Supplemental Education Services Under No Child Left Behind: Who Signs Up, and What Do They Gain?

Carolyn J. Heinrich, Robert H. Meyer and Greg Whitten

EDUCATIONAL EVALUATION AND POLICY ANALYSIS published online 7 May 2010
DOI: 10.3102/0162373710361640

Educational Evaluation and Policy Analysis
Month XXXX, Vol. XX, No. X, pp. xx–xx
DOI: 10.3102/0162373710361640

Supplemental Education Services Under No Child Left Behind: Who Signs Up, and What Do They Gain?

Carolyn J. Heinrich
University of Wisconsin-Madison

Robert H. Meyer
Wisconsin Center for Education Research

Greg Whitten
University of Pittsburgh

Schools that have not made adequate yearly progress in increasing student academic achievement are required, under No Child Left Behind (NCLB), to offer children in low-income families the opportunity to receive supplemental educational services (SES). In research conducted in Milwaukee Public Schools, the authors explore whether parents and students are aware of their eligibility and options for extra tutoring under NCLB, and who among eligible students registers for SES. Using the best information available to school districts, the authors estimate the effects of SES in increasing students' reading and math achievement. Their nonexperimental estimates suggest no average effects of SES attendance on student achievement gains. They use qualitative research to explore possible explanations for the lack of observed effects.

Keywords: *supplemental educational services, student achievement, No Child Left Behind*

In the White House proposal for the No Child Left Behind Act (NCLB), supplemental educational services (SES), or "extra tutoring," are described as a "consequence" or "corrective action" for schools that fail to make adequate yearly progress for disadvantaged students (see http://www.whitehouse.gov/news/reports/no-child-left-behind.html). Describing his view of the role of SES available under NCLB at Woodridge Elementary and Middle Campus, Washington, D.C.,[1] on October 5, 2006, President George W. Bush stated,

> And when we find a child that needs extra help, there's money to do so. And there are options for parents. . . . A parent can enroll their child in a free intensive tutoring program. There's money for that. If your child is not up to grade level early on, there's extra help available for each family to do so. . . .

This project was supported with funding from the Milwaukee Public Schools. The authors would like to thank the following individuals from the Milwaukee Public Schools for their assistance and advice: Mary Ellen Ruzga, Mary Gissal, Mary Dessereau, Bonnie Pionke, Deb Lindsey, Shannon Gordon, Matt Michala, and Cindy Raven. They would also like to thank their colleagues at the University of Wisconsin: Patricia Burch, Huiping Emily Cheng, Annalee Good, John Keltz, Sarah Mason, Ernie Morgan, and Larry Schultz.

> You'd be amazed at the number of districts that don't use this extra tutoring. They don't take advantage of the extra money to help an individual child. Oh, they'll figure out ways to spend it, don't get me wrong. But the money is aimed for helping an individual succeed, and it's the cumulative effect of bringing these students up to grade level that will enable us all to say we're more competitive for the future.

As enacted in Title I, Section 1116(e) of the Elementary and Secondary Education Act (ESEA) and reauthorized by NCLB, schools that have not made adequate yearly progress in increasing student academic achievement for 3 years are required to offer parents of children in low-income families the opportunity to receive extra academic assistance.[2] These services typically include tutoring and remediation in reading and mathematics and are provided outside of the regular school day by public or private (nonprofit or for-profit) organizations, such as public, charter, and private schools; educational service agencies; higher education institutions; faith-based and community-based organizations; and other private businesses. According to the law, the content and educational practices of SES should be aligned with the state's academic content standards (and applicable federal, state, and local health, safety, and civil rights laws) [Section 1116(e)(12)(B)(i)] and should be based on high-quality research evidence of their effectiveness in increasing student academic achievement [Section 1116(e)(12)(C)]. In fact, the law requires states to withdraw approval from SES providers that fail for 2 years to increase student academic achievement.

Research to date has shown, however, that there is little information available on the effectiveness of different organizations entering the market to provide SES, beyond some internal performance evaluations conducted by the larger national providers (Burch, Steinberg, & Donovan, 2007).[3] As a result, states and school districts face considerable challenges in assessing the contributions of SES to students' academic outcomes, both before and after entering agreements with SES providers, which has important implications for the efficient functioning of SES and for the accountability goals of NCLB. Take-up of SES among eligible students has also been low, intensifying concerns among state and school district officials about the effectiveness of SES and complicating their ability to measure it (Government Accountability Office [GAO], 2006).

Public schools in Milwaukee, Wisconsin—the site of this research—account for the vast majority of schools identified for improvement (SIFI) in Wisconsin, four fifths on average from 2002 to 2007. Using administrative data from Milwaukee Public Schools (MPS) and information collected through focus groups with parents, MPS student surveys, interviews with MPS personnel, and field research with SES providers, we undertake analyses to investigate the implementation of SES in Milwaukee and the potential effectiveness of these services in increasing student achievement. In particular, we address the following key questions: Are parents and students aware of their options for extra tutoring under NCLB, and who among the eligible students registers for SES? Based on the best information available to school districts, is it possible to identify the effects of SES in increasing students' reading and math achievement?

We begin now with a brief review of studies on SES implementation and effectiveness and related literature on after-school programs and the public and private educational services market. We next describe the study data, samples, and methods of analysis, followed by a presentation of the data analyses and findings. We conclude with a discussion of the findings and their implications for state educational agencies, school districts, and students who are the intended beneficiaries of SES. In general, we find no average effects of SES attendance on student achievement gains and just a few small effects of the number of SES hours attended.

What (Little) We Know About SES

In his comments at Woodridge Elementary and Middle Campus in Washington, D.C., President Bush suggested that many school districts "don't take advantage of the extra money to help an individual child," implying that they would prefer to spend the Title I funds intended for SES in other ways. Peterson (2005) makes the same point more directly, explaining that school districts can divert SES dollars to their own uses by "suppressing parental demand" and discouraging student participation; in other words, "they have a clear financial disincentive" (p. 44) to enroll

students in SES. School districts acting in their financial interest to limit SES participation would be in direct conflict with the law, however, which requires them "to promote maximum participation by providers to ensure, to the extent practicable, that parents have as many choices as possible" and to notify parents of the availability of SES and allow them to select "any approved provider that they feel will best meet their child's needs" [Section 1116(e)(4)].

State and local educational agencies have, in effect, been charged with the major responsibilities for initiating SES and ensuring compliance with NCLB specifications, including establishing processes for identifying and approving providers and encouraging their services to expand choice for students; developing, implementing, and publicly reporting on standards and methods for monitoring SES quality and provider effectiveness; managing contracts and financial systems for paying providers; and designing procedures for withdrawing approval from unsuccessful providers. For many state and local educational agencies, the development of these procedures for coordinating and evaluating SES, with no new federal funding, is still a work in progress. The Government Accountability Office (GAO, 2006) reported that three fourths of states were experiencing challenges in designing methods to assess academic progress among students and in developing data systems for tracking and analyzing SES information.

In addition, participation in SES among eligible students has generally been low nationwide. The onset of SES was initially slowed by the NCLB stipulation that only schools that had not made adequate yearly progress for 3 consecutive years were required to offer SES. Although participation rates increased from 12% of eligible students receiving services in 2003–2004 to an estimated 23% (3 million) in 2006–2007 (Education Industry Association, 2007; GAO, 2006; Peterson, 2005), the GAO (2006) still reported that low parent and student demand for SES was a challenge in approximately two thirds of the districts they studied.

Take-up of SES is affected by a number of factors that differ across school districts, including the number of students eligible for SES relative to the Title I funds available to school districts, the implementation of the program (e.g., outreach and awareness among staff and students and administrative processes for enrolling students), and the responses of eligible students and their families to information about service offerings. School districts also suggest that it is difficult for SES to compete with other after-school programs and extracurricular activities, particularly for high school students. In Illinois, where typically only 5% of the eligible student population participates in SES, high school program administrators indicated that it was not only difficult to get students registered, but their attendance at SES sessions also declined significantly as the school year progressed (GAO, 2006; Peterson, 2005). Recognizing this problem, some providers offer students incentives to sign up for services and/or to encourage student attendance, ranging from computers to school supplies and gift certificates.

Acknowledging that some challenges in implementing SES are clearly related to incentives and policy design as well as the politics of NCLB, Burch and Good (2009) also show that central to the effectiveness of SES are the details of providers' instructional practices. In their qualitative research on the "instructional core" of SES, they explore "how educators approach and understand their role in the instructional setting" (p. 5), the activities and resources used in instruction, the nature of interactions between students and providers, and the institutional and structural elements that influence choices that SES providers make in establishing instructional practices. As they note, these features of SES are among the least visible to states and school districts, given that SES takes place outside of the regular school classroom and instructional practices can vary significantly not only between providers but also within the same provider depending on the setting and the specific tutor. In fact, the legislation strongly discourages any attempt by states and school districts to regulate instructional choices.

For some of these same reasons that state and local educational agencies have been challenged in their efforts to acquire knowledge of SES content and effectiveness, researchers have also been limited in their ability to conduct rigorous evaluations of SES. Early research on SES was largely descriptive and exploratory, focusing mostly on the challenges of implementing SES in a rapidly evolving market with limited capacity and asymmetric information on both the demand and

supply sides (Anderson & Laguarda, 2005; Burch et al., 2007; Casserly, 2004; Farkas & Durham, 2006; Padilla & Lopez-Torkos, 2006; Steinberg, 2006; Sunderman & Kim, 2004; Vegari, 2007). These researchers have documented the large and growing number of diverse organizations that have entered the market to compete for available SES funds (more than $2.5 billion), their widely varying instructional strategies, curriculums, tutor qualifications and hourly charges, and the dominance of the larger, national for-profit providers.

Empirical Studies of SES Effectiveness

A growing number of studies have sought to empirically estimate the effects of SES. For example, Chatterji, Kwon, and Sng (2006) studied the early effects of SES in one New York elementary school and concluded that program effects were evident only in skills test scores that were aligned with the SES curriculum, and these effects were described as "gross" and "tentative," that is, "confounded with those of other reforms and supports concurrently aiming to raise student achievement" (p. 30). A Chicago Public Schools (CPS) study of students participating in SES in 2003–2004 and 2004–2005 assessed changes in their Iowa Test of Basic Skills scores from one spring to the next and concluded that students receiving at least 40 hours of tutoring had larger gains in reading and mathematics than students who did not receive SES (Ryan & Fatani, 2005). The most recent CPS study (CPS, 2007) estimated generalized linear models to predict state achievement test scores, focusing on SES-eligible students in Grades 4 through 8 who were not English language learners (ELLs), and importantly, who received at least 30 hours of SES tutoring,[4] and reported that SES tutoring increased reading and math achievement among elementary school students (compared to other low-income students attending the same schools but not receiving SES), with gains greater in math than in reading.

Another recently released study (Rickles & Barnhart, 2007) of SES in the Los Angeles Unified School District asked two questions similar to those addressed in this study: How many eligible students used SES (in 2005–2006), and did SES affect their California Standards Test (CST) score gains? As in the Chicago study, the researchers estimated linear regression models to predict the CST scores, controlling for students' 2004–2005 scores and relevant characteristics. Comparable to the prior year's evaluation results, the study authors reported low SES participation (approximately 7% of eligible students), and they found that even among students with the highest levels of SES attendance, the effects of the program were "fairly small" (attributed to improved performance by elementary students). For students with low to medium attendance, no statistically significant effects of SES on student achievement gains were found. Likewise, two studies in Minnesota that used matched samples of students who took the Northwest Achievement Levels Tests (NALT, a national norm-referenced test) to compare students who did not participate in SES with those receiving SES did not find positive effects of SES participation or significant differences among SES providers as assessed by changes in students' annual NALT reading scores (Burch, 2007).

A RAND study (Zimmer, Gill, Razquin, Booker, & Lockwood, 2007) explored the effects of SES across multiple, geographically distinct school districts (Baltimore, Chicago, Long Beach, Los Angeles, Palm Beach, Philadelphia, San Diego, and Washington, D.C.). The study included samples of elementary, middle, and high school students who participated in SES during 1 or more school years, 2002–2003, 2003–2004, and/or 2004–2005 (in the early stages of implementation); consistent with other research, the RAND study reported the highest participation rates among elementary school students. Using a fixed-effects specification to compare changes in students' test scores before and after SES participation with the trajectories of nonparticipating students, they found positive, statistically significant effects of SES on students' reading and math test scores in five of seven districts. A few notable study limitations suggest, however, that the study findings should be weighed cautiously. First, students from all grades and districts were grouped together in estimating SES effects, and not all districts contributed data on students in each grade; the reported findings do not distinguish effects by grade. In addition, the

data supplied by school districts did not allow RAND researchers to observe the number of hours of SES attended or, in some cases, to distinguish between SES registration and attendance.

Another study by Mathematica Policy Research, Inc. (MPR), and COSMOS Corp. is currently collecting data from school districts in which there are more SES applicants than can be served with the available Title I funding to facilitate a regression discontinuity approach to evaluating the effect of SES. In addition, a new study by Burch, Heinrich, and Meyer that extends their Milwaukee-based research to four additional school districts (including CPS) will apply regression discontinuity methods in two of the sites where SES is oversubscribed. These studies will take advantage of the necessity of school districts to establish formulae (based on additional criteria for participation) and generate a cutoff point for students' eligibility for SES (based on available funding) to determine who participates in SES. This design will limit the estimation of the potential effects of SES to students within a specified bandwidth of the cutoff point, and prior research suggests that not all applicants above the cutoff would actually follow through and attend SES. The potential of this research design to produce a relatively accurate and generalizable effect estimate of SES is yet to be determined.

Insights From Other After-School and Tutoring Program Studies

Of course, other after-school study and tutoring (or "out-of-school-time") programs have long been in operation, including federally funded programs, and there is a larger literature on their implementation and effects. Lauer et al. (2006) conducted a recent synthesis of the research evidence on out-of-school-time programs specifically in response to the new NCLB requirements to offer SES. They acknowledged up front that although evaluations of after-school tutoring and related interventions are profuse, relatively few are rigorous in their research design and methodological approach. Selecting only those studies (35) published in peer-reviewed journals and using control or comparison groups to estimate effect sizes (e.g., gain scores), they used meta-analysis techniques to explore the relationship of program focus, duration, timeframe, student grouping, and grade level to program outcomes. They find, based on their review, that out-of-school-time programs can have a positive effect on student achievement (in relation to at-risk students who do not participate), although the effects are not likely to be large enough to close the achievement gap between at-risk students and those who are more advantaged. In addition, effect sizes were larger for programs that were more than 45 hours in duration but became smaller for those longest in duration.

In a random assignment study of a national after-school program highly comparable to SES, Dynarski et al. (2004) found no effects on reading test scores or grades for elementary or middle school students. A follow-up study using these same data (Vandell et al., 2005) attempted to distinguish high and low activity/quality among the after-school programs, and although they reported positive effects on test scores for elementary school students highly active in high quality programs, no statistically significant program effects were identified for middle school students. A more recent experimental study by Black et al. (2008) involved approximately 1,900 second- through fifth-grade students randomly assigned (within after-school centers) to receive either enhanced, adapted models of regular-school-day math and reading instruction in after-school settings or the after-school services that were regularly available at their schools. An implementation study confirmed that the enhanced interventions were implemented as intended, and the authors reported (after the first program year) positive and statistically significant effects for the enhanced math program on student achievement but weak evidence of effects on reading achievement and no effects of the reading or math programs on student engagement, behavior, or homework completion.

A careful review of this literature also reveals, however, that very few of these studies (the Black et al. study being one of the exceptions) measure attendance or make the distinction between planned program duration and actual student attendance or engagement. Although the research generally suggests a positive association between attendance and program effectiveness, measurement of student contact time or

"exposure" in these interventions has been inadequate for precisely estimating these relationships. Furthermore, the apparent link between student motivation (and other individual and family background characteristics) and attendance/engagement in out-of-school-time programs poses challenges for researchers (i.e., the potential for selection bias) in nonexperimentally identifying the effects of different levels of program intensity or duration on student achievement. At the same time, there are other findings from this literature about the nature and quality of instruction and their relationship to program effectiveness that are relevant to SES. The meta-analysis by Lauer et al. (2006) and related research (Elbaum, Vaughn, Hughes, & Moody, 2000; Lou et al., 1996) show the largest average positive effects for programs that use one-on-one tutoring (for reading) and small-group instruction (for mathematics). The positive effect sizes of these instructional approaches on student reading achievement were highest for students in the lower elementary grades and at the high school level, whereas effect sizes for mathematics achievement were larger among middle and high school students (Lauer et al., 2006).

Finally, as suggested above, estimating the effects of SES and other educational interventions on academic achievement is particularly difficult in a nonexperimental context, where participation is a choice that may be affected by supply-side factors as well as student and/or parental characteristics. In fact, a persistent concern is that low-income, less advantaged parents or students will be less capable of navigating the market due to informational constraints, lack of experience, and other limits to choice associated with their poverty. This concern also has been expressed in regard to SES, particularly in light of information asymmetries and procedural difficulties in disseminating information critical to informed choices and so on (Burch, 2009; Steinberg, 2006). The complexity of student selection into multiple stages of participation in educational interventions is an important substantive and methodological issue that we grapple with in our nonexperimental study of SES implementation and effectiveness in MPS. We now describe our data, samples, measures, and methods used in the analysis to address the selection problem and to estimate SES effects on student achievement.

Study Data, Samples, and Measures

The data used in this study include primary data collected in parent focus groups, student surveys, interviews with provider and district staff, and information from observations of SES providers as well as secondary student-level demographics, transcripts, and achievement test data from MPS (see the timeline of data collection in Figure 1). The focus groups with parents of students enrolled in eligible SES schools for 2005–2006 were conducted by Heinrich (lead author) with the support of three doctoral students in August 2006 in a formative stage of the research intended to explore basic issues about how parents learned of SES (or *if* they knew of the program), how they chose a provider for their child, their opinion of the tutoring's effects on their child's academic performance, and any difficulties that they encountered in arranging services. A random, stratified sample of 320 parents/students was drawn first from MPS administrative data to ensure that we included both parents (or guardians) whose children had registered for SES in the 2005–2006 school year (approximately 60%) and those who chose not to participate in SES. Letters were mailed to parents by MPS to invite them to participate in the focus groups, which took place in public libraries easily accessible by public transportation and in zip codes with high concentrations of SES schools. Although the total number of focus group participants ($n = 17$) was low as a percentage of the invited sample (5.3%), the participants in the three groups that met were diverse and the size of the groups was advantageous for meaningful discussion (see Heinrich & Whitten, 2006, for further details).[5] The focus group protocol is shown in Appendix A.

The student surveys, administered in March and April 2007 to those who had registered for SES in the 2006–2007 school year, were developed to collect additional information from students about how they were spending time in their tutoring sessions and how these sessions were affecting their performance in school. That is, the collection of these data was important in obtaining another window on what Burch and Good (2009) describe as the "instructional core" of SES. The surveys also asked students about how they chose their SES provider and what would encourage them to attend more tutoring

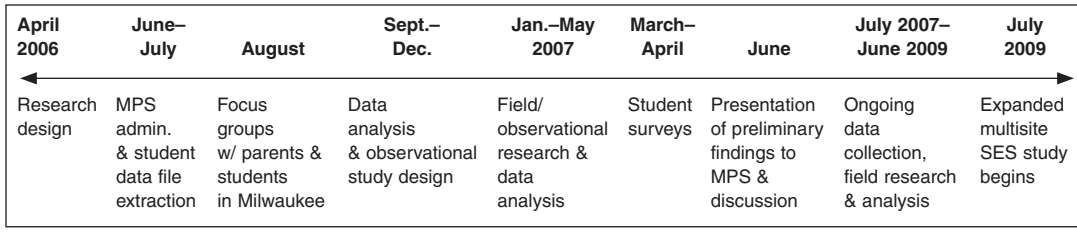

FIGURE 1. *Timeline of study data collection and project research activities.*
Note. MPS = Milwaukee Public Schools; SES = Supplemental Education Services.

sessions. Questions included closed-ended items such as, "In a typical or average SES session, how many minutes do (or did) you spend working *one-on-one* with a tutor in your SES sessions?" as well as open-ended items such as, "What have you learned in your SES sessions this school year?" The survey questions are shown in Appendix B. In establishing the sample frame for these surveys, eight SIFI schools with the largest SES enrollments were selected. Students in seven of these eight schools completed the survey a second time within 1 month (in late March or early April) if they were still participating in SES to assess the consistency of (or changes in) responses. A total of 1,441 students responded to the first survey, and 874 students participated in the second survey. The field research component of this study, including in-depth observations of SES sessions and interviews with tutors and other MPS district staff, was conducted primarily by Patricia Burch and Annalee Good (of the Wisconsin Center for Education Research). We incorporate information from the first phase (January–May 2007) of this qualitative research, involving four SES providers and 22 hours of classroom observation, 15 interviews with provider staff, and six interviews with district-level staff (Burch & Good, 2009; Burch, Good, Heinrich, Meyer, & Whitten, 2007).[6] The four providers studied were selected to allow for diversity in tutoring formats, administrative structures, settings (e.g., home, community, school), and curricular approaches. An observation instrument developed specifically for SES tutoring sessions and based on existing instruments used in research on other after-school programs and state and district SES monitoring forms was used to record details of the classroom-level instructional core, including teacher views on and approaches to instruction, their use of instructional resources, and teacher–student interactions. The observation categories were the facility and associated resources (e.g., tables, desks, computers, blackboards, etc.), participants (type of provider staff present, number of students, etc.), curriculum (content area, skill focus, materials in-use, integration with school day, etc.), instruction (activity focus and type, assessments, etc.), student and tutor engagement, student and tutor interactions, and special student services available (i.e., special education and ELL needs).

With permission and assistance from MPS, we also secured access to students' middle and high school data that come from the administration of standardized tests, databases used by MPS to monitor and manage the SES program, and the MPS eSIS (Electronic Student Information System) database of student transcript and demographic data. The SES program administrative database includes student enrollment and attendance information (with particular SES vendors identified) and other rich student-level data, such as Individual Student Achievement Plans or IAP/ISAP that describe academic goals to be met in tutoring, billing information that allows for the calculation of vendor expenditures on individual students, and student participation in other academic support programs.

To construct measures of student achievement, and particularly, gains in student achievement, we used data from standardized tests administered to MPS students. As the tests used by the school district changed over time, we use data from five different test instruments that were applied in school years 2004–2005, 2005–2006, 2006–2007, and 2007–2008. These data sources are discussed below, and the means and standard deviations of test scores from these data sets are reported in Table 1.

TABLE 1
Sources of Test Score Data by Grade and Year

Grades	2004–2005	2005–2006	2006–2007	2007–2008
5–8	old WKCE or TerraNova	new WKCE AGS	new WKCE	new WKCE
9	TerraNova	TerraNova AGS	none	ThinkLink
10	n.a.	new WKCE AGS	new WKCE	n.a.

Note. WKCE = Wisconsin Knowlege and Concepts Examination.

Means and Standard Deviations of Test Scores by Data Set, Grade, and Year

Grade 8–9–10 Cohort, 2004–2005

Subject/grade	Mean	Standard deviation	Reliability	Variance stabilizing multipliers
Reading				
8[a]	664.89	36.87	0.86	1.00
9[b]	664.57	49.65	0.93	1.35
10[c]	490.42	63.78	0.90	1.73
Math				
8[a]	670.77	40.85	0.86	1.00
9[b]	666.55	54.53	0.90	1.33
10[c]	517.00	49.00	0.86	1.20

a. "Old" Wisconsin Knowledge and Concepts Examination, 2004–2005.
b. TerraNova, 2005–2006.
c. "New" Wisconsin Knowledge and Concepts Examination, 2006–2007.

Wisconsin Knowledge and Concepts Examination ("New" WKCE), 2005–2008

	Reading					
	Mean			Standard Deviation		
Grade	2005–2006	2006–2007	2007–2008	2005–2006	2006–2007	2007–2008
5	452.9			48.7		
6	464.3	467.5		52.4	52.6	
7	477.0	476.7	477.5	47.8	49.5	50.8
8		494.4	489.6		51.4	54.9
9			492.5			58.3
	Math					
	Mean			Standard Deviation		
Grade	2005–2006	2006–2007	2007–2008	2005–2006	2006–2007	2007–2008
5	450.8			42.7		
6	471.6	474.0		43.9	43.2	
7	488.6	497.8	492.7	46.7	43.1	43.4
8		504.8	497.1		46.3	54.8
9			507.1			49.3

TABLE 1 (continued)
AGS Assessment, 2005–2006

		Reading			Math		
Grade	Statistic	Fall	Spring	Gain	Fall	Spring	Gain
6	*M*	429.47	443.41	13.45	490.87	494.77	2.96
	SD	28.07	30.49	23.58	10.03	7.48	6.79
	N	607	484	484	678	530	530
7	*M*	441.36	452.13	9.62	493.81	498.05	3.65
	SD	25.76	26.64	18.51	8.19	6.41	5.72
	N	808	671	671	898	732	732
8	*M*	452.24	460.33	5.98	495.53	500.08	3.08
	SD	26.68	26.99	18.21	10.10	7.45	6.49
	N	1,017	826	825	1,152	872	872
9	*M*	454.59	465.96	7.57	497.72	500.77	1.93
	SD	27.36	27.60	21.47	7.18	7.01	5.82
	N	1,950	1,112	1,112	1,795	1,105	1,105
10	*M*	461.42	471.33	6.73	500.03	502.01	0.95
	SD	28.40	28.88	24.59	8.07	7.13	6.18
	N	1,443	862	862	1,392	891	891
11	*M*	469.59	478.21	4.33	501.78	503.49	0.28
	SD	30.15	31.47	24.56	8.95	8.51	6.88
	N	1,221	748	748	1,275	798	798
12	*M*	479.15	483.29	2.05	503.74	504.23	–0.78
	SD	29.27	29.50	20.91	8.59	9.66	8.30
	N	831	419	419	907	395	395

To evaluate the effect of SES during the 2004–2005 school year, we used reading and mathematics scores from three different sources. Achievement prior to participation in SES was measured in November 2004 using the state assessment, the Wisconsin Knowledge and Concepts Examination (WKCE), in Grade 8, and the TerraNova in Grades 6, 7, and 9. Both tests were scored on what we refer to as the "old" WKCE scale. (A new WKCE scale was introduced in 2005.) Achievement after participation in SES was measured in November 2005 using the new WKCE, administered in Grades 3 through 8 and 10, and the TerraNova in Grade 9.[7]

The state assessments in 2004 (and prior years) and in 2005 (and later years) were scored on different scales: Scores at a given grade were typically lower by several hundred points on the new scale, but the variances in scores were very similar. As a result, it is reasonable to measure growth in student achievement from 2004 to 2005 as the gain in student achievement *relative* to the average gain in achievement.[8] Including grade indicators in the model (the strategy used in this research) addresses the issue of differences in the location (means) of the state test scales. The key requirement is that the variances of the two scales are similar at each grade (although possibly different at different grades).[9]

In 2005–2006, AGS[10] assessments were administered to all students enrolled in schools required to offer SES to their students, with tests in both math and reading administered twice in 2005–2006 (one form given in the fall and another form in the spring). The fall and spring AGS tests were scored on the same scale. Measures of student gain reflect growth from fall to spring during a single school year. In 2006–2007 and 2007–2008, student achievement both before and after participation in SES was measured at all grades (other than Grade 9) using the new WKCE. As in 2004, this assessment was administered in November. Achievement in ninth grade in 2007–2008 was measured using the ThinkLink examination. This assessment was administered in September 2007. Ninth-grade scores were not available for 2005–2006.

It is important to make clear the major differences between the assessment data available for evaluating the effect of SES in school years 2004–2005, 2005–2006, and 2006–2007, given their implications for the interpretation of SES effect estimates. First, the data for 2004–2005 covered SES participation only up through 9th grade; the 10th-grade test score (administered in November 2005) serves as a posttest outcome for 9th-grade participation. Second, the data for 2006–2007 covered SES participation only up to 8th grade as a pretest was not administered in 2006 in 9th grade. In contrast, the AGS assessment was administered in all middle and high school grades in 2005–2006, although a substantial number of middle and high school students did not take the assessment (see Table 1). In contrast, there was very little missing test information for the other assessments because all students were required by the state or MPS to take the state assessment or the TerraNova assessment. Our statistical analysis of the patterns of missing test data in 2005–2006 suggests that they are not missing at random; as expected, the more disadvantaged students (free lunch recipients, those in special education, retained, with lower grade point averages, and more absences) were significantly less likely to have test score information. Consequently, the analyses reported for the 2 years represent somewhat different high school grades and populations of students.[11]

Another important difference between the AGS assessment and the state assessment (and related assessments) is that the growth period is approximately 7 months (October–April) for the AGS assessments (2005–2006) and 12 months (November–November) for the latter assessments. This difference is unlikely to affect our analysis because the services provided by the SES program were concentrated in the period between December and March, well within the test intervals demarcated by the dates of the pretests and posttests in both years.[12] Finally, it is also the case that the units (standard deviations) of the AGS test scale and the state assessment (and related assessments) differ substantially, as indicated in Table 1. This information implies the following: (1) Parameter estimates should generally be larger for analyses based on the state data than the AGS data (2005–2006 only). (2) In the AGS data (2005–2006 only), parameter estimates based on reading scores should generally be larger (by a factor of three) than estimates based on math scores.

Estimation Methods

To estimate the causal effect of SES on student outcomes, we would ideally like to randomly assign some fraction of the eligible students to receive services and others to a control group of students who would be denied access to SES. This would ensure that those benefiting from SES would be statistically equivalent to those not participating across all observed and unobserved student characteristics. With a randomized experimental design, the effect of SES would be calculated simply as the difference between the outcomes of participating students and those in the control groups. This was not an option in this study, however, given the federal mandate to make SES available to as many eligible students as funding allowed and the low SES registration rates that do not beget rationing of services.

In the absence of random assignment, we can only estimate potential effects or associations that are suggestive of a possible causal interpretation of the findings. In doing so, it is critically important to understand and model the process by which those who participate in SES choose to receive services. First, we know that NCLB requires school districts to use the same data to determine eligibility for SES that they use for making within-district Title I allocations, which historically has consisted of information on free school lunch eligibility. School districts are required to notify families of their children's eligibility for SES and to cooperate with approved providers in disseminating information to students and parents about available services. Only if more families request SES than there are funds available to serve them are districts obligated to establish priorities or criteria to determine which eligible students get access to services.[13] At the same time, we know that applying for SES typically takes some initiative on the part of parents and students in following through the steps of registration and choosing a provider. Thus, although we alleviate some selection concerns by using an internal comparison group of students who are also eligible but do not sign up

for SES to compare with those who register to receive services (Heckman, Lalonde, & Smith, 1999), there may still be some selective differences between these two groups of students that we need to adjust for in our analysis.

In this study, we investigated selection into SES using multiple methods. As data are not systematically collected on how parents and students decide to participate in SES (or not) after receiving information about the services, we conducted focus groups with parents and some students to probe and gain insight into their decision making processes (see Appendix A and Heinrich & Whitten, 2006, for further details). Student responses to the surveys administered to eligible students in the 2006–2007 school year were also informative about selection into SES. In addition, we used transcript and administrative data for SES-eligible students to predict their probability of registering for SES and the probability of attending SES (conditional on registration) using logistic regression. The predicted probabilities generated by the logistic regression models subsequently serve an important purpose—this information is used to remove possible bias associated with preintervention differences between students who participated in SES and those who did not register, allowing for more accurate estimation of potential SES effects.

Propensity Score Matching

In their widely cited work on the use of matching to nonexperimentally estimate program effects, Heckman, Ichimura, and Todd (1997) describe "the evaluation problem" as "a missing data problem." Most basically, individuals cannot be observed both participating and not participating in a program at the same time. It is also common, however, even in experimental evaluations, for other problems to arise that lead to missing data. For example, individuals selected to participate (randomly or otherwise) may not follow through and receive services, or follow-up data on their outcomes after participation may not be available for the full sample. In other words, any effort to estimate program effects will almost always compare imperfectly matched participants and nonparticipants.

In attempting to identify the effects of SES in increasing students' reading and math achievement, we compare eligible students who registered for SES with eligible students who did not register; among students registered for SES, we also compare the effects of differing levels of hours of SES attended. We use propensity score matching methods—matching on the predicted probabilities from the logistic regressions noted above—to account for observed differences between the groups, including differences in student characteristics and in the public schools they attended.[14] We also recognize, however, that this approach will be inadequate in the face of any important, unobserved factors that influence both selection into SES and student outcomes. Indeed, Wilde and Hollister (2007) recently compared experimental estimates of the Project STAR class-size reduction experiment on student achievement test scores with estimates produced using propensity score matching and concluded that the nonexperimental results were insufficiently close to experimental effects to place confidence in them, which they attributed to the role of unobservable characteristics. Agodini and Dynarski (2004) reached a similar conclusion in their study that compared experimental and propensity-score effect estimates of dropout prevention programs.

Because our dependent variables (measuring changes in student achievement) are defined as the difference between pre- and posttreatment outcomes (test scores) for each student, we are able to use a "difference-in-differences" version of the matching estimator that allows for a time-invariant (unobserved) difference in outcomes between participants and comparison group members and often performs better than its cross-sectional counterpart, in part because it allows for this unobserved heterogeneity (Mueser, Troske, & Gorislavsky, 2007). The specific matching technique we apply is radius caliper matching, with the caliper set at 0.01. In other words, students who did not register for (or attend) SES are matched with students who participated if their predicted probabilities of registration are within 0.01 of each other.[15] We also impose a "common support" requirement, so that if there is not a match between students in the two groups, the student is dropped from the analysis; no more than two cases were dropped (among both high school and middle school students) due to a lack of a common support. After students are matched, the differences in their changes in achievement

test scores—the change in test scores from the fall before SES to spring or the following fall after SES—are calculated, along with bootstrapped standard errors and bias corrections. In light of recent work suggesting that bootstrapping may not be appropriate with nearest neighbor matching methods (Abadie & Imbens, 2006), we also applied alternative techniques, including local linear matching methods and "nnmatch" in STATA, which yielded comparable results.

Fixed-Effects Model

Primarily for the purpose of sensitivity testing, we use a fixed-effects model specification to account for the possibility that students who participated in SES may differ from nonparticipants in ways that are correlated with achievement growth. Equations 1 and 2 below describe linear growth models that capture the effects of SES in year 1 (η_1) and year 2 (η_2), respectively. Because student achievement is assessed in the fall in Milwaukee, growth during year 1 is captured by year 2 achievement Y_{2i} minus year 1 achievement Y_{1i}. Similarly, growth during year 2 is captured by year 3 achievement Y_{3i} minus year 2 achievement Y_{2i}. In addition to the indicator for any SES attendance, the equations include controls for a vector of individual characteristics (X_i), with coefficient vectors β_1 and β_2, and regular school effects α_1 and α_2 (where S_{1i} and S_{1i} represent vectors of regular school indicators for years 1 and 2, respectively). We hypothesize that the error terms in each equation are composed of a fixed effect u_i and transitory errors e_{1i} and e_{2i}, respectively.

$$Y_{2i} - Y_{1i} = \zeta_1 + \eta_1 SES_{1i} + \beta_1' X_i + \alpha_1' S_{1i} + u_i + e_{1i} \quad (1)$$

$$Y_{3i} - Y_{2i} = \zeta_2 + \eta_2 SES_{2i} + \beta_2' X_i + \alpha_2' S_{2i} + u_i + e_{2i} \quad (2)$$

Ordinary least squares regression estimates of these two equations will yield biased SES coefficients if the unobserved student fixed effect is correlated with participation in SES in either year. To eliminate the fixed effect, we can difference the two equations, yielding the following difference-in-differences model:

$$(Y_{3i} - Y_{2i}) - (Y_{2i} - Y_{1i}) = (\zeta_2 - \zeta_1) + \eta_2 SES_{2i} - \eta_1 SES_{1i} + (\beta_2 - \beta_1)' X_i + \alpha_2' S_{2i} - \alpha_1' S_{1i} + (e_{2i} - e_{1i}) \quad (3)$$

Equation 3 yields separate, unbiased estimates of the potential effect of any SES participation in years 1 and 2 given the maintained model assumptions, although note that the coefficient on SES_{1i} is the negative of the year 1 SES effect.

More efficient estimates of the potential effect of SES can be obtained by aggregating estimates across years 1 and 2 and across grade levels. Because SES providers often serve students at all grade levels, it is quite reasonable to pool information across grade levels (as in the estimates based on matching methods). Below, we report estimates of potential SES effects for students in middle school and students in high school. Both sets of estimates are based on the most recently available data from MPS.

One important requirement of the fixed-effects model presented above is that the test scores in different years and grades must be measured on the same scale. Because the data set used to estimate the potential SES effect for high school students measures achievement on different scales, we explored several methods for rescaling the test scores to meet this requirement. Because all approaches yielded similar results, we present estimates based on one of the simplest approaches, that is, linearly rescaling the test scores so that they have the same variance in all three grades. Because the data set used to estimate the potential SES effect for middle school students in school years 2005–2006 and 2006–2007 measured achievement on the same scale in all years (the new WKCE scale), there was no need to rescale these data.

Who Signs Up for SES?

As the discussion above suggests, there are multiple stages of student/parent selection into SES, including *awareness* through dissemination of information to students and parents by schools about eligibility and availability of services; *registration* of students with a specific SES provider that is encouraged by providers through marketing (i.e., sending out brochures, inviting parents to presentations, and offering incentives to students to register with them); and following registration with a provider, student *attendance* at SES sessions, for which providers invoice the school district to request payment for the number of hours of service provided. Table 2

TABLE 2
Student Supplemental Educational Services (SES) Eligibility, Registration, and Attendance in Milwaukee Public Schools

Academic year	Eligible (middle and high school)	Number registered (% of eligible)	Number attended (% of registered)
2003–2004	6,508	3,707 (57%)	3,333 (90%)
2004–2005	9,433	3,826 (41%)	2,610 (68%)
2005–2006	7,351	3,996 (54%)	2,543 (64%)
2006–2007	8,119	3,897 (48%)	1,315 (34%)

shows the number of middle and high school students who were eligible for SES, registered for SES, and attended any SES session during the 2003–2004, 2004–2005, 2005–2006, and 2006–2007 school years in MPS. Approximately one-half of eligible students registered for SES over the study period, but the number attending any sessions declined substantially over time, from 90% of registered students in 2003–2004 to only one-third in 2006–2007 (or 16% of *all* eligible students in 2006–2007). It is clear that understanding why more eligible students do not progress through registration and attend at least one session is important not only for school districts and providers trying to comply with NCLB and improve student outcomes but also for research efforts to evaluate the effectiveness of SES.

Although the focus groups with MPS parents and students were primarily exploratory with a small number of attendees (n = 17), an important finding was that despite the considerable information made available by schools and providers to inform parents of SES options for their children, not all parents receive or understand this information. Fifteen of the 17 parents had a hard time distinguishing SES from other school-based tutoring or after-school programs and identifying which options were available free-of-cost to their children, and they indicated that they received little or no assistance in making an informed choice for their child. Many of these parents did not receive the booklet on SES options prepared by the school district, and some reported being skeptical of information received through postcards and other direct mailings from SES providers: "I think a lot of it is smoke and mirrors, you know . . . what they really have to offer as far as online services go"; and from another parent, "Actually it was like a little slick. . . . We went up to the school and they were giving away free stuff." It is interesting that these same parents had a clear idea of what factors they *should* be considering in making these choices. They most frequently responded that they would like to know how much one-on-one tutoring their child would receive (time per session and total number of hours), what the student–teacher ratio would be in group SES sessions, and specific information on the tutor qualifications and academic content of SES sessions. Some of this information was available in the district SES booklet, but only two parents reported seeing or using it. Other parents described important logistical concerns about their children's attendance (e.g., transportation) or their ability to accommodate an in-home provider or online service, leading them to choose options based on convenience and familiarity (e.g., a school-based provider that might involve a student's regular school day teacher). Indeed, approximately 18% of eligible students who responded to the 2007 surveys reported that they missed attending SES sessions because of problems getting to and/or from the location of services.

Table 3 presents basic descriptive information on the middle and high school students who were eligible for SES in the 2004–2005, 2005–2006, and 2006–2007 school years by whether they registered for SES or not. It is apparent that females and Black students in middle and high school are more likely to register for SES, whereas White and Asian students and Hispanic students in high school are less likely to register for SES. Over time, the association between free lunch receipt and student registration for SES appears to become stronger. Other than these associations with demographic characteristics, however, most

TABLE 3

Characteristics of Supplemental Educational Services (SES) Registrants Versus Eligible Nonregistrants (middle school and high school students), 2004–2007

	Registered for SES?—Middle school						Registered for SES?—High school					
	2004–2005		2005–2006		2006–2007		2004–2005		2005–2006		2006–2007	
Variable	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Female	0.45	0.45	0.51*	0.46	0.50*	0.43	0.45	0.45	0.52*	0.43	0.50*	0.43
White	0.03*	0.09	0.03*	0.09	0.04*	0.12	0.04*	0.11	0.07*	0.11	0.05*	0.11
Black	0.74*	0.64	0.70*	0.61	0.78*	0.70	0.86*	0.55	0.67*	0.48	0.80*	0.62
Hispanic	0.20	0.20	0.23	0.23	0.16*	0.11	0.07*	0.27	0.23*	0.34	0.11*	0.20
Asian	0.05	0.05	0.01*	0.05	0.01*	0.05	0.02*	0.05	0.02*	0.04	0.02*	0.05
Indian	0.004	0.006	0.003	0.006	0.003	0.006	0.004	0.009	0.005	0.01	0.005	0.01
Other race	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.02	0.01	0.02	0.01
Grade point average	2.00	2.02	1.92	1.92	1.97	1.98	1.91*	1.82	2.00*	1.78	1.92*	1.76
Took foreign language	0.40	0.43	0.35*	0.31	0.22	0.22	0.25	0.24	0.26	0.26	0.36*	0.29
English proficient	0.86	0.86	0.84*	0.88	0.91*	0.95	0.97*	0.87	0.90*	0.87	0.97*	0.95
Free lunch recipient	0.84*	0.79	0.93*	0.86	0.95*	0.91	0.66	0.65	0.81*	0.78	0.87*	0.82
Special education	0.21	0.20	0.19	0.21	0.22	0.23	0.19	0.21	0.18	0.20	0.20	0.20
Prior year absences	16.7*	19.8	15.7*	21.1	17.7*	20.9	21.8*	30.5	21.8*	31.8	23.8*	31.5
Retained	0.46*	0.49	0.53*	0.45	0.38	0.39	0.27*	0.31	0.29*	0.35	0.25*	0.31
Prior SES attendance	0.15*	0.08	0.20*	0.10	0.19*	0.10	0.09*	0.03	0.16*	0.05	0.12*	0.07
Number of observations	1244	2577	1271	1748	993	1132	1367	7052	1723	4431	2255	5136

*The difference between SES registrants and nonregistrants (in middle or high school in a given year) is statistically significant at $\alpha < .05$.

TABLE 4
Factors Influencing Registration Among Students Eligible for Supplemental Educational Services (SES)

	Middle school			High school		
Variable	2004–2005	2005–2006	2006–2007	2004–2005	2005–2006	2006–2007
Female		+22%	+50%		+32%	+31%
White (vs. Black)				–65%		–57%
Hispanic (vs. Black)				–63%		–60%
Asian (vs. Black)	–76%	–81%	–72%	–61%	–66%	–72%
Indian (vs. Black)						
Other race (vs. Black)						
Grade point average						
Grade point average-squared						
Foreign language course						+16%
English proficient						
Free lunch recipient	+31%	+161%	+54%			+43%
Special education student						
Total absences (prior year)	–0.8%	–0.7%	–0.9%	–1.1%	–0.8%	–0.7%
Retained						
Attended SES prior year	+52%	+147%	+92%	+79%	+181%	+88%
Grade year 6 (vs. 8)				n.a.	n.a.	n.a.
Grade year 7 (vs. 8)				n.a.	n.a.	n.a.
Grade year 10 (vs. 9)	n.a.	n.a.	n.a.			
Grade year 11 (vs. 9)	n.a.	n.a.	n.a.			
Grade year 12 (vs. 9)	n.a.	n.a.	n.a.		–31%	–46%
Pseudo R^2	5.0%	11.3%	10.0%	15.9%	8.9%	7.0%
Number of observations	N = 2,178	N = 1,683	N = 1,374	N = 7,225	N = 5,207	N = 6,635

Note. Each number reported is a statistically significant effect (at $\alpha < .05$)—the percentage increase or decrease in the odds of students registering for SES associated with a given variable in a specific academic year for either middle school or high school students.

other statistically significant associations shown in Table 3 suggest that the SES-eligible students who register to receive services are more advantaged than those who do not sign up: They have a lower number of student absences and are less likely to be retained, and among high school students, they have higher grade point averages and are more likely to be English proficient. In addition, those who have attended SES in a prior year are more likely to register again.

We also explored the factors influencing SES registration and attendance decisions in logistic regression models that were used to estimate the propensity scores for matching, where the dependent (or outcome) variable was a binary variable equal to one if a student registered for SES and zero if the student did not sign up. In addition to the student demographic, school performance, and attendance variables shown in Table 3, these models also included controls for students' grade year and school attended. Bearing in mind the focus group findings, we recognize that relevant information about possible parental influences and family constraints on the decision to register (e.g., parental education levels and employment status, location of residence, etc.) is not included in these models, and to the extent that they differ between registrants and nonregistrants and relate to SES outcomes, they may bias the results.

The results of these analyses are presented in summary form in Table 4, and like the simple descriptive analysis, they show some consistent relationships between student characteristics and school experiences and registration for SES among eligible students. In this table, a blank cell indicates that a particular variable was not (statistically) significantly associated with student registrations in that year.[16] If the variable was a statistically significant predictor, the increase in odds of registering for SES associated with that variable is reported. For example, female SES-eligible students are significantly

more likely to register than males (with 22% to 50% higher odds of registering), and Asian students have approximately 60% to 80% *lower* odds of registering than African Americans (the reference category in the model). In several school years, Hispanic and White high school students are also significantly less likely to sign up for SES than African Americans (with 57–65% lower odds of registering). As expected because of federal directives to prioritize economically disadvantaged students, free lunch eligibility becomes a stronger predictor over time, with up to 161% higher odds for middle school students eligible for free lunch, as does prior SES attendance, with up to 181% higher odds for high school students in the 2005–2006 school year. As also seen in the descriptive analysis, student absences from school (measured in the school year prior to the SES enrollment year) lower the odds of registering for SES.

Because the factors that influence student *attendance* at SES sessions (following registration) differed little from those affecting registration, we more briefly discuss the statistically significant findings of these models (available from the authors). Among the registered students, females and those who attended SES in a prior year were significantly more likely to attend any SES sessions, and student absences from school in the prior year were again negative and statistically significant in all models. The new and substantively important findings were that among middle school students, registered Hispanic students (and those who were not English proficient) were significantly more likely to attend. In addition, contrary to the observed relationship between free lunch eligibility and registration, high school students eligible for free lunch were less likely to attend any SES (although this relationship was statistically significant only in 2004–2005). Registered high school students who had been retained were also significantly less likely to attend any SES sessions. Although not reported in Table 4, most of the school indicators in both the registration and attendance models were statistically significant, suggesting that school-based factors (e.g., teacher roles in encouraging attendance, school location, etc.) also influence these decisions. An analysis using Chow tests for differences in coefficients across schools in the models predicting SES attendance suggested that the factors influencing SES attendance likely differed across schools (although no specific factors were identified).[17]

In general, the findings of the SES registration and attendance analysis suggest some clear associations between SES participation and factors such as free lunch eligibility, regular school attendance, and SES attendance in a prior year. The total variation in decisions to register that is explained in these models is relatively low (5–16%), however, and less than a quarter of the variation in decisions to attend SES was explained.[18] It is possible that this reflects a high degree of uncertainty or arbitrariness in these decisions, as suggested by parent comments in the focus groups, although it is also likely that some relevant unobserved or unmeasured factors (e.g., related to student motivation or barriers to participation) are not captured in this analysis. For example, we know from responses to the spring 2007 surveys that for about a quarter of these students, other activities compete for their after-school time, and some (10–15%) reported being drawn to participate by the incentives and rewards that providers offered (e.g., food, prizes, computers, etc.). Based on our analysis, we speculate that if selection on unobservables is operating, it is likely that more motivated or better prepared students (particularly for high school SES) are following through and receiving SES, which would contribute to an upward bias in estimates of SES effects.

The Effects of SES on Student Achievement

Effects of any SES Attendance

We begin with a simple propensity score matching estimation of the effects of *any SES attendance* on student achievement (our "black box" estimation), where the SES "treatment" is measured using a binary variable that indicates any time spent in SES. In each of the models (for middle school and high school students in the 2004–2005, 2005–2006, and 2006–2007 school years), there was strong overlap in the distribution of propensity scores for registrants and nonregistrants, implying that it was relatively easy to find matches between students who registered

TABLE 5
Estimated Effects of Attending any Supplemental Educational Services (SES), 2004–2005 and 2005–2006 School Years

Treatment measure and method	Middle school		High school	
	Change in math test scores	Change in reading test scores	Change in math test scores	Change in reading test scores
2004–2005 school year				
Attended any SES				
1. unmatched	–2.486 (4.562)	–3.368 (5.232)	–10.486 (6.243)	**–14.420 (7.139)**
2. matching	2.024 (5.557)	3.038 (5.916)	–5.427 (8.107)	–4.565 (8.860)
Number of observations	N = 1,562	N = 1,571	N = 1,224	N = 1,262
2005–2006 school year				
Attended any SES				
1. unmatched	–0.529 (0.413)	0.708 (1.202)	0.235 (0.297)	**2.846 (1.132)**
2. matching	–0.232 (0.427)	0.323 (1.099)	–0.372 (0.357)	1.397 (1.099)
Number of observations	N = 1,075	N = 1,016	N = 2,215	N = 2,200
2006–2007 school year				
Attended any SES				
1. unmatched	–0.112 (3.993)	5.798 (4.566)	n.a.	n.a.
2. matching	0.595 (4.343)	4.022 (5.771)	n.a.	n.a.
Number of observations	N = 462	N = 464	only 10th graders were tested (N = 80), and only 7 of them registered for SES[a]	

Note. Standard errors in parentheses; results statistically significant at $\alpha \leq .05$ shown in bold.
a. In light of the very small number of high school students (of the 2,255 registered for SES in the 2006–2007 school year) who were tested, we do not report SES effects for high school students in this year.

for SES and similar students who did not register.[19] The results of balancing tests used to assess the performance of the propensity scores in balancing the distribution of observed characteristics between the treatment and comparison groups (available from the authors) showed that after matching, there were no statistically significant differences in their mean characteristics, with substantial percentage reductions in bias for most variables (as high as 99.4%).

Table 5 presents the estimated effects of attending any SES on changes in reading and math test scores in the 2004–2005, 2005–2006, and 2006–2007 school years[20]; both unmatched and matching estimates are shown. The results show that after matching, there are no statistically significant differences in the changes in test scores for students who attended SES compared with those who did not attend any SES sessions. The standard errors are relatively large and the estimated differences are positive in half of the cases and negative in the others (and again, none are statistically significant).

This finding of no statistically significant estimated effect of *any* SES attendance on changes in student achievement is generally consistent with the limited effects of after-school tutoring programs reported in the literature. As discussed above, however, researchers have also documented some relationship between the level or intensity of services and their effects, which the simple indicator of any SES attendance would not identify. In fact, there is a wide range of total hours of SES attended by students in these samples (from 1 to 110 hours), with the average number of hours attended as low as 13 hours for high school students in 2006–2007 and as high as 30 hours for middle school students in 2004–2005. This may also reflect, at least in part, that vendors charge widely differing hourly rates for their services; the per-student SES funding allocation is the same fixed dollar amount for all providers, so SES vendors that charge higher hourly rates will necessarily provide fewer total hours of SES to their students.

Relationship Between Hours of SES Attended and Student Achievement

We examined the distributions of hours attended for middle and high school students in 2004–2005, 2005–2006, and 2006–2007 (if they attended any hours), based on provider reports of actual student hours attended that are required to receive payment for services.[21] The patterns were very similar across these four groups, with the highest peaks around 25 hours but also with nearly as high spikes in the distributions close to 0 hours (left skewness). This is consistent with both student and vendor reports in our study, as well as in other research, suggesting that getting students to attend (to show up regularly for SES) is an ongoing challenge in the implementation of SES (GAO, 2006). In the literature discussed above, one study of SES (Ryan & Fatani, 2005) and a meta-analysis of after-school programs (Lauer et al., 2006) suggested that effects of tutoring programs were larger for programs that were more than 40 to 45 hours in duration (albeit with effects diminishing for considerably longer hours). In the MPS samples, approximately 8% of middle school students and 17% of high school students attended 40 or more hours of SES in 2004–2005; in 2005–2006, the comparable numbers were 15% of middle school students and 6% of high school students, and in 2006–2007, just 0.07% of middle school students and 0.05% of high school students attended SES for 40 or more hours.

In analyzing the effects of total hours attended on changes in students' math and reading test scores, we first estimated an ordinary least squares (OLS) regression model with the same set of controls for student characteristics and school attended (as shown in Table 3) and with a continuous measure of total hours of SES attended. The results, reported in line 1 of Table 6 for each of the student subgroups, show only one statistically significant, positive estimated effect of total hours attended on the change in high school student reading scores (in 2005–2006); for each additional hour attended, students' reading test scores potentially increase by 0.087 of a test unit. For a student attending SES for 25 hours, this is still a fairly small gain relative to the variability of gains on this test (approximately 10% of a standard deviation in reading gain). In addition, if it is the case that the relationship between hours attended and changes in student achievement is nonlinear, an OLS model is an inappropriate specification.

We subsequently applied propensity score matching to estimate the effects of different levels of SES hours on students' math and reading achievement. In these analyses, we first limited the samples to only students who registered for SES; in other words, students who registered for SES but did not attend any hours were compared with those who registered and attended different levels of hours (eliminating concerns about selection at the stage of SES registration). In 2004–2005, only a quarter of registered middle school students attended SES, whereas two-thirds of registered high school students attended some SES; in 2005–2006, almost two thirds of registered middle school students and slightly more than one half of registered high school students attended at least one session, and in 2006–2007, about one-half of registered middle school students and slightly more than one-third of registered high school students attended at least one session. Lines 2 and 3 in Table 6 show the results of models that compare changes in test scores for students with 20 or more hours of SES who are matched with registered students who did not participate, and registered students with 40 or more hours of SES matched to registered students with no hours attended. Across all of these estimated effects, there are no statistically significant effects of attending more than 20 hours or more than 40 hours of SES.

In an additional set of analyses, we further restricted the samples to include only students who attended some hours of SES, and then we matched students with differing levels of SES: those with greater than 10 hours of SES versus. fewer than 10 hours; greater than 20 hours of SES versus. fewer than 20 hours; greater than 30 hours of SES versus. fewer than 30 hours; and the same for the 40 hours cutoff. In these analyses, we are no longer concerned about selection at the stage of registration (as all are registrants) or at that which determines whether or not students attend any SES; however, there may still be selection into these varying levels of attendance, which the propensity score matching approach will address to the extent

TABLE 6

Estimated Effects of Total Hours Attended Supplemental Educational Services (SES), 2004–2005, 2005–2006, and 2006–2007 School Years

	Middle school		High school	
Treatment measure and method	Change in math test scores	Change in reading test scores	Change in math test scores	Change in reading test scores
2004–2005 school year				
1. # SES hours attended (OLS)	0.046 (0.068) $n = 1562$	–0.017 (0.068) $n = 1571$	–0.127 (0.158) $n = 1224$	–0.148 (0.178) $n = 1262$
SES hours attended (matching)				
2. at least 20 hours (vs. none)	7.727 (5.921) $n = 1419$	5.256 (6.493) $n = 1428$	–24.240 (14.367) $n = 202$	–7.191 (16.262) $n = 207$
3. at least 40 hours (vs. none)	12.757 (7.774) $n = 1273$	11.648 (8.790) $n = 1282$	–26.676 (14.662) $n = 140$	–24.679 (19.241) $n = 145$
If attended > 0 hours:				
4. >10 hours (vs. <10)	19.503 (16.384) $n = 427$	3.603 (14.744) $n = 427$	–1.840 (14.626) $n = 185$	2.051 (18.510) $n = 190$
5. >20 hours (vs. <20)	**23.093 (10.201)** $n = 431$	16.596 (13.442) $n = 431$	5.888 (13.426) $n = 185$	6.050 (16.675) $n = 190$
6. >30 hours (vs. <30)	–4.609 (11.218) $n = 431$	–5.598 (12.868) $n = 431$	5.307 (12.387) $n = 183$	–0.450 (14.424) $n = 188$
7. >40 hours (vs. <40)	0.156 (11.992) $n = 416$	–3.913 (11.982) $n = 416$	2.689 (12.123) $n = 183$	–14.796 (15.035) $n = 188$
2005–2006 school year				
1. # SES hours attended (OLS)	–0.005 (0.013) $n = 1075$	–0.010 (0.040) $n = 1016$	0.007 (0.011) $n = 2215$	**0.087 (0.042)** $n = 2200$
SES hours attended (matching)				
2. at least 20 hours (vs. none)	–0.055 (0.441) $n = 366$	0.117 (1.672) $n = 343$	0.246 (0.504) $n = 626$	0.328 (1.992) $n = 595$
3. at least 40 hours (vs. none)	0.175 (0.814) $n = 216$	0.547 (2.694) $n = 200$	–1.832 (1.094) $n = 346$	–5.220 (3.546) $n = 333$
If attended > 0 hours:				
4. >10 hours (vs. <10)	1.226 (0.672) $n = 304$	0.721 (3.523) $n = 282$	1.008 (0.556) $n = 505$	2.722 (3.364) $n = 468$
5. >20 hours (vs. <20)	–0.697 (0.656) $n = 304$	–0.762 (2.211) $n = 282$	1.402 (0.715) $n = 508$	0.519 (2.295) $n = 482$
6. >30 hours (vs. <30)	1.086 (0.845) $n = 307$	–2.694 (2.307) $n = 285$	0.787 (0.593) $n = 505$	–1.419 (1.947) $n = 468$
7. >40 hours (vs. <40)	0.403 (0.705) $n = 293$	0.479 (2.284) $n = 271$	–1.116 (0.872) $n = 470$	–4.534 (3.350) $n = 421$
2006–2007 school year				
1. # SES hours attended (OLS)	0.023 (0.190) $n = 462$	–0.045 (0.222) $n = 464$	n.a.	n.a.
SES hours attended (matching)			n.a.	n.a.
2. at least 20 hours (vs. none)	0.310 (7.381) $n = 370$	–4.178 (6.356) $n = 379$	n.a.	n.a.
3. at least 40 hours (vs. none)	–11.199 (22. 639) $n = 285$	–5.820 (11.812) $n = 289$	n.a.	n.a.
If attended > 0 hours:			n.a.	n.a.
4. >10 hours (vs. <10)	2.359 (8.719) $n = 462$	–6.606 (6.876) $n = 463$	n.a.	n.a.
5. >20 hours (vs. <20)	–2.501 (7.072) $n = 462$	–5.539 (6.893) $n = 463$	n.a.	n.a.
6. >30 hours (vs. <30)	0.991 (12.573) $n = 462$	–4.974 (9.849) $n = 464$	n.a.	n.a.
7. >40 hours (vs. <40)	0.032 (35.163) $n = 462$	10.368 (22.727) $n = 464$	n.a.	n.a.

Note. Standard errors in parentheses; results statistically significant at $\alpha \leq .05$ shown in bold. OLS = ordinary least squares.

that selection is on observable characteristics. The results of these matching analyses are summarized in lines 4 through 7 of Table 6. Among all of these estimates, there is only one statistically significant finding for the group of middle school students in 2004–2005 receiving at least 20 hours of SES (compared to fewer than 20 hours). A final set of matching analyses conducted separately by grade for Grades 8 through 12 (but not reported here) likewise did not change the predominant finding of no statistically significant effects of SES attendance on students' math and reading achievement.

Recall that the fixed-effects or "double difference" model (from Equation 3 above) produces separate, unbiased estimates of the estimated effects of any SES participation on math and reading achievement (given the assumptions of the model). The results of these models are shown in Tables 7 (2004–2006) and 8 (2005–2007). The top part of Table 7 shows the mean, variance, and reliability of student achievement in Grades 8, 9, and 10 for the period 2004–2006 and the multipliers used to produce equal test variances for all three tests. Table 7 also presents the parameter estimates and standard errors of primary interest in this study, that is, the coefficients of the SES indicators in the fixed-effects models for Grades 8 and 9. As indicated in the table, the SES estimates are not statistically different from zero. The middle part of Table 8 shows the means and standard deviations for reading and math achievement for three grades for each of the three middle school cohorts for the period 2005–2007. Table 8 also presents parameter estimates and standard errors of the coefficients of the SES indicators in the fixed-effects model for these cohorts, along with a single pooled estimate that optimally combines the information from the separate cohorts. These results provide no evidence of statistically significant estimated SES effects. It is clear that these results are consistent with those of the propensity score matching models, showing no statistically significant estimated effects of SES participation on students' math or reading achievement at any grade level. In light of a growing consensus among scholars (Lockwood & McCaffrey, 2007; McCaffrey, Han, & Lockwood, 2008; Meyer & Christian, 2008) that, if basic model assumptions are met, fixed-effects models produce estimates with relatively less bias and noise (compared, for example, to OLS models that predict test scores using students' baseline achievement scores and levels, such as in the Chicago and Los Angeles studies), we view the fixed-effects model results as providing strong, suggestive support for our general finding of no estimated effects of any SES participation on MPS middle and high school students' achievement.

Qualitative Study Findings on SES Effectiveness

The qualitative components of this study—student surveys, observations of SES providers, and interviews with SES provider and school district staff—yielded valuable information and insights for understanding the empirical findings of no SES effects on student achievement. One obvious problem is that students are not attending a sufficient number of SES hours, with less than 1% of middle and high school students attending for at least 40 hours in the 2006–2007 school year. Perhaps of greater concern, though, is what students are doing in the SES sessions that they attend. Data collected in the MPS student surveys showed that they attend an average of 2.4 SES sessions per week of an average length of slightly more than 60 minutes. In a set of five questions that asked students how they spent their time in SES sessions—the number of minutes working one-on-one with a tutor, working on their own (e.g., on homework or in other self-directed activities), in group activities led by a tutor, in group activities without a tutor, and socializing or other nonacademic uses of time—they reported spending the most time in self-directed activities (30 minutes on average). In their observations of instruction in SES sessions, Burch, Good, et al. (2007) and Burch and Good (2009) saw that many of these activities take the form of "more school"; that is, they consist of a lot of desk time and worksheets, with few opportunities for richer activity-based programming and other such activities. In one example reported by Burch and Good (2009), they describe how "students exclusively sat at desks in every tutoring session observed over two years, and the only exceptions to packet or

TABLE 7
Fixed-Effects Model Results for Grade 8–9–10 Cohort, 2004–2006

	Achievement Grade 9 – 8	Achievement Grade 10 – 9	Fixed effects (double difference)
Reading			
SES 8th grade	–1.58		–2.35
(standard error)	(2.31)		(3.38)
SES 9th grade		0.33	0.27
(standard error)		(2.03)	(3.41)
Sample size: 4300			
Math			
SES 8th grade	–2.67		–5.91
(standard error)	(2.32)		(3.53)
SES 9th grade		–0.47	–4.13
(standard error)		(2.25)	(3.59)
Sample size: 4228			

Note. SES = supplemental educational services.

TABLE 8
Pooled Fixed-Effects Model Results for Middle School Cohorts, 2005–2007

	Reading			Math		
Grade cohort	Estimate	Standard error	Sample size	Estimate	Standard error	Sample size
5 & 6	5.10	(7.93)	4679	–1.12	(6.64)	4680
6 & 7	–1.02	(4.38)	4892	–0.47	(3.89)	4883
7 & 8	–1.80	(4.99)	3848	–5.66	(4.40)	3864
All middle school grades	–0.41	(2.19)		–2.49	(1.94)	

worksheet-based activities were occasional class discussions, a Jeopardy-type review game, and an activity where students did math questions related to a documentary" (p. 8).

Another problem identified by Burch and Good through their observational study is that even in SES sessions where the level of student engagement is high, the work in which they are engaged tends to be disconnected from the school curriculum, contributing to a lack of continuity in their daytime and after-school learning environments. This finding was consistent with MPS student responses to a set of survey questions that inquired about subject areas where they needed help and whether the content and activities of SES sessions helped them to improve in these areas. For example, slightly more than a quarter of students who reported needing help on reading worked on their reading, and less than 10% of students reported writing an essay or paragraph during a SES session. Overall, less than 30% of these MPS students responded that the SES sessions had been very helpful to them in improving their performance in school. Planned curriculum frequently does not match curriculum in use, and direct connections to the school day curriculum are more likely to happen by chance, such as in the case where a tutor is also a teacher in the child's school. As one school principal expressed in an interview (Burch, Good, et al., 2007), "This is my problem with the whole SES program. . . . How can they expect to move children academically if you don't look at the curriculum that they're being taught during the day and connect that with what they're being taught at nighttime during tutoring? It's gonna be . . . a worksheet heaven" (p. 16).

Although the SES providers studied by Burch and Good all had some level of individualization in their tutoring sessions, and two-thirds of MPS students reported receiving at least some one-on-one tutoring in their SES sessions, Burch and Good also observed a high degree of variability

in the frequency and type of interactions between SES instructors and students, ranging from simply being available for students' questions or online "chatting" to side-by-side direction of the students' work (Burch, Good, et al., 2007). The tutors observed also differed considerably in their credentials and experience, and yet as one school district SES coordinator explained, SES providers are paid for each hour of SES that students attend, regardless of what students do in the sessions or what they gain. In other words, there is little in the way of "quality control" at the school district level, that is, beyond the process of SES provider certification at the state level.

SES providers, for their part, may have a legitimate grievance in stating that inconsistent or waning student attendance makes it difficult for them to help students advance and to improve the effectiveness of instruction. At the same time, more than a quarter of students reported in the MPS survey that they were not learning anything in the sessions, and thus, it is possible that if SES instruction were more effective, better attendance would follow. Student responses to the survey question asking what would encourage them to attend more sessions were mixed, though, and the most common response suggested more food, prizes, games, or rewards, rather than improvements in academic content.

Discussion and Conclusion

Any new program requires time to work out the early implementation challenges and to settle on an effective service delivery model. Although after-school tutoring was not a novel intervention itself, the context in which SES programs were developed and administered under NCLB was new for state and local educational agencies and their contracted providers. As Sunderman and Kim (2004) explained, school districts were required to translate complex provisions of NCLB into viable programs that did not conflict with existing policies, while relying heavily on the private sector to supply the core services and present adequate competition and choice. This demanded the support of school staff and teachers, parents, and community members at a time when the opportunity costs of Title I funds were high. In addition, NCLB requires state and local educational agencies to assess SES providers' effectiveness in increasing student achievement and to disseminate this information to parents of children who are eligible to receive services, and our study shows just how difficult this is to do, that is, to identify with confidence the effectiveness of SES vendors.

Six years since NCLB mandated the provision of SES, what do we know about its implementation and effectiveness? Although this study focuses on a single, urban school district, our findings are generally consistent with the growing body of field research and related studies that are investigating SES programs in medium- to large-size urban districts. First, although the typical demands of outreach and implementation in a new program might have accounted for the low *initial* take-up of SES, our research suggests that the number of eligible students registering for SES has leveled off and that attendance among those who registered is declining over time, particularly among older students. And although it is positive to find that MPS students eligible for free lunch (i.e., from lower income families) are significantly more likely to sign up (among those eligible for SES), the empirical evidence suggesting that free lunch eligible students are significantly less likely to attend after registering is discouraging. The empirical analyses also suggested that students with more absences or who had been retained were less likely to register and/or attend SES. In effect, students who are more likely to have higher levels of academic need for SES may be missing or declining the opportunity to receive the extra tutoring and individual help that NCLB intended to provide.

In inquiring about what would encourage students to attend more SES, the students indicated that they are primarily responding to incentives and prizes (i.e., computers, gift cards, fun and games, food and candy). The focus group findings suggested that parents had very different concerns in choosing participation or a SES provider for their children, for example, student–teacher ratios, tutor qualifications, how much one-on-one tutoring their child would receive, and specific academic content. Yet, they also reported that they rarely had this type of information available to use in deciding what was best for their child. Most of the information that is currently available on SES programs is supplied by the vendors,

and as Burch, Steinberg, and Donovan (2007) show, the limited capacity of state and local educational agencies for monitoring provider activities and performance results in little more than "lip service" to accountability requirements.

In Milwaukee, SES vendors' most successful tool in recruiting students to register and participate in SES was diluted in the 2006–2007 school year with a new policy that limits the use of incentives to encourage attendance to those deemed educational (e.g., books, educational software, magazines, museum field trips, etc.). The policy explicitly prohibits vendors from offering more popular incentives such as iPods, mall gift cards, movie passes, and pizza parties (see http://dpi.wi.gov/esea/pdf/ses_incentives_policy.pdf). Although we are not able to establish a causal relationship, we suggest that it is highly likely that there is some link between the dramatic drop-off in student attendance at SES programs in the 2006–2007 school year (down to 34% from 64% in 2005–2006, as shown in Table 1) and these new restrictions on incentives. Alternatively, based on these student reports and the empirical analyses in this study that produced little evidence of the potential effectiveness of SES in increasing student achievement, one might also speculate that parents and students are, in fact, choosing rationally in not registering for or attending SES. Using propensity score matching techniques and fixed-effects models to adjust for student selection into SES, we failed to find any statistically significant *average* effects of SES on student math and reading (test score) gains. And although one statistically significant effect of total SES hours attended (or treatment intensity) was found for high school students in 2005–2006 (on reading gains), the effect was substantively small (relative to total variation in reading score changes), and none of the other seven subgroup estimates was close to statistical significance.

Our dismal conclusion does not necessarily imply that SES programs should be discontinued in Milwaukee or elsewhere or that SES could never be effective for some students. Other studies discussed in this article find some effects of SES on elementary school students' achievement. With colleagues at the Wisconsin Center for Research, we are also continuing with research that focuses on getting "inside the black box" to better understand why SES programs are not currently more effective in four additional sites, and state and local educational agencies are also eager to increase their understanding of how these programs' effectiveness might be improved through policy and market governance changes, such as the use of performance-based contracting. The federal government should also consider granting a waiver to allow for an experimental evaluation of the effectiveness of SES if provision of these services continues to be mandated. This may be the only way we can surmount some of the important limitations faced in this research, including not having complete test data for measuring student achievement in each school year and the possibility that we may not be adequately controlling for student selection into SES registration and attendance. And of course, this study is based on research from a single urban school district, and although this poorer and predominantly minority population of SES-eligible students is very comparable to the larger national population of eligible students, the cross-state and -district variation in SES provider markets and program administration that is relevant to program outcomes might very well limit the wider applicability of these study findings.

Notes

[1]White House press release: http://www.whitehouse.gov/news/releases/2006/10/20061005-6.html.

[2]Students attending Title I schools identified for improvement are given the option to transfer to another public school or to receive supplemental educational services, depending on the eligibility of the student and the status of the school. These options continue until the school has made adequate yearly progress (AYP) for 2 consecutive years. A district must set aside an amount equal to 20% of its Title I allocation to fund both SES and transportation for students who elect to attend other schools under school choice. This set-aside cannot be spent on administrative costs for these activities, and the district may reallocate any unused set-aside funds to other Title I activities after ensuring that all eligible students have had adequate time to opt to transfer to another school or apply for SES.

[3]Farkas and Durham (2006) add that these provider self-evaluations are too weak in their research design and methodology to be viewed as reliable.

[4]The 30-hour cutoff was applied because it is the fewest number of hours that SES providers were approved to offer CPS students.

[5]Of the 17 focus group attendees, 13 were female 10 were Black, and 4 were Hispanic; 12 had completed their high school degree or GED; 6 were single and never married; and 9 had children who had attended SES in the prior school year.

[6]Burch and Good continued the field research in the 2007–2008 and 2008–2009 school years, but as those observations do not correspond with the quantitative data used in this study, we do not incorporate them here.

[7]The 2004 state assessment and the TerraNova assessments were both developed by CTB/McGraw-Hill and scored on the same developmental scale.

[8]Note that gains in student achievement can be computed only for students who were enrolled in the district (and tested) in subsequent school years.

[9]We experimented with a model in which the pretest variable was included as an explanatory variable (with appropriate corrections for measurement error in this variable). This analysis confirmed that it was legitimate to impose the restrictions implied by a gain model.

[10]AGS is a company that publishes student assessment and curriculum materials, with a particular focus on supporting students who are at risk or performing below grade level.

[11]In future work, we will explore using grades received in mathematics, English, and other courses as student outcomes possibly affected by participation in SES. The primary advantage of these student outcomes is that they are available for all students.

[12]We conjecture that participation in SES affects achievement growth primarily during the period in which a student participates in SES. If the effects of SES participation persist over the entire school year, then the measured effects of SES may be smaller in models based on the AGS assessments than in models based on the other tests, because the AGS tests do not span the entire school year. We assume (as is conventional in most achievement growth models) that SES participation affects achievement growth only during the school year (or years) in which a student participates in SES.

[13]NCLB also requires school districts to give priority to the lowest achieving eligible students if sufficient funds are not available to serve all those eligible for SES; at least in the past, this has not been a concern in many school districts due to low participation rates [Section 1116(b)(10)(C)].

[14]It is important that the factors included in any first-stage model are observed prior to the intervention or are measures of characteristics that are stable or deterministic with respect to time, such as demographic characteristics.

[15]Radius caliper matching uses not only the nearest neighbor within each caliper but also all of the comparison group members within the caliper.

[16]We do not report the statistically significant school indicators.

[17]Chow tests in models predicting total SES hours attended rejected functional/structural differences across schools.

[18]The proportion of variation in the logistic regressions predicting SES registration and attendance was calculated using a pseudo-R^2 measure. These values for the registration models are shown in Table 3; the highest pseudo-R^2 for attendance was 23.6%.

[19]For example, for high school students in 2005–2006, predicted probabilities of SES registration ranged from 0.02 to 0.77 for students who did not attend any SES and 0.04 to 0.75 for those who attended; for middle school students, these same probability ranges were 0.02 to 0.80 and 0.04 to 0.81, respectively.

[20]In the 2006–2007 school year, only 10th graders were tested in high school, and only 80 10th graders were registered for SES and tested; thus, we do not report SES effects for high school students in this year.

[21]Because some students transfer from one SES provider to another (as many as four times in 1 year), these variables record the hours and SES sessions of each student with each vendor.

References

Abadie, A., & Imbens, G. W. (2006). *On the failure of the bootstrap for matching estimators* (NBER Technical Working Paper 325).

Agodini, R., & Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention programs. *Review of Economics and Statistics*, *86*(1), 180–194.

Anderson, L. M., & Laguarda, K. G. (2005). *Case studies of supplemental services under the No Child Left Behind Act: Findings from 2003–04*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

Black, A. R., Doolittle, F., Zhu, P., Unterman, R., Grossman, J. B., & Warner, E. (2008). The evaluation of enhanced academic instruction in after-school programs: Findings after the first year of implementation. *U.S. department of Education National Center for Education Evaluation Report,* 2008-4021.

Burch, P. (2007). *Supplemental educational services under NCLB: Emerging evidence and policy*

issues. Boulder: Educational Policy Research Unit, University of Colorado.

Burch, P. (2009). *Hidden markets: The new educational privatization* (The Critical Social Thought Series). New York: Routledge.

Burch, P., & Good, A. (2009). *Getting to the core: The role of instructional setting in federally mandated tutoring*. Working paper presented at the 2009 American Educational Research Association Conference, San Diego, CA.

Burch, P., Good, A., Heinrich, C. J., Meyer, R. H., & Whitten, G. (2007, August). *Summary of preliminary findings: Milwaukee Public Schools supplemental education services evaluation*. Report to Milwaukee Public Schools, Wisconsin Center for Education Research.

Burch, P., Steinberg, M., & Donovan, J. (2007). Supplemental educational services and NCLB: Policy assumptions, market practices, emerging issues. *Educational Evaluation and Policy Analysis*, *29*, 115–133.

Casserly, M. (2004). Choice and supplemental services in America's great city schools. In F. M. Hess & C. E. Finn (Eds.), *Leaving no child behind? Options for kids in failing schools* (pp. 191–212). New York: Palgrave Macmillan.

Chatterji, M., Kwon, Y. A., & Sng, C. (2006). Gathering evidence on an after-school supplemental instruction program: Design challenges and early findings in light of NCLB. *Education Policy Analysis Archives*, *14*(12), 1–44.

Chicago Public Schools. (2007). *The 2007 supplemental educational services program: Year 4 summative evaluation*. Chicago: Chicago Public Schools Office of Extended Learning Opportunities, Research, Evaluation and Accountability.

Dynarski, M., James-Burdumy, S., Moore, M., Rosenberg, L., Deke, J., & Mansfield, W. (2004). *When schools stay open late: The national evaluation of the 21st Century Community Learning Centers Program: New findings*. Washington, DC: U.S. Department of Education.

Education Industry Association. (2007, August 24). *Supplemental education services (SES) and the 2007–2008 school year: The "No Child Left Behind" program can make a difference for more than 3 million students nationwide*. Rockville, MD: Author.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*(4), 605–619.

Farkas, G., & Durham, R. (2006). *The role of tutoring in standards-based reform* (Working paper). Department of Sociology and Population Research Institute, The Pennsylvania State University).

Government Accountability Office. (2006). *No Child Left Behind Act: Education actions needed to improve local implementation and state evaluation of supplemental educational services* (GAO Report 06-758). Washington, DC: Author.

Heckman, J. J., Ichimura, H., Todd, P., (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies,* 65(2), 261–294.

Heckman, J. J., Lalonde, R. J., & Smith, J. A. (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (pp. 1865–2097). New York: Elsevier.

Heinrich, C. J., & Whitten, G. W. (2006). *Supplemental education services (SES) in Milwaukee Public Schools (MPS): Focus group report*. Madison: Wisconsin Center for Education Research.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Athorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*, 275–313.

Lockwood, J. R., & McCaffrey, D. F. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics*, *1*(1), 223–252.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*, 423–458.

McCaffrey, D. F., Han, B., & Lockwood, J. R. (2008, February 28–29). *From data to bonuses: A case study of the issues related to awarding teachers pay on the basis of their students' progress*. Prepared for the conference on "Performance Incentives: Their Growing Impact on American K–12 Education," Nashville, TN.

Meyer, R. H., & Christian, M. (2008, February 28–29). *Value-added and other methods for measuring school performance*. Prepared for the conference on "Performance Incentives: Their Growing Impact on American K–12 Education," Nashville, TN.

Mueser, P. R., Troske, K. R., & Gorislavsky, A. (2007). Using state administrative data to measure program performance. *Review of Economics and Statistics*, *89*(4), 761–783.

Padilla, C., & Lopez-Torkos, A. (2006, April 7–11). *Early implementation of NCLB Title I choice and supplemental services provisions*. Center for Education Policy, SRI International. Paper presented at the 2006 annual meeting of

the American Educational Research Association, San Francisco, CA.

Peterson, P. (2005). Making up the rules as you play the game. *Education Next*, *5*(4), 42–49.

Rickles, J. H., & Barnhart, M. K. (2007). *The impact of supplemental educational services participation on student achievement: 2005–06* (Publication No. 352). Los Angeles, CA: Los Angeles Unified School District Program Evaluation and Research Branch, Planning, Assessment and Research Division

Ryan, S., & Fatani, S. (2005). *SES tutoring programs: An evaluation of the second year—Part one of a two part report* (Policy report). Chicago: Office of Research, Evaluation and Accountability, Chicago Public Schools.

Steinberg, M. S. (2006). Private educational services: Whom does the market leave behind? *PolicyMatters*, *4*(1), 17–22.

Sunderman, G. M., & Kim, J. (2004). *Expansion of federal power in American education: Federal-state relationships under the No Child Left Behind Act, year one*. Cambridge, MA: The Civil Rights Project at Harvard University.

Vandell, D. L., Reisner, E. R., Brown, B. B., Dadisman, K., Pierce, K. M., Lee, D., & Pechman, E. M. (2005). *The study of promising after-school programs: Examination of intermediate outcomes in year 2*. Madison: Wisconsin Center for Education Research.

Vegari, S. (2007). Federalism and market-based education policy: The supplemental educational services mandate. *American Journal of Education*, *113*, 311–339.

Wilde, E. T., & Hollister, R. (2007). How close is close enough? Evaluating propensity score matching using data from a class size reduction experiment. *Journal of Policy Analysis and Management*, *26*(3), 455–478.

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 educational services, and student achievement*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

Authors

CAROLYN J. HEINRICH is a professor and the director of the La Follette School of Public Affairs and an affiliated professor of economics at the University of Wisconsin-Madison, 1225 Observatory Drive, Madison, WI 53706; cheinrich@lafollette.wisc.edu. Her research focuses on human capital development, social welfare policy, public management, and program evaluation.

ROBERT H. MEYER is a research professor at the Wisconsin Center for Education Research (WCER) and the director of the Value-Added Research Center (VARC), 1025 West Johnson Street, Room 769, Madison, WI 53706; rhmeyer@wisc.edu. He is known for his research on value-added modeling and innovative statistical methods for education studies.

GREG WHITTEN, formerly at the Value-Added Research Center (VARC) of the University of Wisconsin-Madison, is a Ph.D. candidate at the University of Pittsburgh's Department of Economics, WW Posvar Hall, Pittsburgh, PA 15260; gww2@pitt.edu. In addition to his work on supplemental educational services at VARC, he developed processes for analyzing and reporting survey data on teacher practices and small-classroom management.