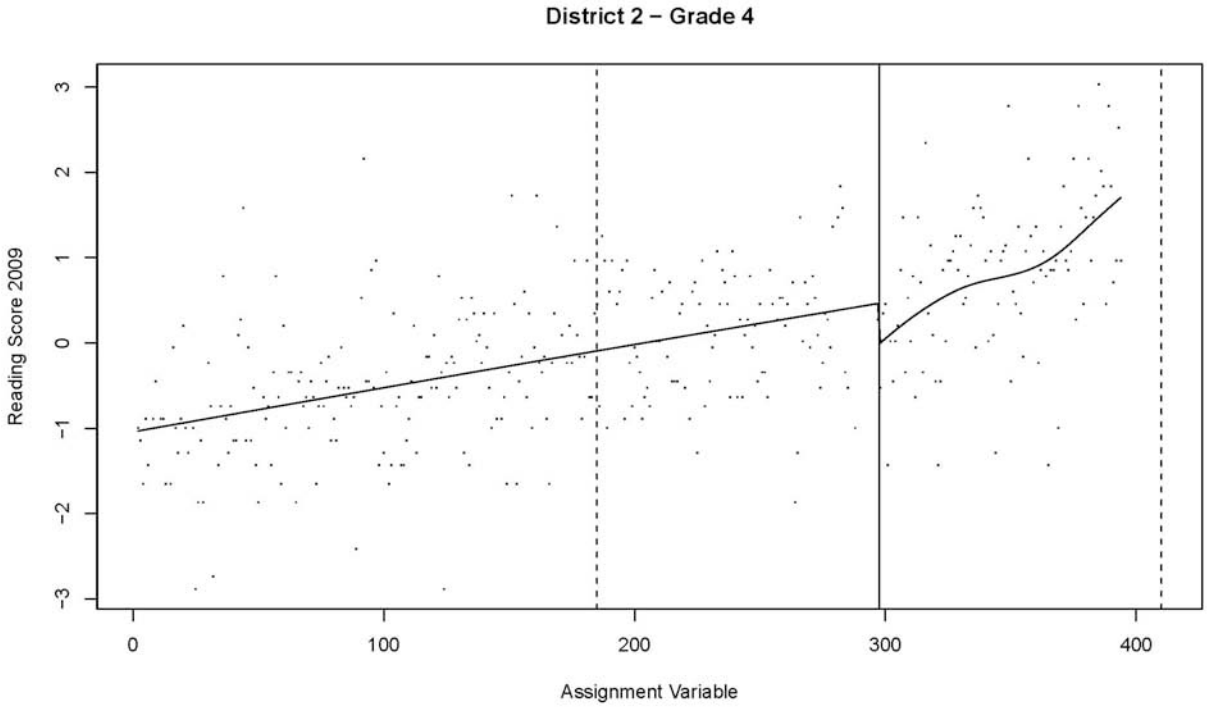
Figure F.8a. Math Score versus Assignment Variable for Mini-Study 8



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on pages A.3-A.4.

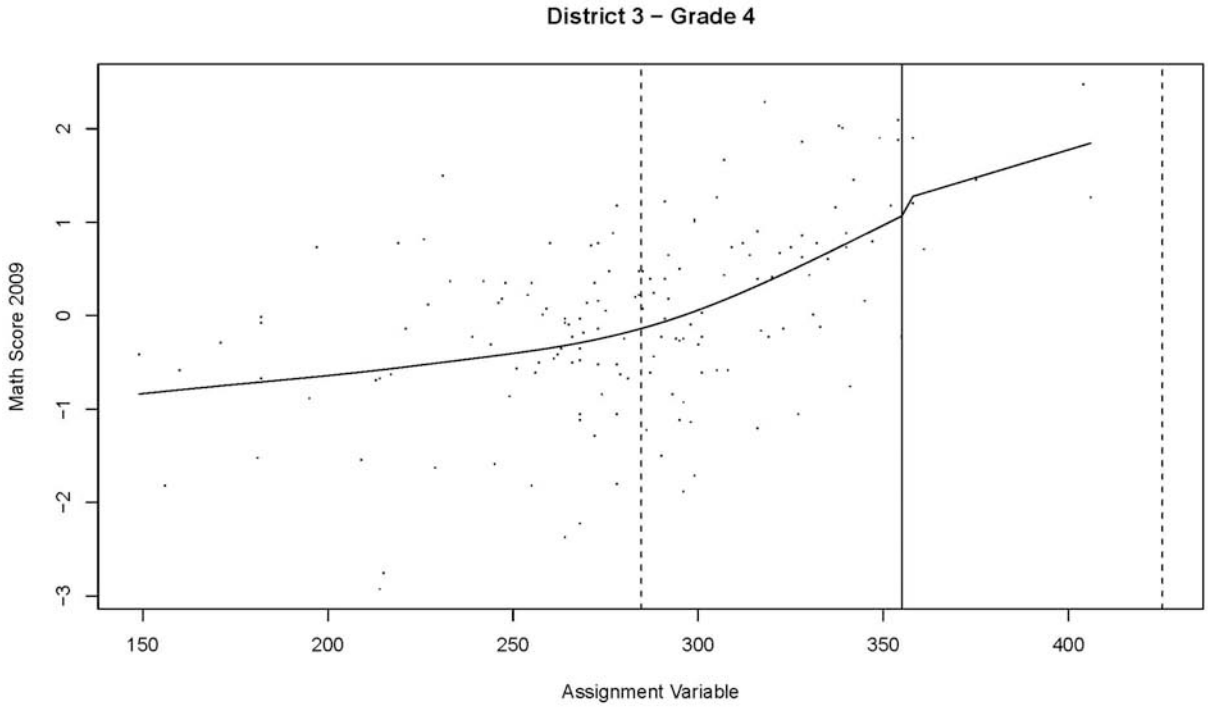
Figure F.8b. Reading Score versus Assignment Variable for Mini-Study 8



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on pages A.3-A.4.

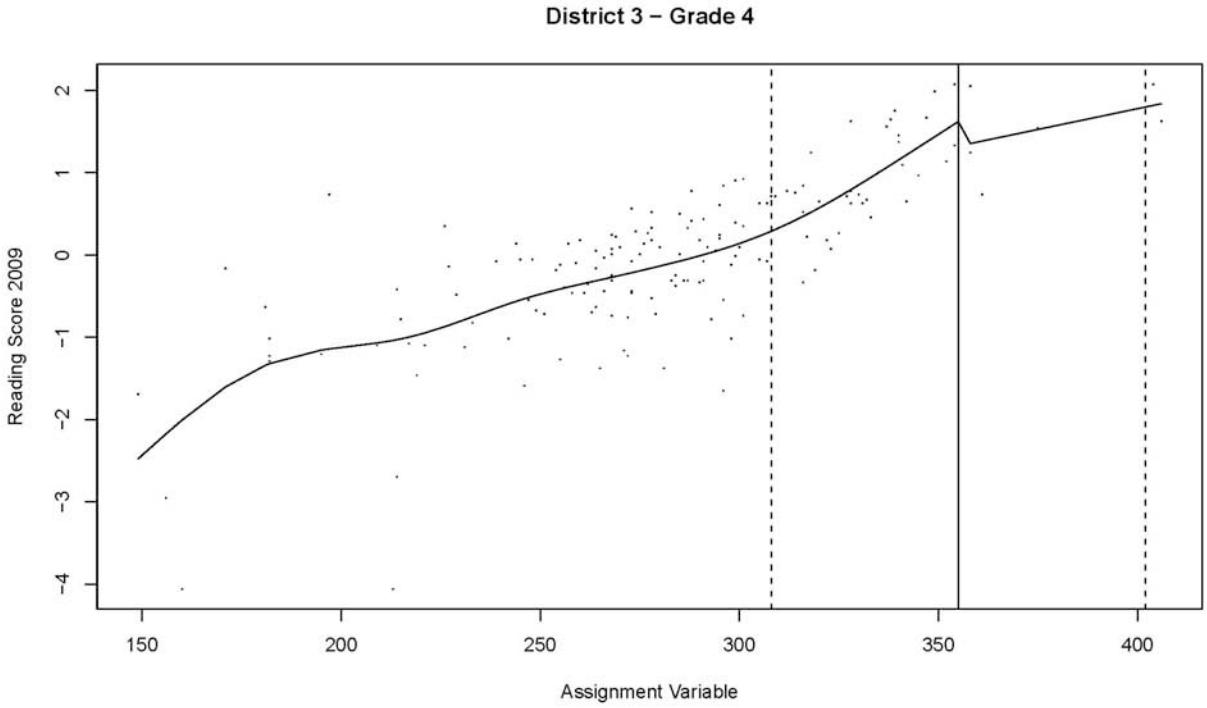
Figure F.9a. Math Score versus Assignment Variable for Mini-Study 9



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

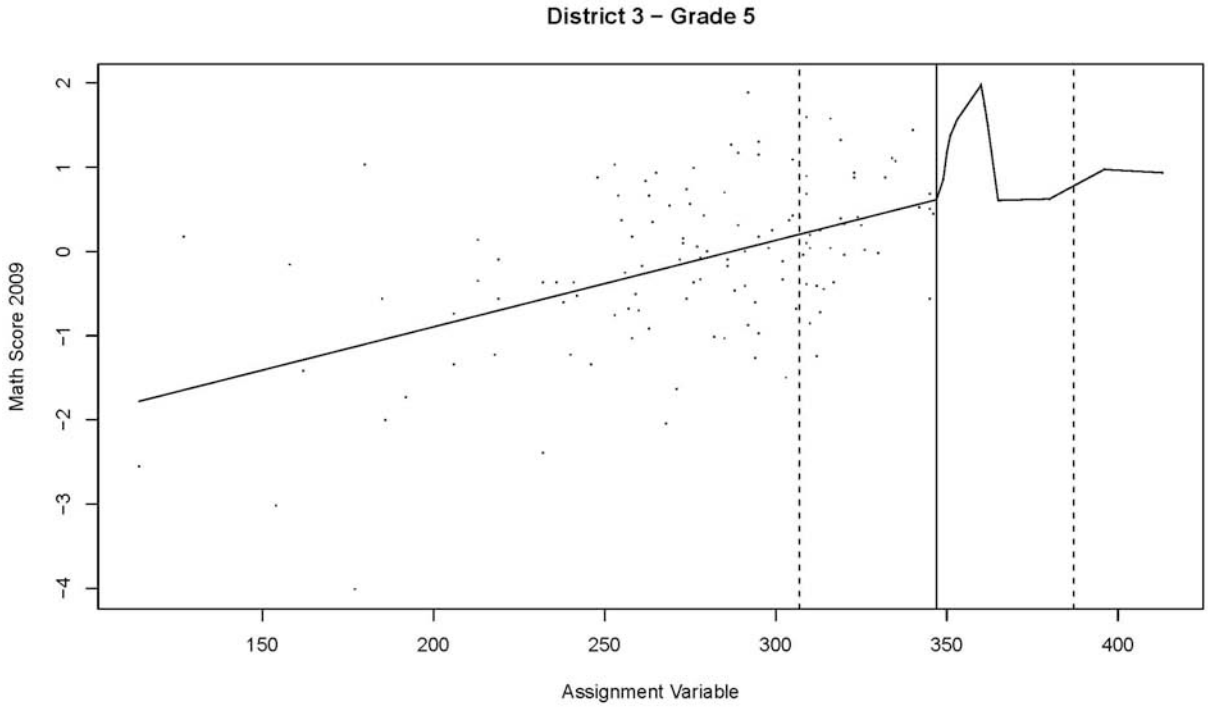
Figure F.9b. Reading Score versus Assignment Variable for Mini-Study 9



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

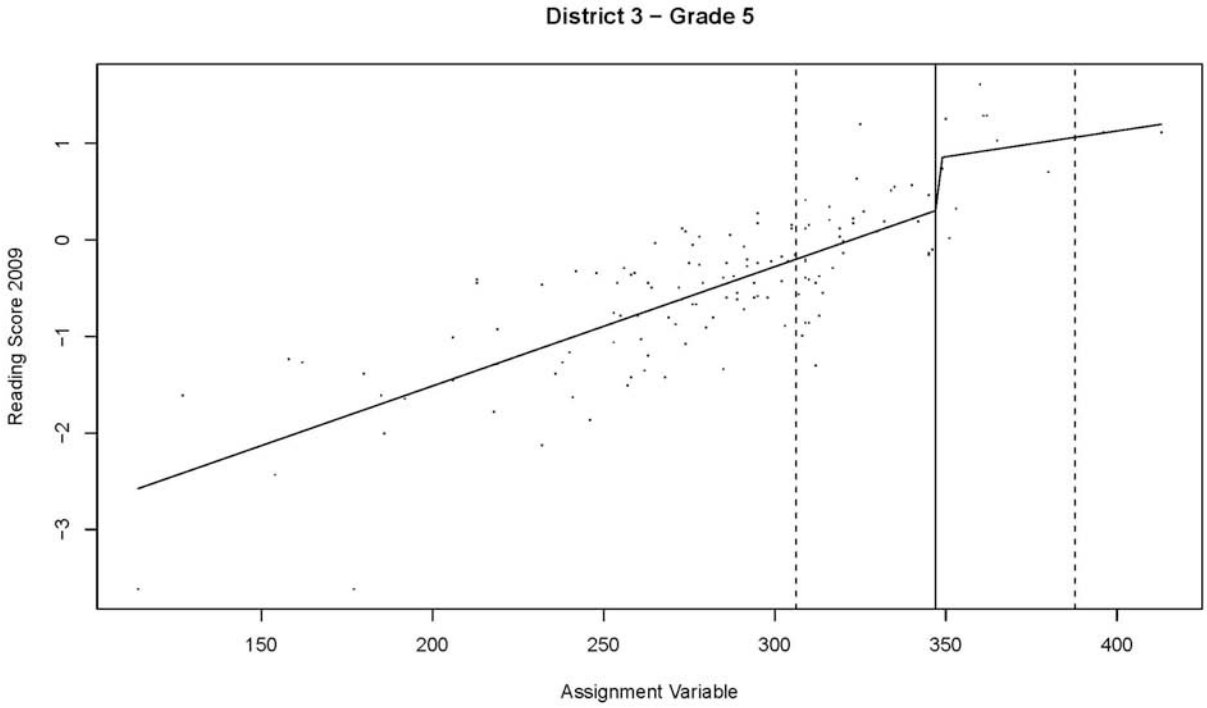
Figure F.10a. Math Score versus Assignment Variable for Mini-Study 10



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

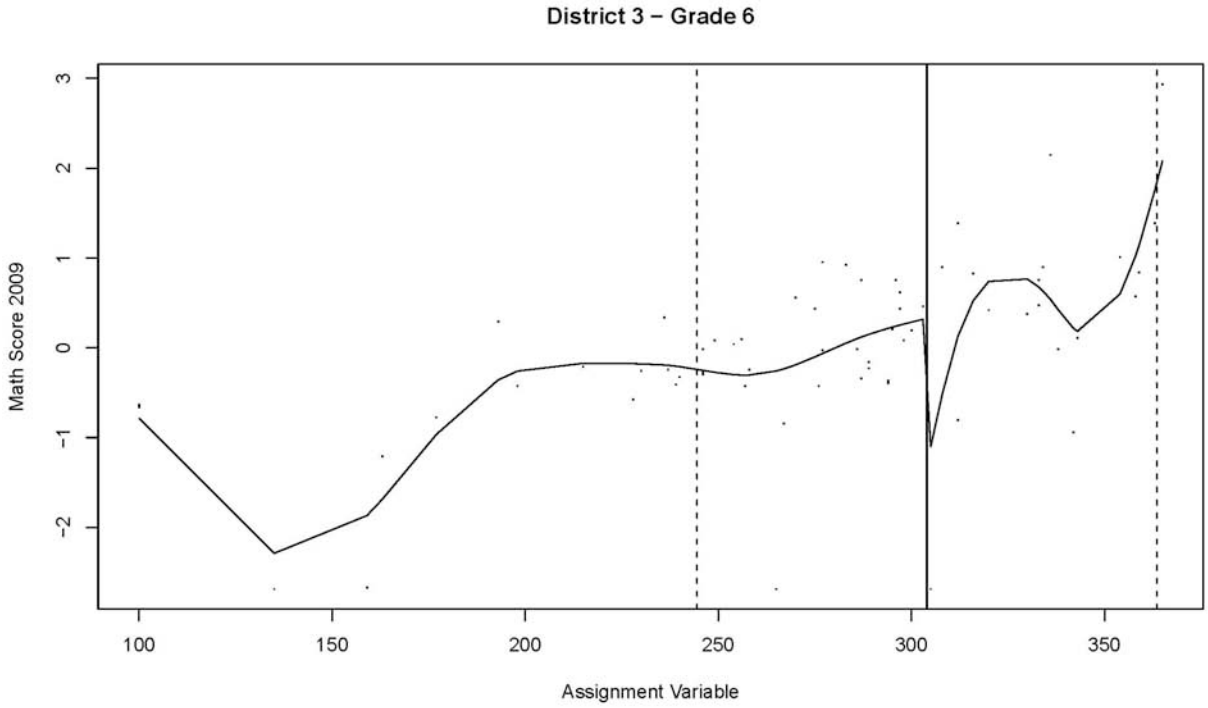
Figure F.10b. Reading Score versus Assignment Variable for Mini-Study 10



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

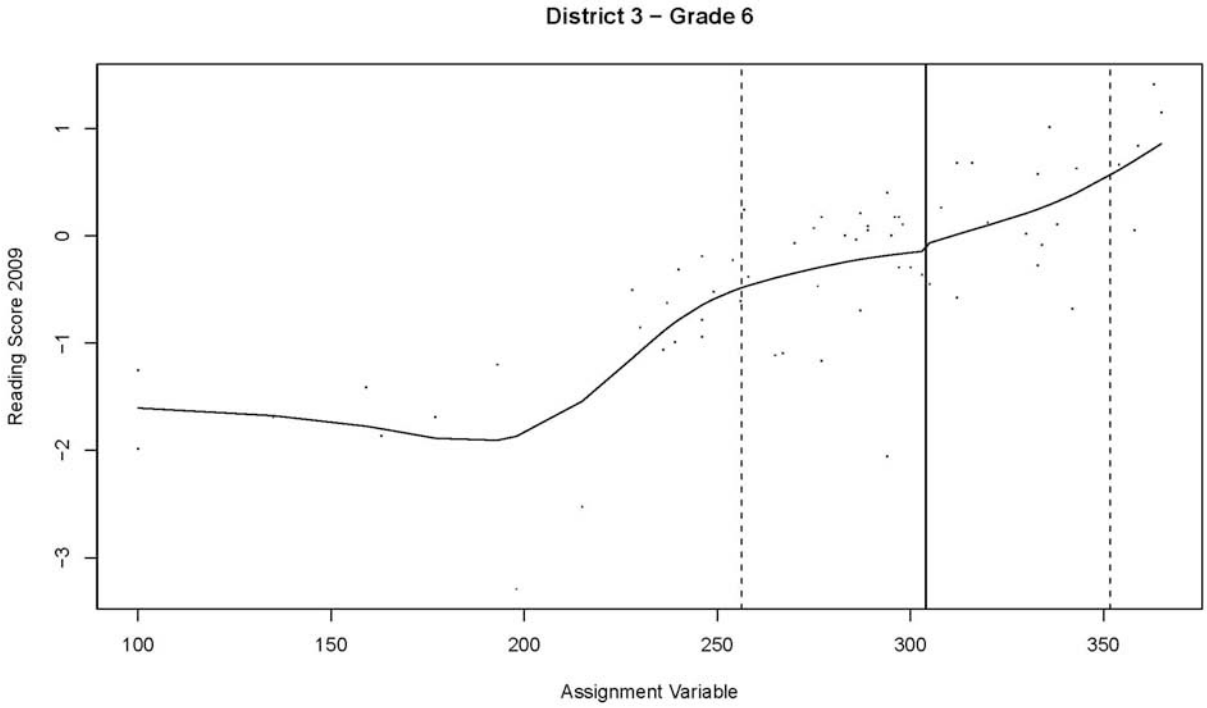
Figure F.11a. Math Score versus Assignment Variable for Mini-Study 11



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

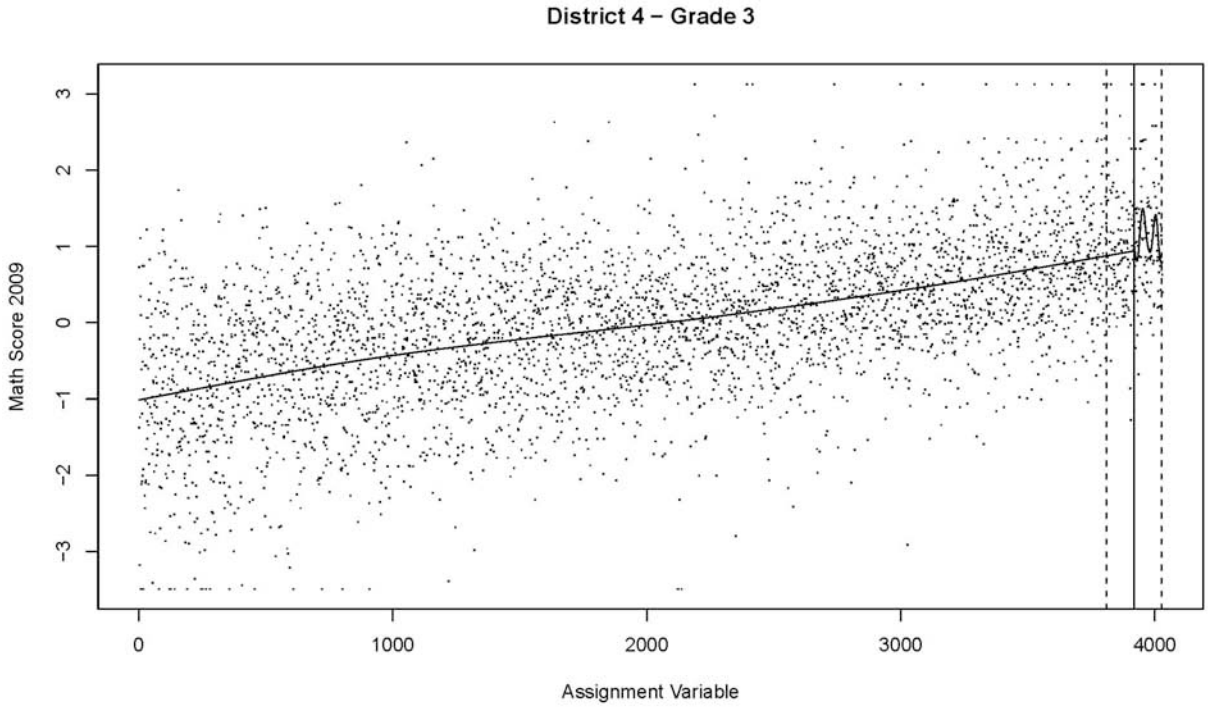
Figure F.11b. Reading Score versus Assignment Variable for Mini-Study 11



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

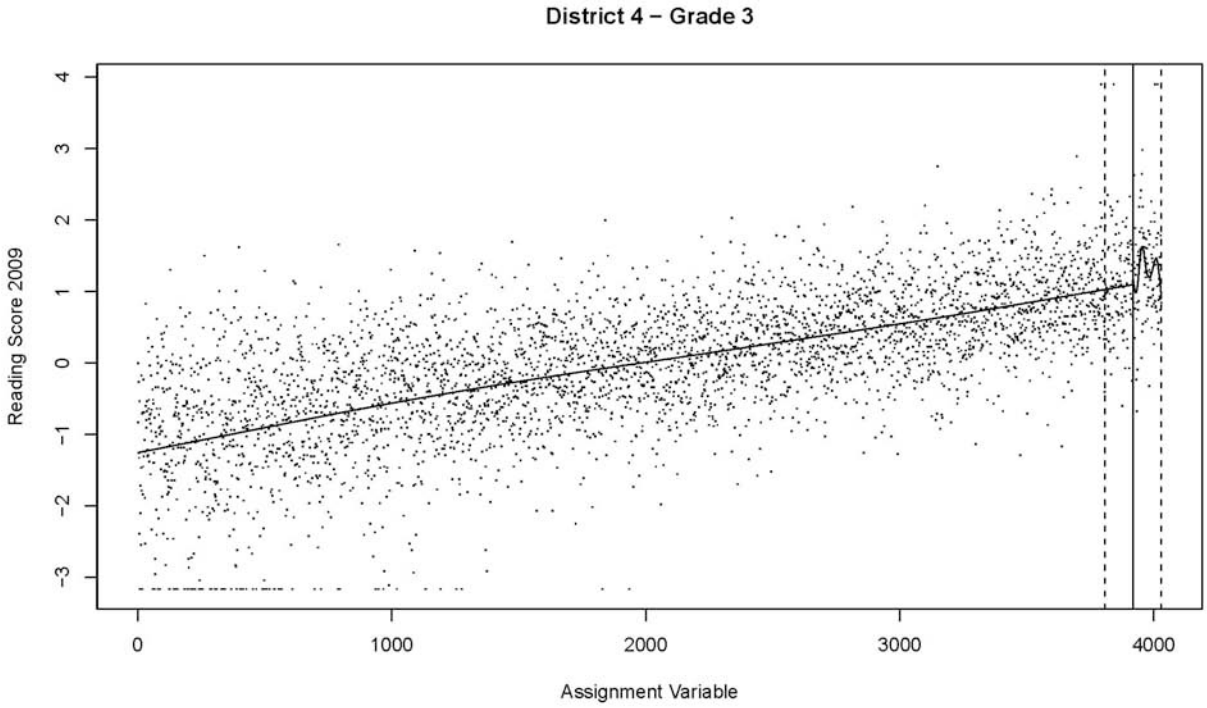
Figure F.12a. Math Score versus Assignment Variable for Mini-Study 12



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

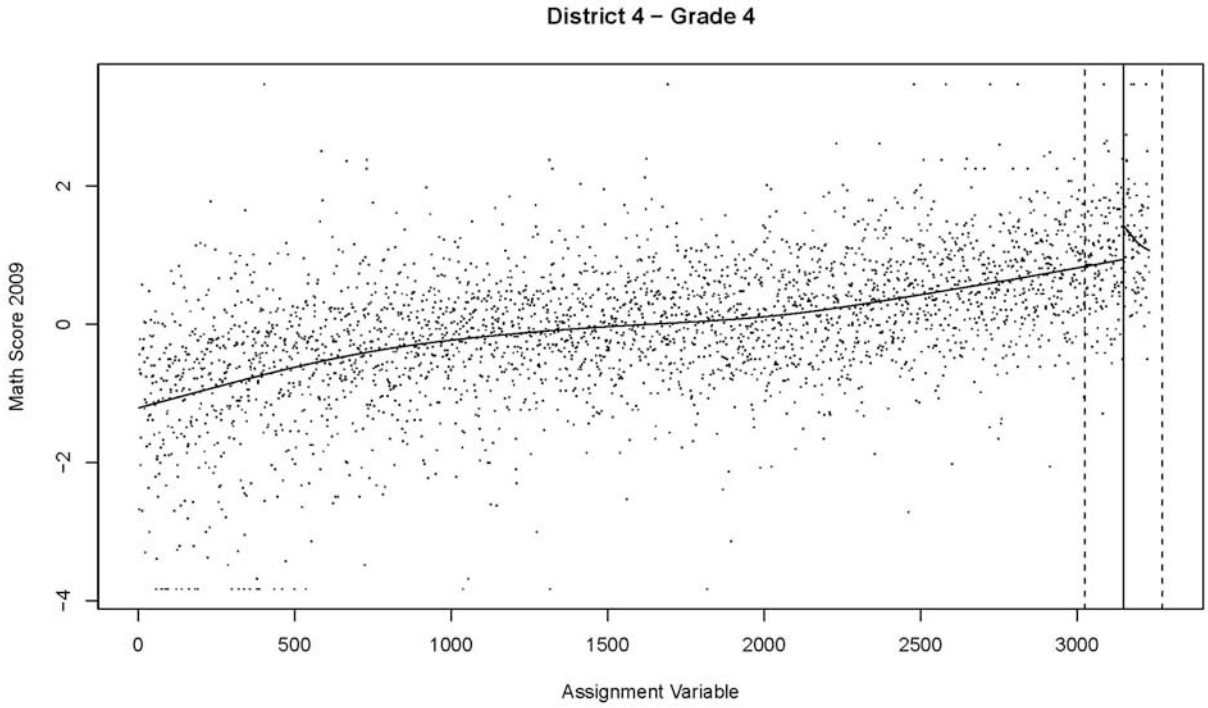
Figure F.12b. Reading Score versus Assignment Variable for Mini-Study 12



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

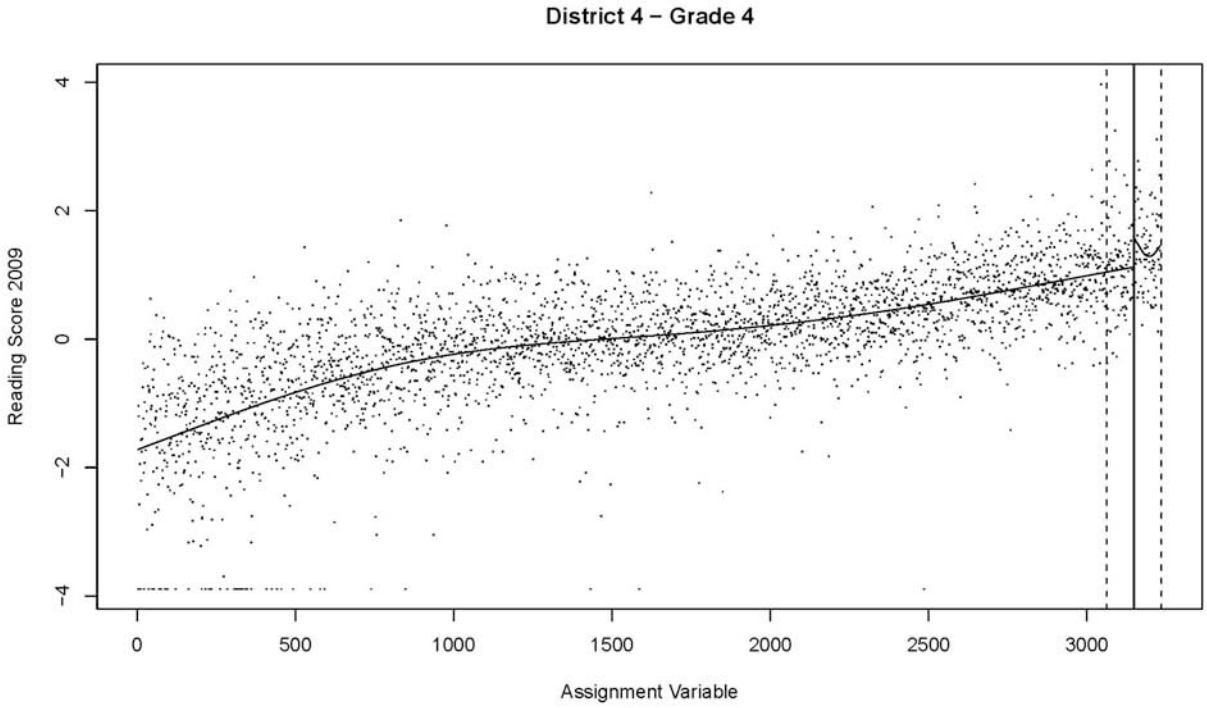
Figure F.13a. Math Score versus Assignment Variable for Mini-Study 13



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

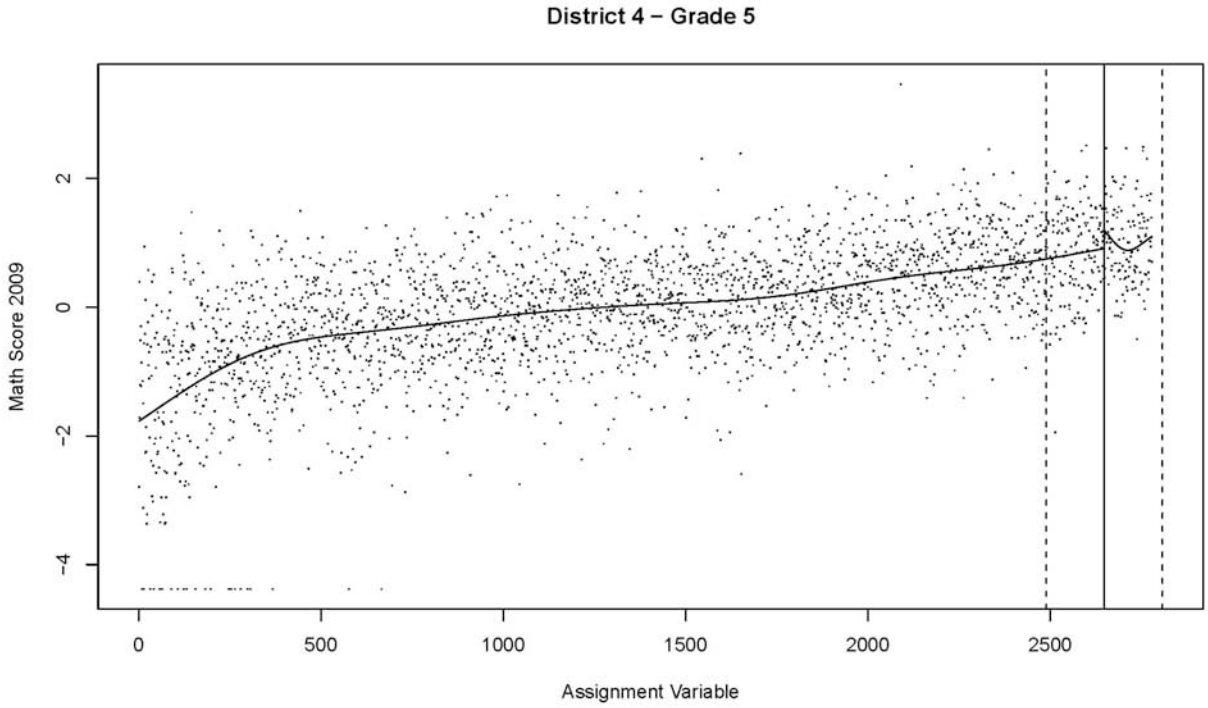
Figure F.13b. Reading Score versus Assignment Variable for Mini-Study 13



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

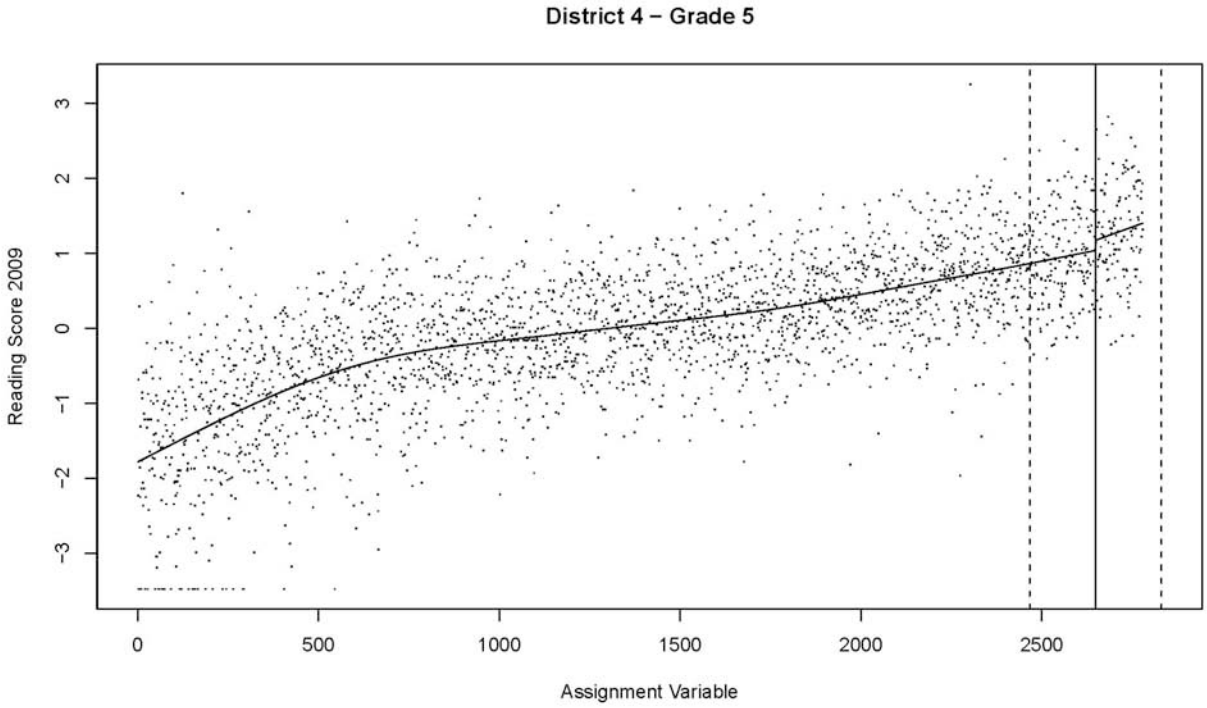
Figure F.14a. Math Score versus Assignment Variable for Mini-Study 14



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

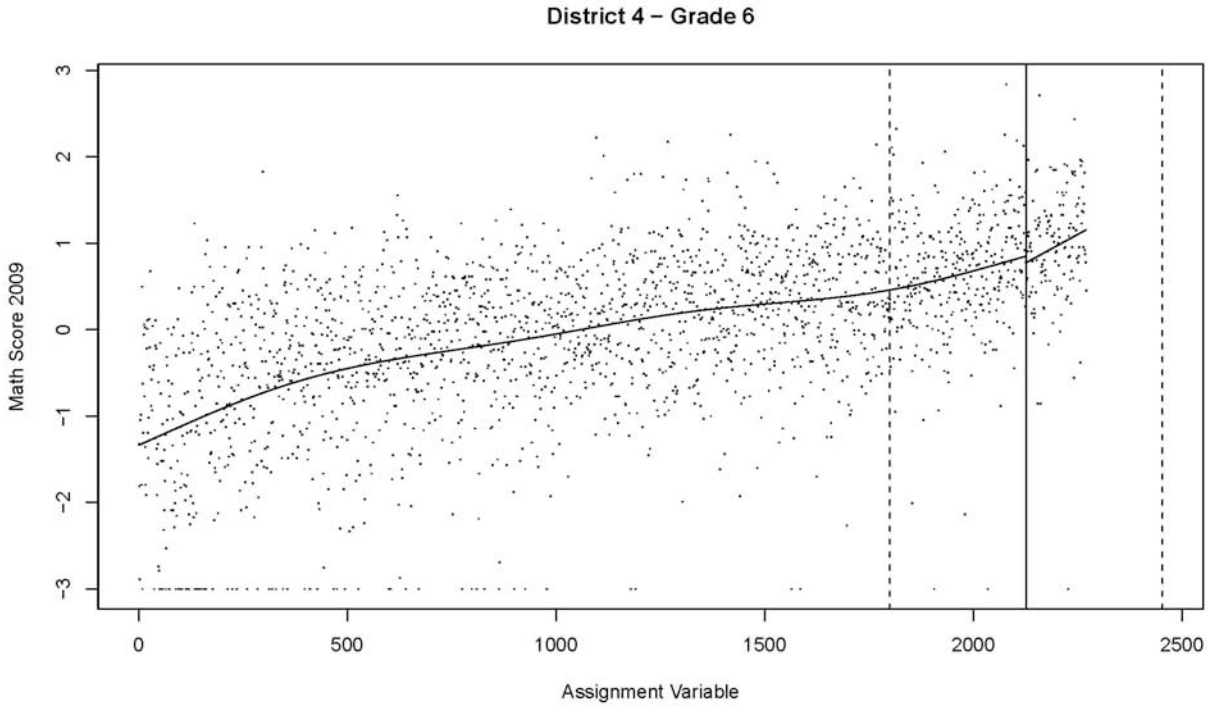
Figure F.14b. Reading Score versus Assignment Variable for Mini-Study 14



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

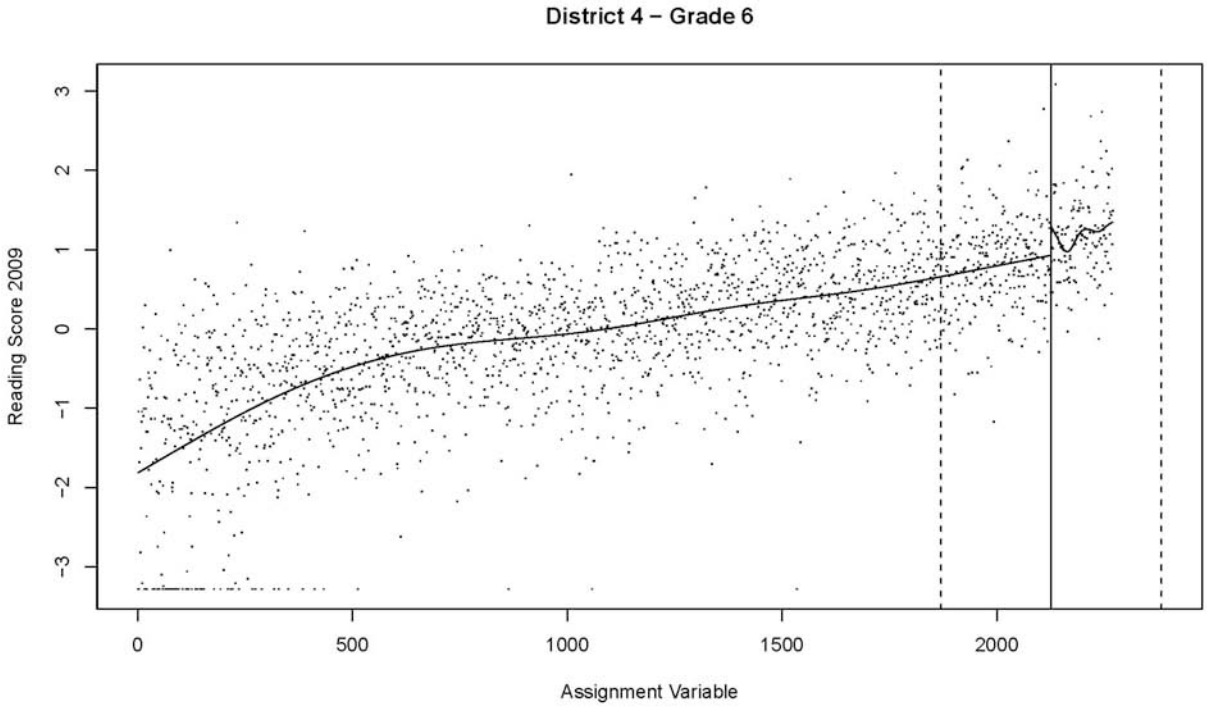
Figure F.15a. Math Score versus Assignment Variable for Mini-Study 15



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

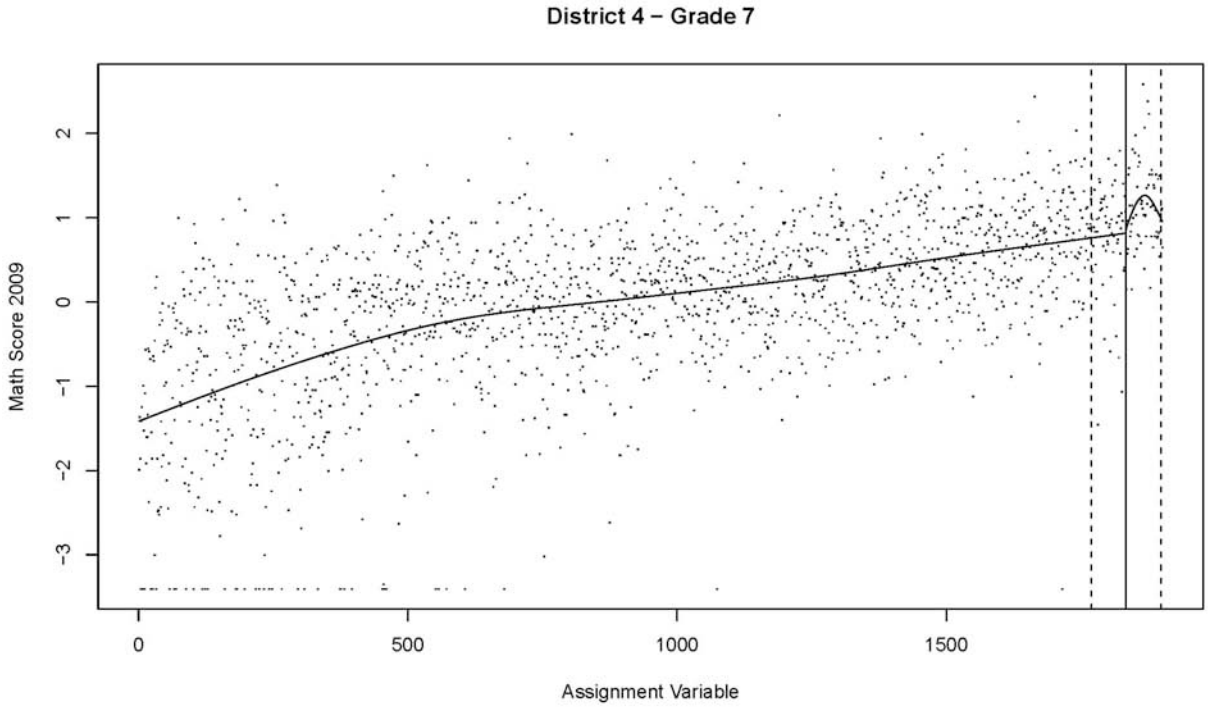
Figure F.15b. Reading Score versus Assignment Variable for Mini-Study 15



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

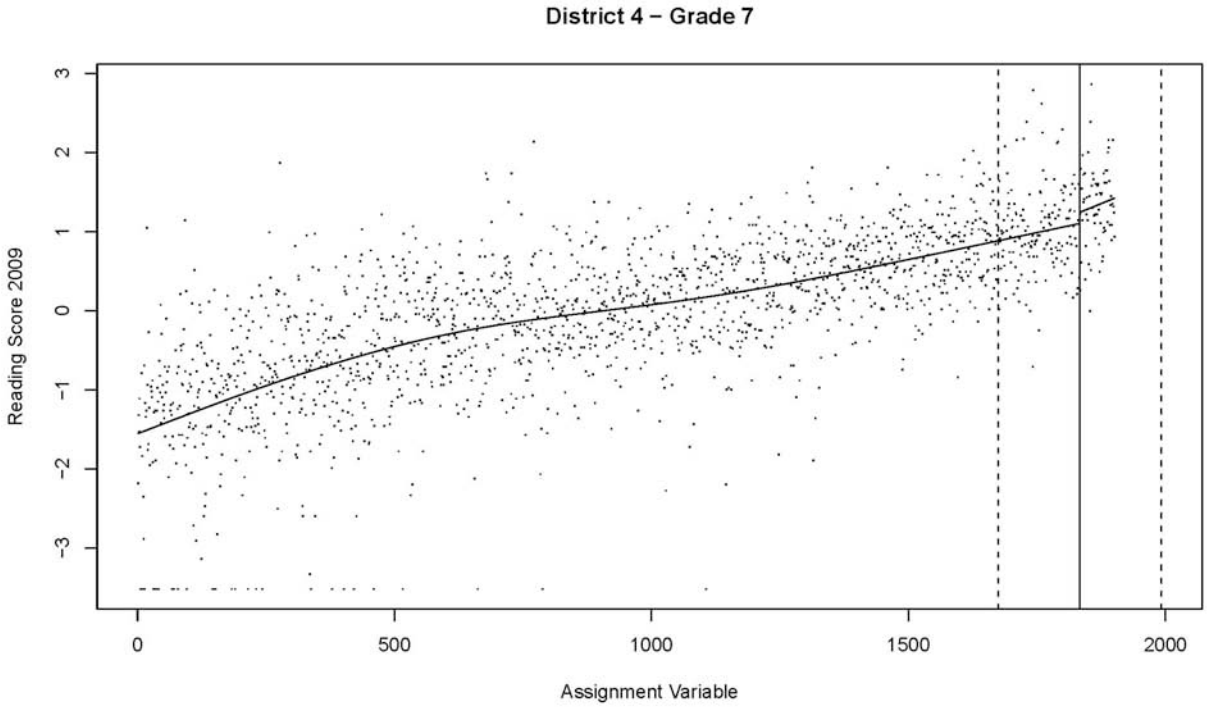
Figure F.16a. Math Score versus Assignment Variable for Mini-Study 16



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

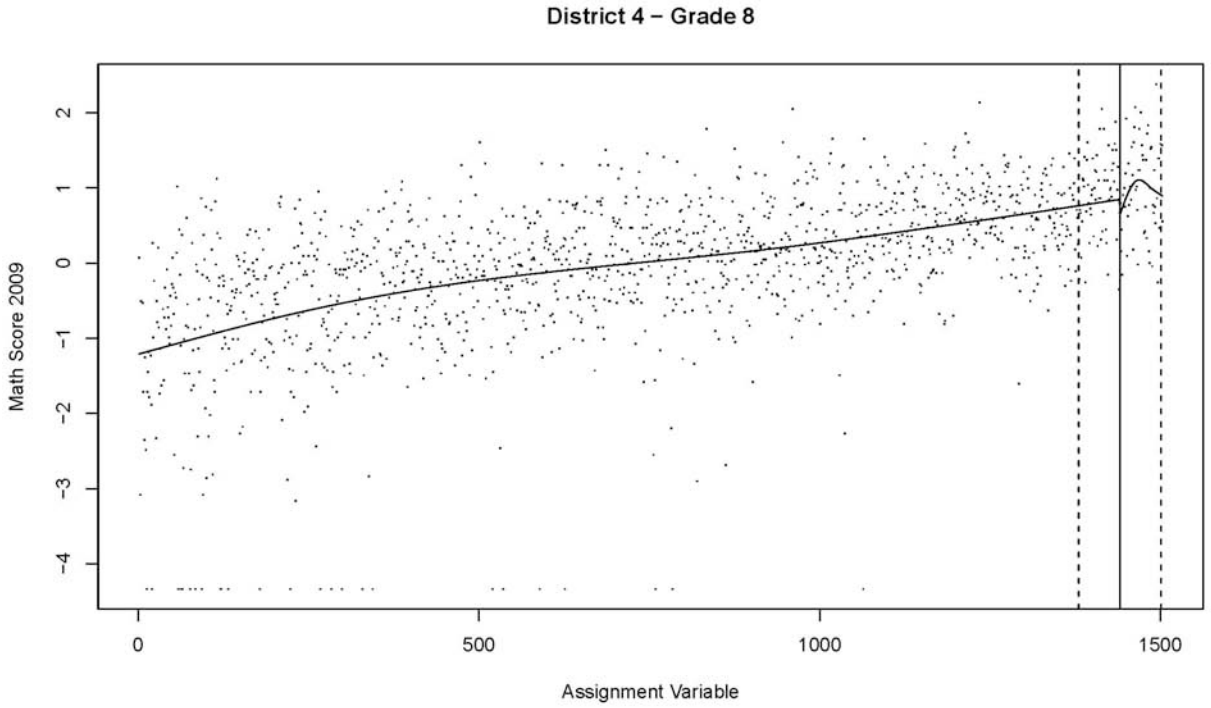
Figure F.16b. Reading Score versus Assignment Variable for Mini-Study 16



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

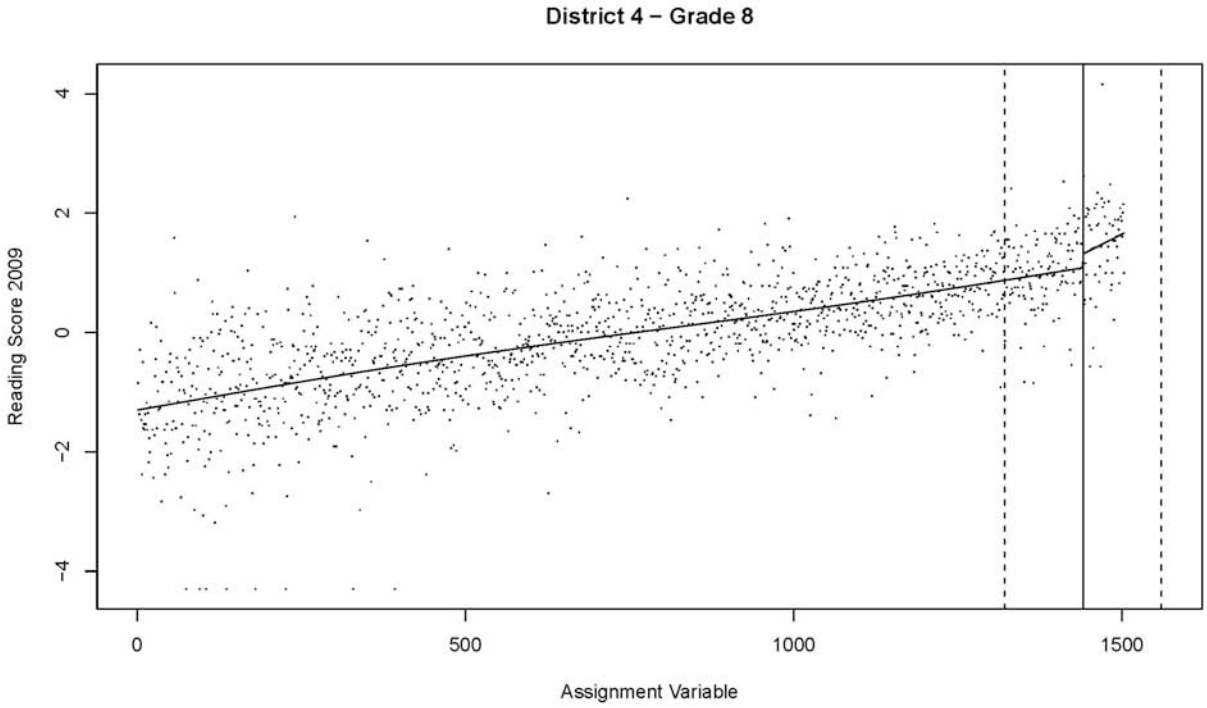
Figure F.17a. Math Score versus Assignment Variable for Mini-Study 17



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

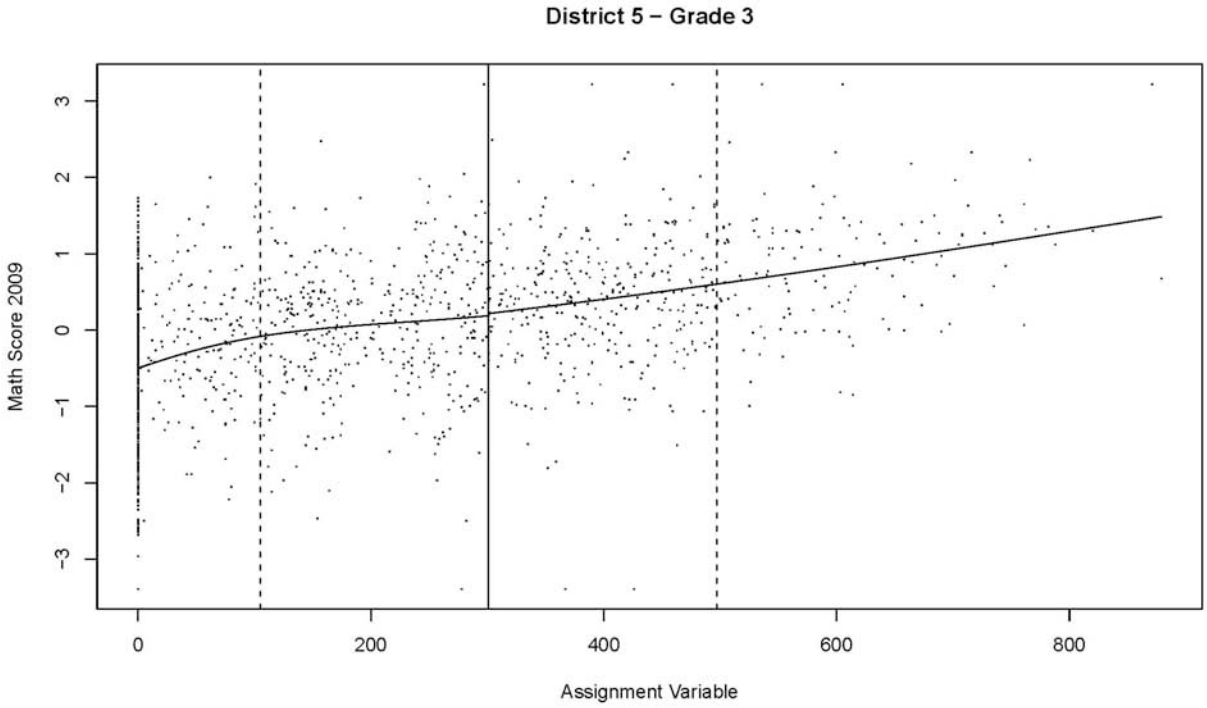
Figure F.17b. Reading Score versus Assignment Variable for Mini-Study 17



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

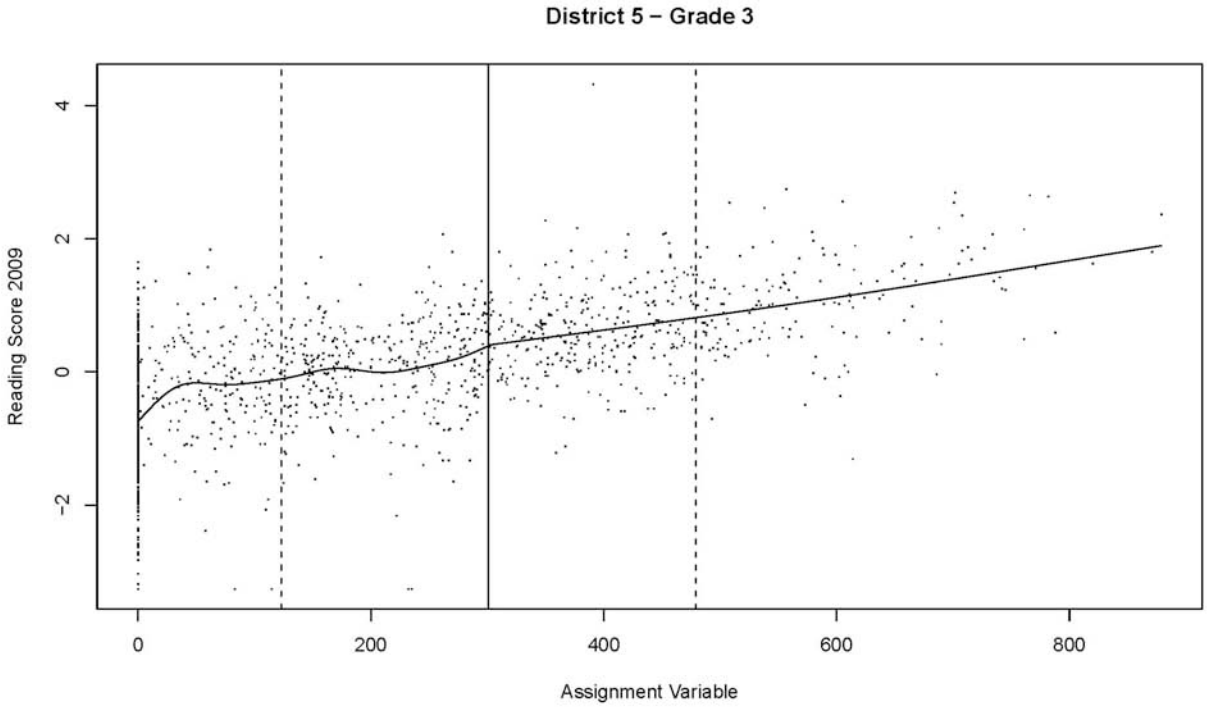
Figure F.18a. Math Score versus Assignment Variable for Mini-Study 18



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

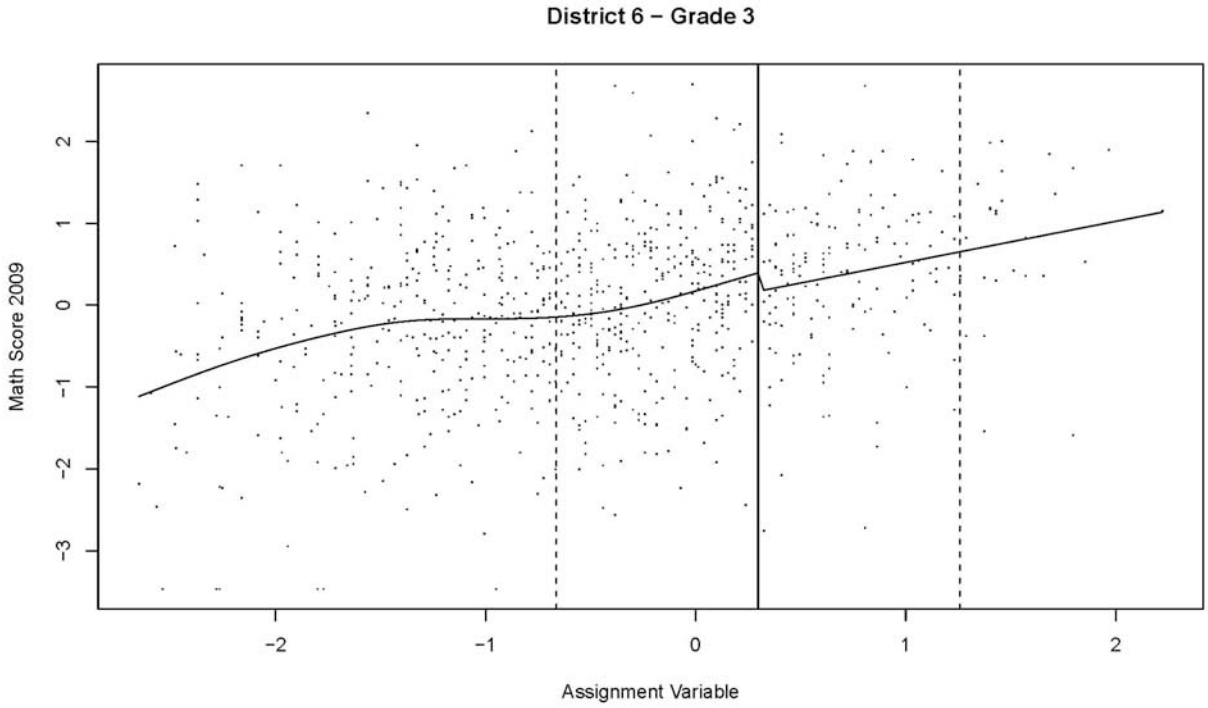
Figure F.18b. Reading Score versus Assignment Variable for Mini-Study 18



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

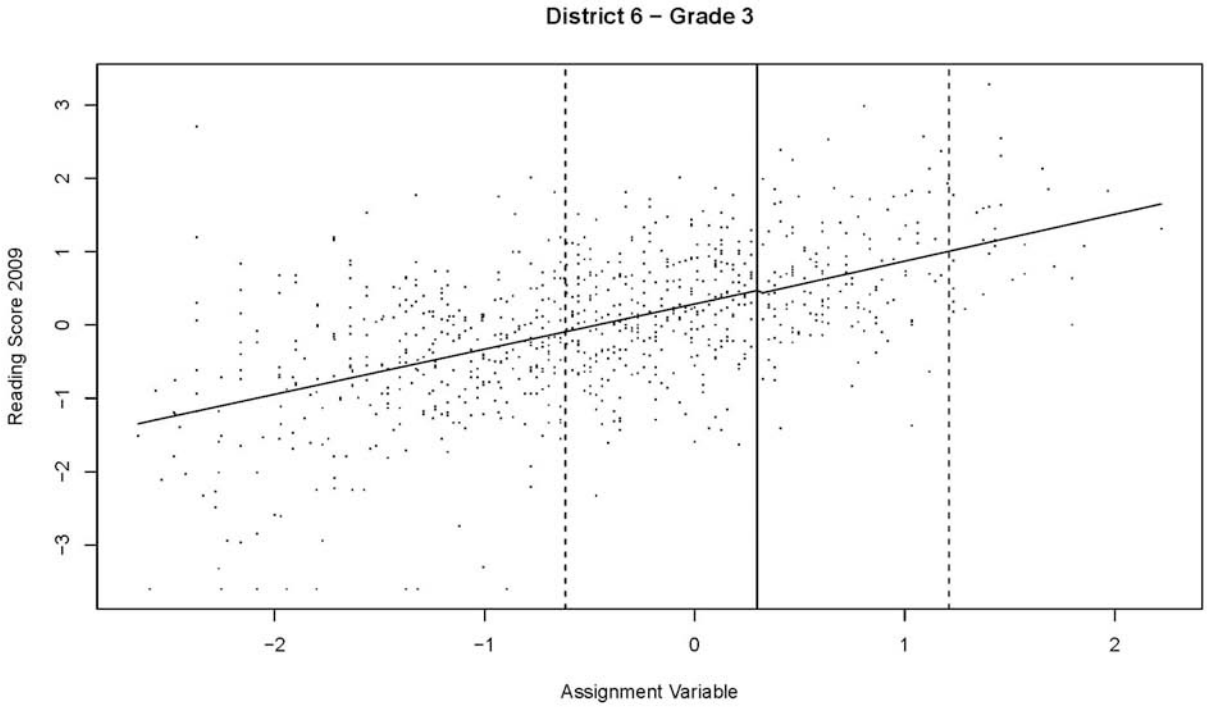
Figure F.19a. Math Score versus Assignment Variable for Mini-Study 19



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

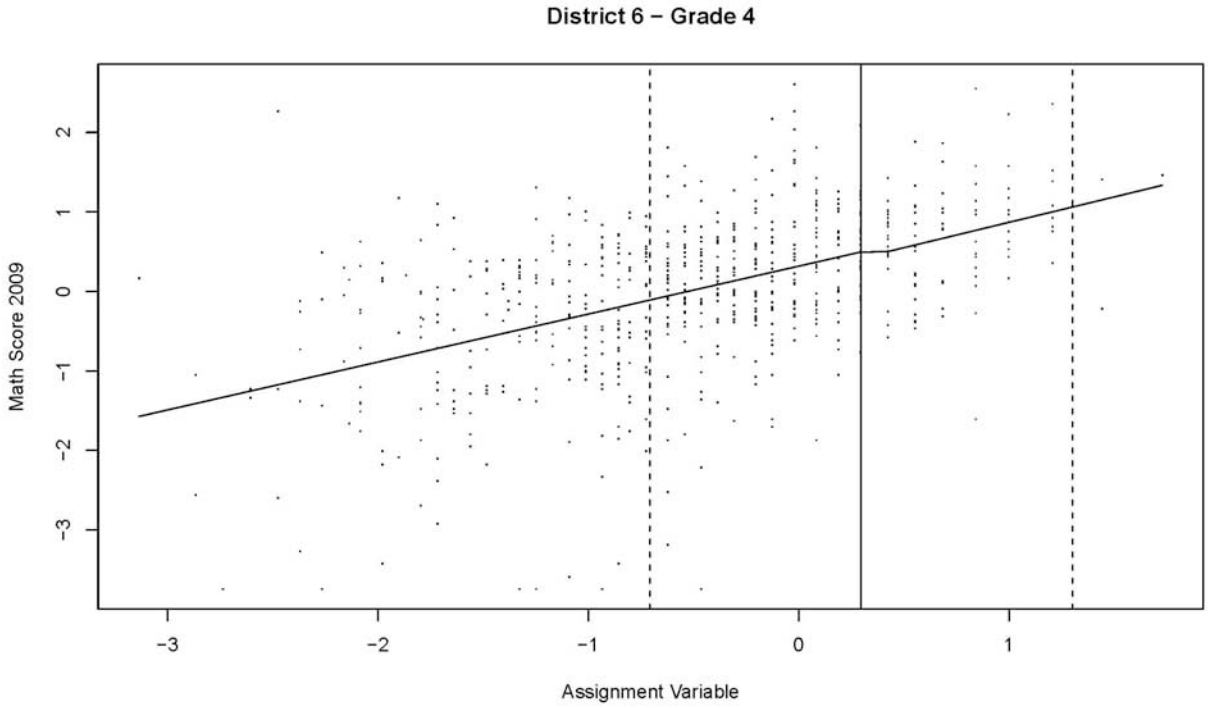
Figure F.19b. Reading Score versus Assignment Variable for Mini-Study 19



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

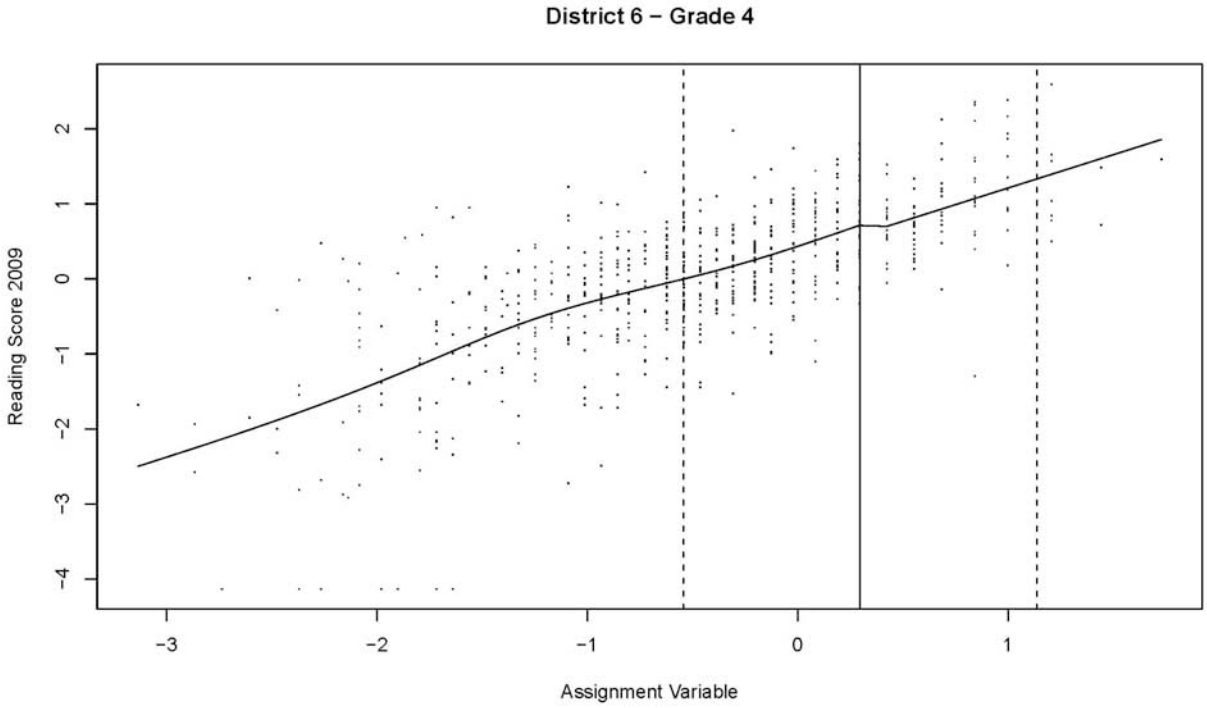
Figure F.20a. Math Score versus Assignment Variable for Mini-Study 20



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

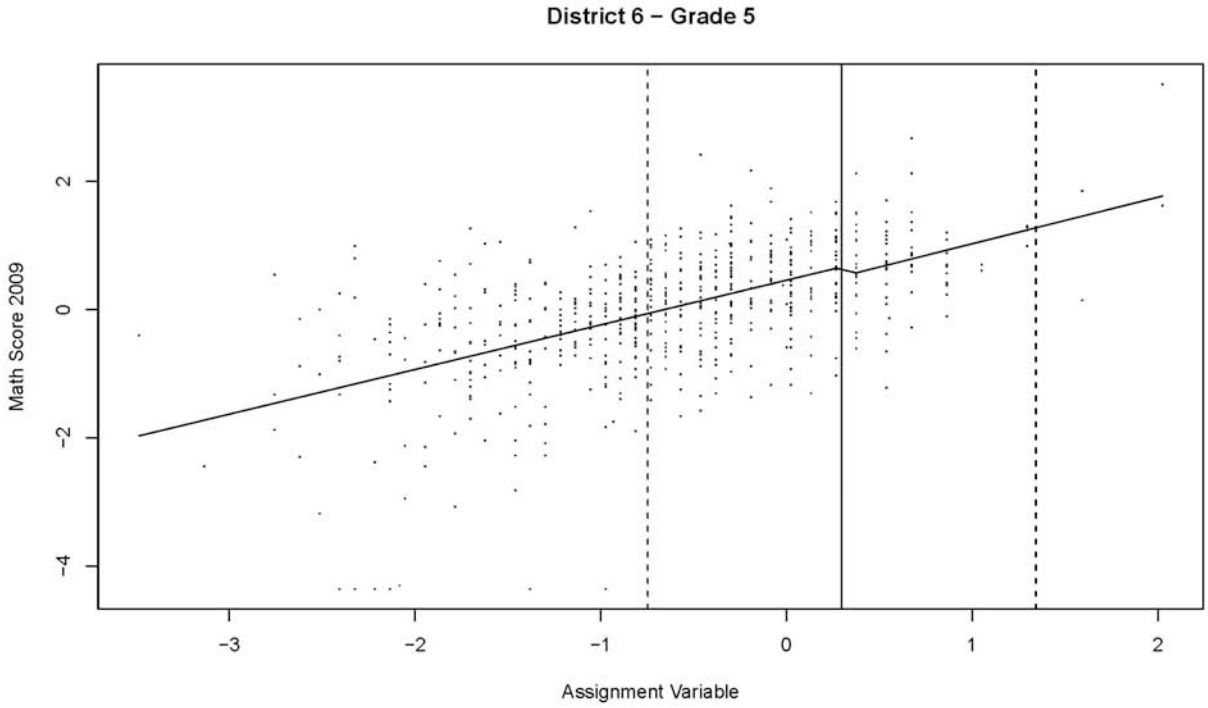
Figure F.20b. Reading Score versus Assignment Variable for Mini-Study 20



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

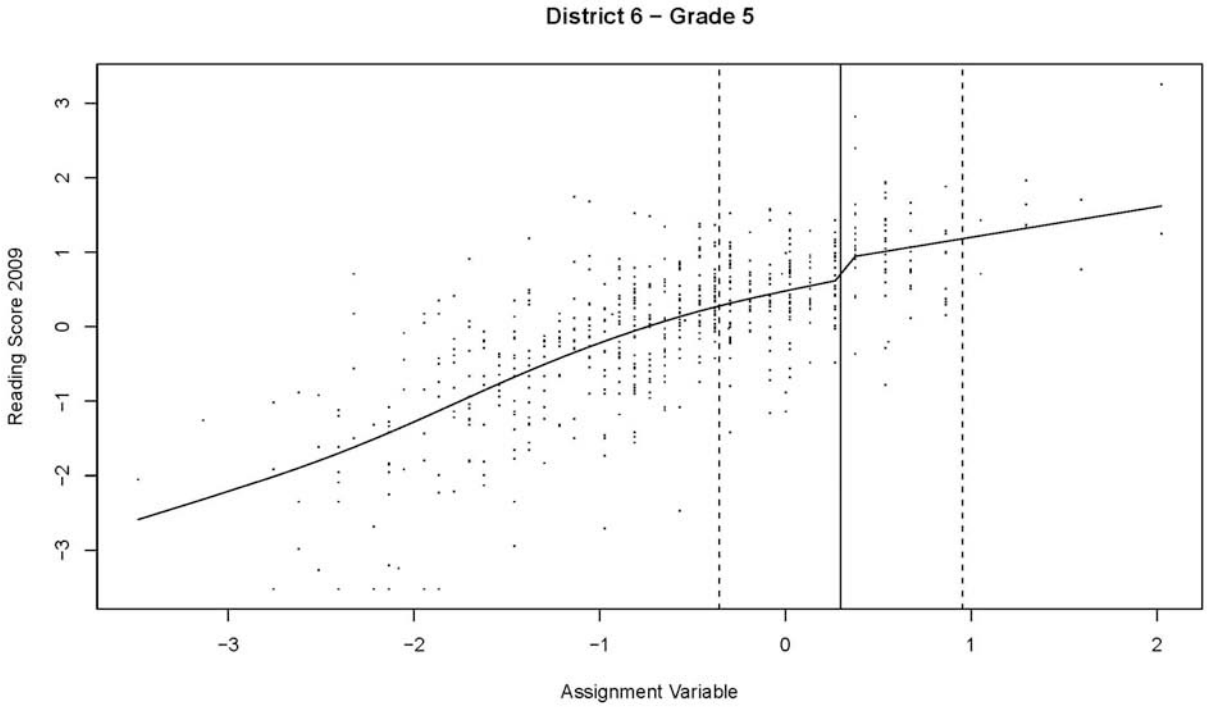
Figure F.21a. Math Score versus Assignment Variable for Mini-Study 21



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

Figure F.21b. Reading Score versus Assignment Variable for Mini-Study 21



Source: School district records.

Notes: Test scores are standardized to have a mean of 0 and a standard deviation of 1. The assignment variable is described in Appendix A on page A.5.

APPENDIX G

SES SURVEY DATA COLLECTION METHODS

A. Instrument Development

Instrument development for the Supplemental Educational Services (SES) Provider Survey began in October 2008. Mathematica worked collaboratively with the Institute of Education Sciences (IES) to design the instrument, building from the questionnaire used in the National Longitudinal Study of No Child Left Behind, SES provider survey, in the 2006–2007 school year (Vernez et al. 2009), and adapting it to address the particular research questions for the current study. Mathematica and IES reviewed multiple drafts of the instrument, preparing a pilot-ready, self-administered, paper-and-pencil questionnaire by the end of 2008.

1. Pilot Study

In January 2009, we completed a pilot test of the questionnaire with eight SES providers that were located in districts originally contacted to participate in the main study but were later deemed ineligible. The pilot test was used to identify the questions that respondents did not understand clearly or consistently. District contacts gave us provider names; before sending them the questionnaire, we called to see if they would participate in the pilot. Participants were chosen to cover a variety of provider characteristics, for example, serving one school district or multiple school districts or offering online instruction, to make sure that the questionnaire was able to capture the different environments and was understandable to the full range of SES providers.

Using FedEx, we mailed to the pilot participants two copies of the questionnaire, a student participation spreadsheet, instructions for completing the forms, and a pre-labeled FedEx return envelope. We asked participants to return a completed copy of the questionnaire to Mathematica in the pre-labeled envelope and to keep the second blank copy and the student participation spreadsheet for reference during a debriefing call we would schedule with them. Upon receipt of completed questionnaires, we contacted respondents by email to schedule the debriefing at their convenience. The debriefing discussion, which was based on a protocol developed for the pilot study, was conducted by Mathematica project staff; in it, we asked respondents specific questions about their experience completing the instrument. Each debriefing lasted approximately 1 to 1.5 hours. Pilot respondents were paid $125 for returning a completed questionnaire and participating in the debriefing. One of the key findings from the pilot test (documented separately) was that more than half the providers said that they were not allowed to give out individual level information on the student participation spreadsheet and that the information would be available through the district, not through the provider. We revised the plan to collect this student-level information from the districts.

In February 2009, after the pilot was completed, a final version of the self-administered questionnaire was prepared for administration. From that version, we developed a telephone interviewer version of the questionnaire, with questions adapted for ease of interviewer administration.

2. Questionnaire Content

The questions were designed to obtain descriptive data intended to show the variation and range among providers. More importantly, the data were intended for analytical purposes, to see whether impact estimates on student achievement were related to SES provider characteristics, such as instructional staff experience, services offered, required training, or communications

between the SES providers and others. The questionnaire was divided into four content sections: Provider Characteristics, Staff Characteristics, Services, and Communication.

a. Provider Characteristics

The section on provider characteristics asked organization-level questions, including the kind of organization (such as faith-based, for-profit, etc.), the number of school districts and states in which the organization operated, the number of years it offered services, and the different locations at which it offered services.

b. Staff Characteristics

The goal of this section was to gather information on the instructional staff the organization hires for SES, including staff requirements, training offered, kind and number of years of experience, and hourly wage they are paid.

c. Services

This section of the questionnaire collected information about the services offered by a provider in a particular district. Questions included the grades and subjects in which SES were offered, information received about students in advance of tutoring (such as state assessment scores or report card grades), what and how many assessments were conducted on SES students, and the contribution of parents and teachers to the process of developing student learning plans. There were also questions about what the providers considered in developing their own curricula (such as state standards or local curricula), the sources of their curricula, what teaching aids they used, how rigid or flexible were the SES sessions, session duration and frequency, and how the sessions were offered (face-to-face, one-on-one, small groups, etc.).

d. Communication

The final section of the questionnaire was about communication between the SES provider and parents, teachers, school staff, and district staff. It was intended to focus on communication regarding instructional matters such as progress and performance, rather than administrative and management communications such as billing or logistics. Questions asked about mode, frequency, and topics of communications, as well as barriers or challenges to communication, such as language barriers or bad contact information. This section also included a question about the nature of the relationship between the district staff and the organization, characterizing the district staff on a continuum from difficult, unresponsive, and adversarial to supportive, responsive, and collaborative. There was a final question that measured the organization's awareness of the curriculum materials (title and publisher) used in the district in two grades, 3rd and 7th, in two subjects, reading and math.

B. Provider Survey Sample Frame

The sample frame was compiled from information sent to Mathematica by each of the school districts that met all study criteria (namely, being oversubscribed for the 2008–2009 school year and using a continuous measure of prior achievement to assign students to treatment and comparison group). School districts gave Mathematica the names of state-approved providers that supplied tutoring services to students in the district under the SES program during

the 2008–2009 school year. Mathematica compiled contact information for each provider, including organization name, address, contact name, email address, and any phone numbers for the contact person as well as an alternate contact name and phone number. Mathematica staff conducted additional research when districts provided incomplete information or when no street address was given, thereby limiting our ability to send out the materials by FedEx as planned.[1]

We combined the separate provider lists into one complete list to determine if there were provider organizations that served more than one district. If the contact and organization name for a district matched another contact and organization name for a different district, that provider organization was considered a "multi-district" provider. The multi-district providers received a slightly different version of the letter that described why there were multiple questionnaires in their packets and how long it would take to complete them. Each provider organization was assigned an eight-digit ID number that was used to keep track of returns and follow-up attempts.

C. Data Collection

Data collection for the provider survey began at the end of June 2009 and extended until the second week of November 2009. Efforts to reach providers included three FedEx mailings (the original mailing plus two nonrespondent follow-up mailing), two email follow-up reminders, and telephone follow-up calls encouraging each nonrespondent to return its questionnaire by fax or FedEx or to complete it on the telephone with the interviewer during the follow-up call. Additional targeted follow-up efforts were directed to the providers serving the most students, including an email or telephone call from the district contact asking the provider to participate in the study. Table G.1 summarizes the mailing and follow-up efforts.

Mailing packets for the first two mailings included the following items:

- Letter from IES, on Department of Education (ED) letterhead and electronically signed by the project officer, explaining the use and importance of the data collection
- Study endorsement letter from the school district on district letterhead, electronically signed by the SES coordinator (or other representative)
- Questionnaire booklet labeled with the name of the district served and provider's unique ID number and barcode
- Check request form preprinted with the name of the district served and provider's unique ID number[2]
- Pre-addressed, prepaid FedEx return envelope

The questionnaire packets for the third mailing contained the same materials included in with the first and second shipments, except that the ED letter was replaced with a letter from the

[1] In the end, we used USPS Priority Mail for providers whose only address was a post office box.

[2] Survey respondents were paid $30 in appreciation for their time and effort to complete the questionnaire. Respondents were provided with a check request form to indicate where the check should be sent, and they returned those forms to Mathematica with the completed questionnaires.

Mathematica survey director, stating that we had unsuccessfully tried to make contact numerous times and encouraging the respondent to complete and return the questionnaire. Multi-district providers received one large envelope with separate questionnaire packets for each district served. After the first mailing and throughout the interview process, we attempted to find the correct address or other contact information for the providers whose questionnaires had been returned.

As shown in Table G.1, the telephone follow-up period lasted 16 weeks in an effort to improve participation. Telephone interviewers were given a five-hour training session that covered a description of the study and goals, a question-by-question review of the provider questionnaire, review of general interviewing guidelines and procedures, and a role-playing practice session. Follow-up efforts were focused on large providers, those that served the greatest number of SES students in our study sample, because they would have the greatest impact on the statistical power of the correlational analyses. These providers received personal phone calls from the survey director and additional email solicitations; at the request of Mathematica, they were also contacted directly by the district contacts, who urged them to participate in the study.

Table G.1. Summary of the Survey Mailings and Followup

Mode of Contact Effort	To Whom	Date(s)
Mailing 1 (mostly FedEx)	All providers in original sample	June 26, 2009
Mailing 2 (mostly FedEx)	Nonrespondents to initial mailing	July 7, 2009
Phone followup	Nonrespondents	July 20, 2009– November 5, 2009
Email followup 1	Providers with whom we had never made contact or made indirect contact only (spoke with a gatekeeper)	August 4, 2009
Mailing 3	Nonrespondents to previous efforts (excluded providers who had recently said they would return the questionnaire)	August 19, 2009, and September 7, 2009 (depending on start of school in district)
Email followup 2	Nonrespondents to the third mailing	August 26–27, 2009
District followup	At our request, six districts contacted their largest providers by email or phone to encourage their participation.	September and October, 2009

Source: Mathematica Policy Research.

D. Response Rate and Weighting

As noted in the data collection section, we made a concerted effort to gain cooperation from the providers who served the most students. The goal was to match the provider survey data with the individual student records for the students it served; so that more students would have matched provider survey data for analysis, it was critical that the largest providers participate in the survey. Table G.2 shows response rates, both unweighted (with each provider counting as one response) and student-weighted (each provider weighted by the number of students in the

data file that the provider served during the 2008–2009 school year),[3] for the six districts included in the data analysis.

Table G.2. Survey Response Rate

	A	B	C	D	E
	Ineligible (Did Not Serve Students Included in Analysis)	Completed Survey	Did Not Complete	Unweighted Response Rate (B/B+C)	Student-Weighted Response Rate
District 1	0	4	0	100%	100%
District 2	12	31	5	86.11%	96.45%
District 3	10	12	6	66.67%	72.79%
District 4	0	45	16	73.77%	90.29%
District 5	24	31	8	79.48%	82.43%
District 6	6	21	7	75.00%	73.20%
Total	**52**	**144**	**42**	**77.42%**	**88.77%**

Source: Mathematica Policy Research.

Notes: The total from columns B and C differs from the number of providers listed in Chapter III, Table III.1. Providers that had not been identified by the districts, but were later determined to be providers, did not receive a survey but are included in the sample for the achievement analysis.

E. Data Processing

Four quality control staff received a two-hour training session on how to review and edit the data recorded on the completed and returned questionnaires. The training included a description of the study purpose and goals, a review of the general editing guidelines, special instructions for this questionnaire (reviewing question by question), and a review of administrative processes. Quality control was ensured during a practice session in which participating staff edited questionnaires that had been mocked up with different kinds of errors. The staff being trained then reviewed the edited questionnaires with the trainers to be sure that the quality review and editing had been done correctly and to address questions about the process.

Completed questionnaires were processed as they were returned, and the processing consisted of the following steps:

1. Questionnaire receipt (through a database that included all cases in our sample frame).
2. Review of the questionnaire and edit for internal inconsistencies based on the questionnaire skips as well as assignment of missing data codes (to distinguish erroneously skipped items from properly skipped items). Ten percent of quality control staff work was reviewed by a supervisor.

[3] For students served by only one provider, the student-based provider weight increases by one for each student served. If a student was served by multiple providers, the student weight assigned to each provider is a fraction equal to the proportion of that student's total tutoring hours that were provided by that provider.

3. Callbacks to providers who skipped items that had been identified as critical for inclusion in the analysis.
4. Key data entry with 100 percent verification of all numeric items.
5. Data cleaning for consistency and creation of new variables.

Staff processing the data were trained to determine if the curriculum used by the provider was the same as the district's reported curriculum. Staff looked at each of the curriculum materials listed by the SES provider and compared it to the name/publisher of the district's reported materials. Codes were assigned for an exact match, a possible or partial match, or a non-match or if the district did not use a set or available curriculum. Successful matching was rare. Table G.3 shows that looking across all curricula that providers used, 5 were an exact match for its district/subject/grade, 12 were possible matches, and the vast majority (388) were non-matches.

Table G.3. Results of Curriculum Alignment Coding

Coding Result	Number
Exact match	5
Possible match	12
Non-match	388
No set curriculum	34

Source: Mathematica Policy Research.

APPENDIX H

SUPPLEMENTAL MATERIALS TO CHAPTER III: SES PROVIDER CHARACTERISTICS AND STUDENTS' SES EXPERIENCES

This appendix includes supplemental tables referenced in Chapter III, presenting information on SES provider characteristics and students' SES experiences based on information supplied by providers that served students in the analysis sample and that responded to our survey. We show unweighted data—that is, each provider is given an equal weight to portray the characteristics and practices of the average provider—as well as data weighted by the number of students the provider serves. The student-based weighting is described in Appendix G.

Table H.1. Content Areas Offered through SES, by Grade Level

	Unweighted Frequency	Unweighted Percentage	Student-Weighted Percentage
Primary Elementary Grades			
Reading/language arts/English	135	94.4	99.1
Mathematics	115	80.4	79.5
Social studies/history	+	+	+
Science	+	+	+
Writing	24	16.8	9.8
Homework help	29	24.2	10.8
Test-taking skills	60	50.0	34.8
Study/organizational skills	55	45.8	24.2
Other subjects	8	6.7	3.7
Intermediate Elementary Grades			
Reading/language arts/English	96	69.6	83.4
Mathematics	86	62.3	72.9
Social studies/history	0	0.0	0.0
Science	0	0.0	0.0
Writing	14	10.1	6.4
Homework help	19	15.1	6.1
Test-taking skills	44	34.9	29.8
Study/organizational skills	35	27.8	15.4
Other subjects	6	4.8	3.5
Middle School Grades			
Reading/language arts/English	108	77.7	85.6
Mathematics	96	69.1	75.6
Social studies/history	+	+	+
Science	+	+	+
Writing	18	12.9	6.3
Homework help	23	18.1	6.9
Test-taking skills	54	42.5	30.6
Study/organizational skills	46	36.2	19.4
Other subjects	5	3.9	3.4

Source: SES provider survey.

Notes: The weighted numbers are based on number of students the provider served, as reported in the student data; if a student was served by more than one provider, the weighting factor is proportional to the hours served by each provider.

The percentage calculations are based on non-missing data only, so the denominator varies depending on the number of missing values. Reading/language arts/English, math, social studies, science, and writing were asked about in one question and the other subjects in another, once for each grade grouping. It is among the six questions that the number of missing answers varies. Thus the same frequency can have a different percentage associated with it.

+ Suppressed to protect respondent confidentiality.

Table H.2. Student Learning Plan (SLP) and Teacher Involvement and Other School Staff Involvement

	Unweighted Frequency	Unweighted Percentage	Student-Weighted Percentage
Teacher Contribution to Development of SLP—Approved Plan but Did Not Contribute Substantively to Development			
Most (more than 50%)	28	20.9	22.8
Some (25–50%)	15	11.2	34.2
Few (less than 25%)	30	22.4	19.9
None	61	45.5	23.2
Total	134	100.0	100.0
Missing	10	--	--
Teacher Contribution to Development of SLP—Contributed Substantively to Content of the Plan			
Most (more than 50%)	21	15.2	9.1
Some (25–50%)	24	17.4	40.9
Few (less than 25%)	39	28.3	29.8
None	54	39.1	20.3
Total	138	100.0	100.0
Missing	6	--	--
Teacher Contribution to Development of SLP—No Involvement in Developing or Approving Plan			
Most (more than 50%)	46	33.8	21.5
Some (25–50%)	12	8.8	14.8
Few (less than 25%)	23	16.9	45.9
None	55	40.4	17.8
Total	136	100.0	100.0
Missing	8	--	--

Source: SES provider survey.

Note: The weighted numbers are based on number of students the provider served, as reported in the student data; if a student was served by more than one provider, the weighting factor is proportional to the hours served by each provider.

Table H.3. Frequency of Topics Addressed with Teachers

	Unweighted Frequency	Unweighted Percentage	Student-Weighted Percentage
Does Not Communicate with Teacher	6		
Student Attendance at SES			
At least once a week	15	12.5	6.3
A few times a month	11	9.2	17.4
At least once a month	60	50.0	56.6
A few times a year	23	19.2	17.6
Never	11	9.2	2.1
Total	120	100.0	100.0
Missing	18	--	--
Student Progress Related to Classroom Instruction			
At least once a week	6	4.8	1.3
A few times a month	13	10.5	17.9
At least once a month	61	49.2	30.4
A few times a year	30	24.2	42.4
Never	14	11.3	8.0
Total	124	100.0	100.0
Missing	14	--	--
Upcoming Standardized Tests or Student Progress Related to SES Activities or Goals			
At least once a week	8	5.9	2.7
A few times a month	23	17.1	21.6
At least once a month	84	62.2	71.1
A few times a year	17	12.6	3.8
Never	3	2.2	0.8
Total	135	100.0	100.0
Missing	3	--	--
Recent Academic Topics Addressed in SES			
At least once a week	6	4.7	2.1
A few times a month	22	17.2	21.4
At least once a month	76	59.4	58.0
A few times a year	15	11.7	12.3
Never	9	7.0	6.2
Total	128	100.0	100.0
Missing	10	--	--
Upcoming Academic Topics to Be Covered in School			
At least once a week	3	2.8	3.2
A few times a month	12	11.0	14.1
At least once a month	42	38.5	23.7
A few times a year	23	21.1	44.6
Never	29	26.6	14.3
Total	109	100.0	100.0
Missing	29	--	--
Student Behavior or Motivation During Sessions			
At least once a week	12	10.2	4.2
A few times a month	12	10.2	15.7
At least once a month	51	43.2	55.2
A few times a year	25	21.2	16.4
Never	18	15.3	8.4
Total	118	100.0	100.0
Missing	20	--	--

Table H.3 *(continued)*

	Unweighted Frequency	Unweighted Percentage	Student-Weighted Percentage
Specific Learning Needs of the Student			
At least once a week	13	9.9	2.9
A few times a month	13	9.9	18.4
At least once a month	73	55.7	64.4
A few times a year	28	21.4	13.5
Never	4	3.1	0.9
Total	131	100.0	100.0
Missing	7	--	--

Source: SES provider survey.

Note: The weighted numbers are based on number of students the provider served, as reported in the student data; if a student was served by more than one provider, the weighting factor is proportional to the hours served by each provider.

Table H.4. Did SES Provider Know the District Curriculum—Summary and by District

	Unweighted Frequency	Unweighted Percentage	Student-Weighted Percentage
All Districts			
One correct	20	13.9	10.1
Two correct	11	7.6	6.8
Three correct	6	4.2	5.3
Four correct	5	3.5	24.7
All wrong or don't know	92	63.9	48.4
Missing	10	6.9	4.7
Total	144	100.0	100.0

Source: SES provider survey.

Note: The weighted numbers are based on number of students the provider served, as reported in the student data; if a student was served by more than one provider, the weighting factor is proportional to the hours served by each provider.

Table H.5. Frequency of Communication to Parents About Child's Progress, by Mode

	Unweighted Frequency	Unweighted Percentage	Student-Weighted Percentage
By Telephone			
At least once a week	15	10.6	4.3
A few times a month	31	21.8	21.4
At least once a month	37	26.1	20.8
A few times a year	42	29.6	50.0
Did not use phone	17	12.0	3.5
Total	142	100.0	100.0
Missing	+	--	--
By Email			
At least once a week	5	3.5	0.6
A few times a month	6	4.3	10.9
At least once a month	8	5.7	3.7
A few times a year	23	16.3	11.3
Did not use email	99	70.2	73.5
Total	141	100.0	100.0
Missing	3	--	--
In Person			
At least once a week	49	34.3	16.9
A few times a month	26	18.2	20.2
At least once a month	28	19.6	15.3
A few times a year	31	21.7	44.7
Did not meet in person	9	6.3	2.8
Total	143	100.0	100.0
Missing	+	--	--
Notes Sent Home with Students			
At least once a week	12	8.6	5.5
A few times a month	25	17.9	26.1
At least once a month	49	35.0	27.1
A few times a year	31	22.1	33.0
Did not send notes home	23	16.4	8.2
Total	140	100.0	100.0
Missing	4	--	--
By Postal Mail or Other Mode of Communication			
At least once a week	7	4.9	1.9
A few times a month	16	11.3	17.2
At least once a month	52	36.6	44.5
A few times a year	38	26.8	23.4
Did not use other mode	29	20.4	13.0
Total	142	100.0	100.0
Missing	+	--	--

Source: SES provider survey.

Note: The weighted numbers are based on number of students the provider served, as reported in the student data; if a student was served by more than one provider, the weighting factor is proportional to the hours served by each provider.

+ Suppressed to protect respondent confidentiality.