

**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  
Wisconsin Center for Education Research

March 2008

*Preliminary: please do not quote or cite without permission.*

This project was supported with funding from the Milwaukee Public Schools. We 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. We would also like to thank our colleagues at the University of Wisconsin: Patricia Burch, Huiping Emily Cheng, Annalee Good, John Keltz, Sarah Mason, Ernie Morgan, and Larry Schultz.

*“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. ... 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.”*

President George W. Bush, October 5, 2006, discussing supplemental educational services available under the *No Child Left Behind Act* at Woodridge Elementary and Middle Campus, Washington, D.C.<sup>1</sup>

## **Introduction**

In the White House proposal for the No Child Left Behind (NCLB) Act, 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 <http://www.whitehouse.gov/news/reports/no-child-left-behind.html>. 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 three years are required to offer parents of children in low-income families the opportunity to receive extra academic assistance.<sup>2</sup> These services typically include tutoring and remediation in reading/language arts and mathematics and are provided outside of the regular school day. According to the law, the content and educational practices of SES should be aligned with the state’s academic content standards and based on high-quality research evidence of their effectiveness in increasing student academic achievement [Section 1116(e)(12)(C)].

---

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

<sup>2</sup> 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 AYP for two consecutive years. A district must set aside an amount equal to 20 percent 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 all eligible students have had adequate time to opt to transfer to another school or apply for SES.

In arranging for SES, state educational agencies are required to identify organizations that provide services consistent with state and local instructional programs (and applicable federal, state, and local health, safety, and civil rights laws) [Section 1116(e)(12)(B)(i)] and that have an established record of effectiveness in increasing student academic achievement. These may include any public and/or private (non-profit 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. As research to date has shown, however, there is very little information available on the effectiveness of different organizations that are entering the market to provide SES, beyond some internal performance evaluations conducted by the larger national providers (Burch, Steinberg and Donovan, 2007).<sup>3</sup> Correspondingly, states and school districts are facing considerable challenges in assessing provider effectiveness, both before and after entering agreements with SES providers, which has important implications for the efficient functioning of the SES market and for the accountability goals of NCLB. In fact, the law requires states to withdraw approval from providers that fail to increase student academic achievement for two years, but without adequate data, methods or expertise to assess the contributions of SES to students' academic outcomes, compliance with this provision is difficult to realize in practice.

This research was motivated largely by concerns of state and school district officials in Wisconsin who are charged with meeting NCLB requirements to offer SES to students in schools identified for improvement (SIFI) and to measure their effectiveness in increasing student achievement. In Wisconsin, schools in Milwaukee account for the vast majority of SIFI schools, ranging from 89 percent of all SIFI schools in 2005-2006 to 71 percent last year (81 percent on

---

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

average from 2002-2007). As the only Wisconsin school district that has been required to offer SES to date, Milwaukee Public Schools (MPS) is the site of this research. Using administrative data from MPS, documentation on SES providers, and data collected in student surveys and parent focus groups, we undertake analyses to explore the evolving SES provider market in Milwaukee, the effectiveness of SES in increasing student achievement, and the relationship between SES provider effects and their market shares in Milwaukee. In particular, we address the following key questions in investigating how SES operates in Milwaukee: 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, can we 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 impacts of SES attendance on student achievement gains and just a few small effects of the number of SES hours attended (total hours and hours with specific providers).

### **What (little) we know about SES**

For some of the 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. First, the onset of

SES was slowed by the NCLB stipulation that only schools which had not made adequate yearly progress for three consecutive years were required to offer SES. In the 2002-03 school year, less than 100,000 students nationwide participated in SES, although participation rates subsequently increased from 12 percent of eligible students receiving services in 2003-2004 to 19 percent in 2004-2005 (Peterson, 2005; Government Accountability Office, 2006).

In addition, state and local educational agencies were charged with demanding new responsibilities for initiating services and ensuring compliance with NCLB specifications. These include establishing processes for identifying and approving providers and encouraging their participation to expand choice for parents and students; developing, implementing, and publicly reporting on standards and methods for monitoring SES quality and provider effectiveness, and designing procedures for withdrawing approval from unsuccessful providers.<sup>4</sup> For many state and local educational agencies, the development of these procedures for coordinating and evaluating SES is still a work in progress. The Government Accountability Office (GAO, 2006) reported that three-fourths of states are experiencing challenges in designing methods to assess academic progress among students and in developing data systems for tracking and analyzing SES information. The GAO observed that at the time of its survey (completed in July 2006), no state had produced a report presenting a conclusive assessment of SES providers' effects on student academic achievement.

Independent research on SES up to now has been 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 (Burch, Steinberg, and Donovan, 2007; Vegari, 2007; Farkas and Durham, 2006; Padilla and Lopez-

---

<sup>4</sup> NCLB requires the withdrawal of providers that fail, for two consecutive years, to contribute to increasing the academic proficiency of the students they serve.

Torkos, 2006; Steinberg, 2006; Anderson and Laguarda, 2005; Casserly, 2004; Sunderman and Kim, 2004). These researchers have documented the large and growing number of diverse organizations that have entered the market to compete for available SES funds (up to 20 percent of \$12.7 billion in 2005-06, or over \$2.5 billion) and the dominance of the larger, national for-profit providers (Burch, Steinberg, and Donovan, 2007; Farkas and Durham, 2006). They describe providers' widely varying hourly charges for SES and attempt, primarily through case studies, to examine relationships among service costs, tutor qualifications, tutoring session length, instructional strategies and curriculums, and recruiting and overhead costs (Burch, 2007; Burch, Steinberg, and Donovan, 2007; Farkas and Durham, 2006; Anderson and Laguarda, 2005). Although we will come back to some of the specific findings of these studies as we present results from our own analyses, these researchers have reported important problems and hurdles in implementing SES, including low student enrollment and attendance, curriculum clarity and alignment problems, lack of knowledge and communication among parents, providers and schools, inadequate monitoring and oversight of providers, and other problems related to market incentives and competition.

A handful (but growing number) of studies have sought to empirically estimate the effects of SES, although most have been fairly limited in scope and/or design. For example, Chatterji, Kwon and Sng (2006: 30) 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.” A Chicago Public Schools (CPS) study of students participating in SES in 2003-04 and 2004-05 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 and Fatani, 2005). The report also identified providers with higher than (district) average student reading scores. This same report was also criticized, however, for failing to adjust for differences between students who received SES and those who did not participate, implying that selection bias may have compromised the findings (Burch, 2007).

The most recent study of SES in CPS (Chicago Public Schools, 2007) focused on SES-eligible students in grades four to eight who were not English Language Learners, and importantly, who received at least 30 hours of SES tutoring.<sup>5</sup> The researchers estimated models to predict 2007 Illinois Standards Achievement Test (ISAT) reading and math scale scores, controlling for students' baseline achievement scale scores and levels, demographics, grade level and SES participation. They reported that SES tutoring in CPS increases 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. Model interactions suggested that students with disabilities gained more through SES participation, and that 6th and 7th graders and male participants benefited more from SES in their math achievement. Again, specific SES providers were identified as significantly more effective in producing gains in reading and math achievement.

Another recently released study (Rickles and 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 utilized SES (in 2005-06), and did SES affect their California Standards Test (CST) score gains? As in the Chicago study, the researchers estimated linear regression models

---

<sup>5</sup> The thirty hour cutoff was applied because it is the fewest number of hours that SES providers were approved to offer CPS students.

to predict the CST scores, controlling for students 2004-05 scores and relevant characteristics (gender, English language and special education status, participation in Gifted and Talented Education programs, free/reduced lunch receipt, grade retention, and parental educational attainment). Comparable to the prior year's evaluation results, the study authors reported low SES participation (approximately 7 percent 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.

A RAND study (Zimmer et al., 2007) is the only one we know of (to date) that has explored the effects of SES across multiple, geographically distinct school districts (Baltimore, Chicago, Long Beach, Los Angeles, Palm Beach, Philadelphia, San Diego and Washington DC). The study included samples of elementary, middle and high school students who participated in SES during one or more school years, 2002-03, 2003-04 and/or 2004-05 (in the early stages of implementation); consistent with other research, the RAND study reported the highest participation rates among elementary school students. To estimate the effects of SES on student achievement, researchers used a fixed-effects specification to compare changes in students' test scores before and after SES participation to 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. In light of observed differences in SES effects across grade and by intensity of participation (hours of SES) in other studies, one should be careful not to generalize broadly based on the RAND study results.

Burch (2007) also reviewed 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 to those receiving SES. These studies did not find positive effects of SES participation or significant differences among SES providers as assessed by changes in students' annual NALT reading scores.

Other after school study and tutoring (or "out-of-school-time") programs have long been in operation, of course, 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 afterschool

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 afterschool 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 careful review of this literature also reveals, however, that very few of these studies measure attendance or make the distinction between planned program duration and actual student attendance or engagement. While 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 (Lou et al., 1996; Elbaum et al., 2000) shows 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, while 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 among policymakers and researchers is that low-income and 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 has already been expressed in regard to SES, particularly in light of information asymmetries and procedural difficulties in disseminating information critical to informed choice, i.e., curriculum, costs, quality, effectiveness, etc. (Steinberg, 2006). In their study of parental decision making and charter school choice and quality in Texas, Hanushek et al. (2007) found that black students were significantly more likely to attend a charter school than students of other race or ethnic groups, and charter schools performed significantly worse on average than regular public schools in increasing student achievement. However, after accounting for student selection into charter schools (i.e., prior achievement levels and unobserved student differences), those attending charter schools made more rapid progress in their academic achievement than students who stayed in regular public schools. They also reported that despite the absence of publicly available information on school effectiveness, parents appeared to take into account “school value added” and school quality in making entry and exit decisions, with no significant differences by income in their responsiveness.

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 Milwaukee Public Schools.

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 and student surveys and secondary student-level demographics and achievement test data. The focus groups with parents of students enrolled in eligible SES schools for 2005-2006 were conducted early in our research 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 sample of 320 parents/students was first drawn from MPS administrative data, using zip codes to identify families as north side or south side residents and to select one half of the sample from each of these geographic areas. In addition, the sample was stratified to ensure that we included both parents whose children had registered for SES in the 2005-06 school year and those who chose not to participate in SES, with approximately 60% of the sample in each geographic area drawn randomly from parents whose children participated in SES. The focus groups took place in August 2006 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 percent 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 and Whitten, 2006 for further details).

The student surveys, administered in March and April of 2007 to those who had registered for SES in the 2006-2007 school year, were developed to collect additional

information from the students' perspective about how they were spending time in their tutoring sessions and how these sessions were affecting their performance in school. The surveys also asked students about how they chose their SES provider, if they had assistance in making their choice, and about what would encourage them to attend more tutoring 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?" In establishing the sample frame for the surveys, we selected eight SIFI schools with the largest SES enrollments. Students in seven of these eight schools completed the survey a second time in late March or early April if they were still participating in SES. A total of 1441 students responded to the first survey, and 874 students participated in the second survey. Due to requirements for protecting the confidentiality of student responses, no student identifying information was recorded, precluding the linking of data from the first survey with those of the second survey. Having two rounds of the survey, however, allowed for an analysis of the consistency of responses, particularly students' reports of activities with specific SES vendors.

With permission and assistance from Milwaukee Public Schools (MPS), we also secured access to their middle and high school level data that comes from the administration of standardized tests, administrative 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 (A+) 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. The range of years for which these data are available varies, however, as the system was not constructed to archive data. SES attendance data exists from the 2003-2004 school year forward, but the 2003-2004 data are incomplete.

To construct measures of student achievement, and particularly, gain in student achievement, we used data from standardized tests administered to MPS students. Currently, data are available for analysis for the 2004-2005 and 2005-2006 school years, and we expect to soon have data for the 2006-2007 school year. To evaluate the effect of SES during the 2004-2005 school year, we used reading and mathematics scores that were administered in November 2004 (prior to participation in SES) and in November 2005 (after participation in SES). In 2004 students were tested with the Wisconsin state assessment, the Wisconsin Knowledge and Concepts Examination (WKCE), in grades 8 and 10 and with the TerraNova assessment in grades 6, 7 and 9. In 2005 the state assessment was expanded from grades 4, 8, and 10 to grades 3-8 and 10, MPS assessed 9<sup>th</sup> grade students with the TerraNova. The 2004 state assessment and the TerraNova assessments were both developed by CTB/McGraw-Hill and scored on the same developmental scale. The 2005 state assessment was scored on a somewhat different scale than prior state assessments: scores were lower by several 100 points, but the variance in scores was very similar. Given that the variance in test scores of the 2004 and 2005 test scores were comparable at each grade, it is reasonable to measure growth in student achievement as the gain in student achievement *relative* to the average gain in achievement.<sup>6</sup> Including grade indicators in the model (the strategy used in this paper) also addresses the issue of differences in the location (means) of the 2004 and 2005 test scales. The key requirement is that the variance of the

---

<sup>6</sup> Note that gain in student achievement can be computed only for students who were enrolled in the district (and tested) in subsequent school years.

two scales are similar at each grade (although possibly different at different grades)<sup>7</sup> In 2005-2006, AGS tests 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 AGS tests were not used in any other school year. 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.

It is important to make clear the major differences between the assessment data available for these two school years, given their implications for the interpretation of SES effect estimates. First, the data for 2004-2005 covered SES participation only up through ninth grade, since the 10<sup>th</sup> grade test score (administered in November 2005) serves as a posttest outcome for 9<sup>th</sup> grade participation. In contrast, AGS scores were administered in all middle and high school grades. And while there are essentially no missing test data in 2004-2005 (since all students were required by the state or MPS to take the state assessment or TerraNova assessments in 2004-2005), a substantial number of middle and high school students did not take the AGS assessment in 2005-2006. Our statistical analysis of the patterns of missing test data in 2005-06 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 two years represent somewhat different high school grades and populations of students.<sup>8</sup>

---

<sup>7</sup> 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) (Fuller, 1987). This analysis confirmed that it was legitimate to impose the restrictions implied by a gain model.

<sup>8</sup> 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.

Another important difference between the assessments available in the two years is that the growth period is 12 months (November to November) for 2004-2005 and approximately 7 months (October through April) for 2005-2006. This difference is unlikely to affect our analysis since 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. It is reasonable to assume that the achievement growth of matched SES and non-SES students was similar during months in which SES services were not provided.

Finally, it is also the case that the units of the test scales in 2004-05 and 2005-06 differ substantially, as indicated in the chart below that reports the standard deviations in student gains for all middle school and all high school students. This information implies the following: (1) parameter estimates should generally be larger for analyses based on the 2004-2005 data than the 2005-2006 data (by as much as a factor of 10); (2) in 2004-2005, parameter estimates should generally be larger for high school students than middle school students (about double in size), and (3) in 2005-2006 parameter estimates based on reading scores should generally be larger (by a factor of three) than estimates based on math scores.

<b>Standard Deviation in Student Achievement Gain</b>	<b>2004-2005</b>		<b>2005-2006</b>	
	Mathematics	Reading	Mathematics	Reading
Level				
Middle School	50	50	6	17
High School	80	80	6	20

With the available administrative, transcript and test data, we constructed four separate datasets for each of four academic years, 2003-2004, 2004-2005, 2005-2006, and 2006-2007. However, we had sufficiently complete data for use in modeling the effects of SES on student achievement gains for only two of these years, 2004-2005 and 2005-2006. It was essential to have data from the prior school year to use in predicting SES registration and attendance (for the

purposes of controlling for student selection into SES), and since data are not fully available for 2002-2003, we do not estimate SES effects for the 2003-2004 school year. In addition, as indicated above, the test score data for 2006-2007 are still being prepared for analysis.

### **Estimation methods**

In a nonexperimental study, in which there is no random assignment process that generates statistically equivalent treatment and control groups, it is important to understand and model the process by which those who participate in the intervention (SES in this case) choose to receive services. For example, if only the more motivated students (or alternatively, those most in need of tutoring) take the steps required to register for SES, then a simple comparison of eligible students who register for SES with those who do not sign up would overlook important differences between the two groups of students and bias the estimates of SES effects.

In this study, we investigated selection into SES using multiple methods. As data are not systematically collected on how parents' and students' come to decisions to sign up for 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 to take up SES (or forego the option) (see Heinrich and Whitten, 2006 for details). We next used data from transcripts of SES-eligible students and administrative data systems to predict their probability of registering for SES and the probability of attending at least one SES session, conditional on registration, using logistic regression. The predicted probabilities generated by the logistic regression models subsequently serve a second important purpose—this information is used to remove possible bias associated with pre-intervention differences between students who registered for SES and those who did not register, allowing for more accurate identification of SES effects.

### *Propensity score matching*

More specifically, in attempting to identify the effects of SES in increasing students' reading and math achievement, we use propensity score matching methods to account for observed student selection into SES. The propensity score—which takes into account student characteristics and the public schools they attended<sup>9</sup>—reduces the matching problem to a single dimension, although we recognize that this approach will be inadequate in the face of any important, unobserved factors that influence both selection into SES and its effects on students. 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 impacts to place confidence in them, which they attributed to the role of unobservable characteristics. It is also worth noting, as the authors acknowledged, that only a limited number of covariates were available for use in their study, and some of their treatment subgroup sizes used in matching were very small.

The specific matching technique we apply in this study is radius caliper matching, with the caliper set at 0.01. In other words, students who did not attend SES are matched with students who attended if their predicted probabilities of registration are within 0.01 of each other. We also impose a “common support,” so that if there is not a match between students in these 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

---

<sup>9</sup> 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.

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 and Imbens, 2006), we also applied local linear matching methods, which yielded comparable results.

*Fixed effect model*

In addition, we also test an alternative, 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 8<sup>th</sup> grade ( $\eta_8$ ) and 9<sup>th</sup> grade ( $\eta_9$ ), respectively. Since student achievement is assessed in the fall, growth during 8<sup>th</sup> grade is captured by 9<sup>th</sup> grade achievement  $Y_{9i}$  minus 8<sup>th</sup> grade achievement  $Y_{8i}$ . Similarly, growth during 9<sup>th</sup> grade is captured by 10<sup>th</sup> grade  $Y_{10i}$  achievement minus 9<sup>th</sup> grade achievement  $Y_{9i}$ . In addition to the indicator for any SES attendance, the equations include controls for a vector of individual characteristics ( $X_i$ ), with coefficient vectors  $\beta_8$  and  $\beta_9$ , and regular school effects  $\alpha_8$  and  $\alpha_9$  (where  $S_{8i}$  and  $S_{9i}$  represent vectors of regular school indicators for 8<sup>th</sup> and 9<sup>th</sup> grade, respectively). We hypothesize that the error terms in each equation are composed of a fixed effect  $u_i$  and transitory errors  $e_{8i}$  and  $e_{9i}$ , respectively.

$$Y_{9i} - Y_{8i} = \zeta_8 + \eta_8 SES_{8i} + \beta_8' X_i + \alpha_8' S_{8i} + u_i + e_{8i} \quad (1)$$

$$Y_{10i} - Y_{9i} = \zeta_9 + \eta_9 SES_{9i} + \beta_9' X_i + \alpha_9' S_{9i} + u_i + e_{9i} \quad (2)$$

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 8<sup>th</sup> or 9<sup>th</sup> grade. To eliminate the fixed effect we can difference the two equations, yielding the following “double difference” model:

$$(Y_{10i} - Y_{9i}) - (Y_{9i} - Y_{8i}) = (\zeta_9 - \zeta_8) + \eta_9 SES_{9i} - \eta_8 SES_{8i} + (\beta_9 - \beta_8)' X_i + \alpha_9' S_{9i} - \alpha_8' S_{8i} + (e_{9i} + e_{8i}) \quad (3)$$

Equation (3) yields separate, unbiased estimates of the effect of any SES participation in 8<sup>th</sup> and 9<sup>th</sup> grade given the maintained model assumptions, although note that the coefficient on  $SES_{8i}$  is the negative of the 8<sup>th</sup> grade SES effect. Given that the fixed effect model will not yield precise SES estimates if the 8<sup>th</sup> and 9<sup>th</sup> grade SES variables are highly correlated, we have applied this model to students in 8<sup>th</sup> and 9<sup>th</sup> grade in the 2004-05 and 2005-06 school years. Very few of these students participated in SES in both grades, as very few students were enrolled in SES-eligible schools in both grades (one a middle school grade, the other a high school grade).<sup>10</sup>

One important requirement of the fixed effects model presented above is that the test scores in 8<sup>th</sup>, 9<sup>th</sup>, and 10<sup>th</sup> grade must be measured on the same scale. We explored several methods for rescaling the test scores in order to meet this requirement. Since 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.

We now present the empirical findings of this study, beginning with the analysis of who participates in SES in Milwaukee Public Schools.

### **Who signs up for SES?**

In his comments at Woodridge Elementary and Middle Campus in Washington, DC, 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 1 funds intended for SES in other ways. Peterson (2005:44) makes the same point more directly, explaining that school districts can divert SES dollars to their own uses by “suppressing parental demand” and

---

<sup>10</sup> One additional benefit of this sample is that the decision to participate in SES in 9<sup>th</sup> grade for most students was not influenced by information obtained from participating in the program in 8<sup>th</sup> grade.

discouraging student participation; in other words, “they have a clear financial disincentive” to enroll students in SES. School districts acting in their financial interest to limit SES participation would be in direct conflict, however, with the law, 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)].

Participation in SES among eligible students has, to date, been low nationwide. The GAO (2006) reported that low parent and student demand for SES was a challenge in approximately two-thirds of districts of the districts they studied. School districts suggest that it is difficult for SES to compete with other afterschool programs and extracurricular activities. In Illinois, where only 5 percent of the eligible student population was recently participating in SES, high school officials 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 and “iPods” to school supplies and gift certificates.

In assessing student participation in SES, the above discussion suggests that there are multiple stages of student/parent selection into services that should be investigated and/or modeled. First, school districts work with approved SES providers to disseminate information to students and parents about available services. Specifically, school districts are required to send a letter explaining the SES option to all families with eligible students (i.e., children from low-income families enrolled in public schools that have not made adequate yearly progress for three years). Most SES providers do their own marketing to parents and students as well, sending out

brochures, inviting parents to presentations, and offering incentives to students to register with them. Parents and students, sometimes with the guidance of teachers or other school staff, subsequently decide whether or not to take up SES; if so, they typically register with a specific provider at the same time. Following registration with a provider, students choose to attend SES sessions, and providers invoice the school district to request payment for the number of hours attended by each student.

Table 1 shows the number of middle and high school students who were eligible for SES, registered for SES, and attended any SES session during the 2003-04, 2004-05, 2005-06 and 2006-07 school years in Milwaukee Public Schools (MPS). Consistent with the findings of other field studies, the fraction of eligible students who register for SES is small, and among those, the number attending any sessions is declining substantially over time (from 90% to only one-third in the most recent academic year). Clearly, understanding why a relatively low number of eligible students progresses through registration and opts to 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. That is, prior research suggests that the potential for selective differences among students who progress to these different stages of the enrollment process is strong.

Perhaps one of the most important findings of the focus groups with MPS parents and students was that although considerable information is made available by schools and providers to inform parents of SES options for their children, not all parents receive or understand this information. Some 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. Many 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 on-line 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.*”

Interestingly, 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. Although some of this information was available in the district SES booklet, few 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 on-line 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).

In exploring empirically the factors influencing SES registration and attendance decisions, we estimated separate logistic regression models for middle and high school students for the 2004-2005, 2006-2006 and 2006-2007 school years. In the first set of models (see Table 2), the dependent (or outcome) variable is a binary variable equal to one if a student registered for SES and equal to zero if the student did not sign up. The predictor variables consist primarily of student demographics and special needs characteristics (gender, race/ ethnicity, free lunch eligibility, English proficiency and special education status), student school performance and attendance measures (GPA, number of absences from the prior school year, a retention indicator and foreign language course indicator), and controls for prior SES attendance, their grade year

and school attended. Because these data come from student records, no additional family composition or parental characteristics information is available for the analysis. Bearing in mind the focus group findings, we recognize that relevant information about 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 may bias the results.

The results of these analyses are presented (in summary form) in Table 2. Across time and middle and high school years, 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 did not (statistically) significantly influence middle or high school student decisions in that year.<sup>11</sup> If the variable was a statistically significant predictor, the increase in odds of registering for SES associated with that variable are 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 60-80% *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% to 65% lower odds of registering). Another consistent predictor that becomes stronger over time is free lunch eligibility; middle school students eligible for free lunch are significantly more likely to sign up for SES in all school years (with up to 161% higher odds), and this becomes evident for high school students as well in the 2006-2007 school year. This relationship is expected, given that NCLB requires school districts to use the same data to determine eligibility for supplemental educational services that they use

---

<sup>11</sup> We do not report the statistically significant school indicators.

for making within-district Title I allocations, which has historically consisted of free school lunch data.<sup>12</sup>

A strong and consistently negative predictor of SES registration is student absences from school (measured in the school year prior to the SES enrollment year). For each additional absence, both middle school and high school students have 0.7-1.1% lower odds of registering for SES. Given the findings from the field work component of this study showing that a majority of MPS providers' instructional strategies fall into the "*more school*" category (worksheets, packets, deskwork, homework, etc.), it is not surprising that those students who choose not to attend regular school also forego "*more school*" in the form of SES. In addition, high school students in the 12<sup>th</sup> grade are significantly less likely to register for SES than 9<sup>th</sup> grade students. Alternatively, an encouraging finding for SES program implementers is that students who attended SES the prior school year had substantially higher odds of registering for SES again, as high as 181% for high school students in the 2005-2006 school year.

Since the factors that influence student attendance at SES sessions (following registration) differ little from those affecting registration, we more briefly discuss the relatively sparse 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

---

<sup>12</sup> NCLB requires that if sufficient funds are not available to serve all eligible children, school districts must give priority to the lowest-achieving eligible students [Section 1116(b)(10)(C)]. School districts are encouraged to notify all eligible families of their children's eligibility, and only if more families request supplemental educational services than there are funds available to serve them, should the district set priorities or criteria to determine which eligible students can get services.

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 2004-05). Registered high school students who had been retained were also significantly less likely to attend any SES sessions.

In general, the findings of the SES registration and attendance models suggest that MPS is having some success in encouraging those eligible to register for SES, with free lunch students significantly more likely to register and those who have previously attended SES more likely to sign up again. At the same time, those students with poor attendance records are significantly less likely to register or attend, and free lunch eligible students and those who have been retained (among those who register) are also significantly less likely to attend any SES sessions. Clearly, some self-selection is taking place at multiple stages of these processes, and given that less than a quarter of the total variation in decisions to register or attend SES is explained in any of these models,<sup>13</sup> it is likely that there are some relevant unobserved or unmeasured factors (e.g., related to student motivation or barriers to participation) that are not captured in these models. For example, we know from the focus groups that the timing and location of the sessions and other activities that compete for students' after-school time influence these decisions, and student responses to the spring 2007 surveys indicated that many were encouraged to attend SES by the incentives and rewards providers offered (e.g., food, prizes, computers, etc.). An analysis using Chow tests for differences in coefficients across schools in the models predicting SES attendance suggested that the factors influencing *any* SES attendance likely differed across schools (although no specific factors were identified); alternatively, the Chow tests in models predicting total SES hours attended rejected functional/structural differences across schools.

---

<sup>13</sup> The proportion of variation explained was calculated using a pseudo-R<sup>2</sup> measure, reported in SAS and Stata. The lowest pseudo-R<sup>2</sup> value was 6.2% for high school attendance in 2004-05, and the highest pseudo-R<sup>2</sup> was for registration in 2004-05 (23.6%).

### **The effects of SES on student achievement**

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. Next we estimate the effects of differing levels of SES attendance (as measured in total hours attended) on student achievement. We then follow with the fixed-effects model estimation of the effects of any SES participation. Finally, we use ordinary least squares methods (assuming linear selection on observables) to examine the effects of specific SES providers on student achievement.

In performing the first set of analyses to estimate the effects of attending any SES sessions, we used the predicted probabilities (or propensity scores) from the models predicting student registration (shown in Table 2) to match students who registered SES with those who did not. These first-stage models and matching estimation of the effects of SES were performed separately for middle school and high school students in the 2004-05 and 2005-06 school years. To illustrate, Figures 1 and 2 show the propensity score distributions (predicted probabilities) for high school and middle school students in 2005-06. It is evident in these graphs that the range of predicted probabilities of registration are very similar for students who did not attend any SES sessions (the distribution on the left hand side) and those students who attended at least one SES session (the graphic on the right hand side). In addition, as expected, the distribution on the left is skewed more toward zero, reflecting the lower average probability of registration for students who did not attend. In general, the comparable ranges of these distributions imply that we can readily match (middle and high school) students who did not register for SES to similar students who did register in estimating the effects of SES (as the average difference in achievement test score *changes* between the matched pairs/groups of students).

### *Effects of any SES attendance*

Table 3 presents the estimated effects of attending any SES on changes in reading and math test scores in the 2004-05 and 2005-06 school years; both unmatched and matching estimates are shown. The results show quite clearly that after matching, there are no statistically significant differences in the changes in test scores for students who attended SES compared to 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 impact of *any* SES attendance on changes in student achievement is generally consistent with the limited impacts 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 impacts, 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 23 hours for high school students in 2005-06 and as high as 30 hours for middle school students in 2004-05. It is also important to note that vendors charge widely differing hourly rates for their services, with studies documenting rates as low as \$25 per hour to as high as \$80 per hour in the same district (Burch, Steinberg and Donovan, 2007; Steinberg, 2006). Since information on provider cost structures is not publicly available, it is difficult to assess the relationship between hourly rates charged and the quality of services provided. At the same time, since the per-student SES funding allocation is the same fixed dollar amount for all providers, SES vendors that charge higher hourly rates will necessarily provide fewer total hours of SES to their students. And if higher hourly rates reflect higher quality

services, then it may not be the case that a higher number of hours of SES attended will be linearly and positively related to increases in student achievement.

#### *Relationship between hours of SES attended and student achievement*

Figures 3a.-d. show the distributions of hours attended for middle and high school students in 2004-05 and 2005-06 (if they attended any hours); these data are based on provider reports of actual student hours attended that are required to receive payment for services.<sup>14</sup> The patterns are very similar across the four graphs, 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 and a meta-analysis of after-school programs suggested that effects of tutoring programs were larger for programs that were more than 40-45 hours in duration (albeit with effects diminishing for considerably longer hours). In the MPS samples, approximately 8 percent of middle school students and 17 percent of high school students attended 40 or more hours of SES in 2004-05; in 2005-06, the comparable numbers were 15 percent of middle school students and just 6 percent of high school students attending 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 2) and with a continuous measure of total hours of SES attended. The results, reported in line 1 of Table 4 for each of the four student subgroups, show only one statistically significant, positive effect of total

---

<sup>14</sup> Since some students transfer from one SES provider to another (as many as four times in one year), these variables record the hours and SES sessions of each student with each vendor.

hours attended on the change in high school student reading scores (in 2005-06); for each additional hour attended, students' reading test scores 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 percent 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 to those who registered and attended different levels of hours (eliminating concerns about selection into SES registration). In 2004-05, only a quarter of registered middle school students attended SES, while two-thirds of registered high school students attended some SES; in 2005-06, almost two-thirds of registered middle school students and just over one half of registered high school students attended at least one session. Lines 2 and 3 in Table 4 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 sixteen 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 vs. less than 10 hours; greater than 20 hours of SES vs. less than 20 hours; greater than 30 hours of SES vs. less than 30 hours, and the same for the 40

hours cut-off. In these analyses, we are no longer concerned about selection at the stage of registration 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 that selection is on observable characteristics. The results of these matching analyses are summarized in lines 4-7 of Table 4. In the 2004-05 school year, there is a fairly consistent pattern among the results suggesting that SES effects are positive and largest (on test score changes) for the group receiving at least 20 hours of SES (compared to less than 20 hours), although only the estimate for changes in middle school students' math scores is statistically significant. In the 2005-06 school year, there are no statistically significant relationships or clear patterns observed in these results. A final set of matching analyses conducted separately by grade (for grades 8-12) 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 effects of any SES participation in 8<sup>th</sup> and 9<sup>th</sup> grade on math and reading achievement, assuming that test scores are on a scale for which differencing scores across grades is reasonable. The results of these models are shown in Table 5. The top half of Table 5 shows the mean, variance, and reliability of student achievement in grades 8, 9, and 10 and the multipliers used to produce equal test variances for all three tests. The bottom half of Table 5 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. Clearly, these results are consistent with those of the propensity score matching models, showing no statistically significant effects of SES participation (in 8<sup>th</sup> or 9<sup>th</sup> grades) on students' math or reading

achievement. In light of a growing consensus among scholars (Lockwood and McCaffrey, 2007; McCaffrey et al., 2008; Meyer and 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-effect model results as providing strong, confirmatory support for our general finding of no effects of any SES participation on MPS middle and high school students' achievement.

#### *SES provider effectiveness*

In a final analysis of the effects of SES hours attended, we estimated OLS models in which the treatment was measured as the total number of SES hours attended with a specific provider. For each student who attended any SES, the total number of hours with the provider with which he or she attended SES for the most total time is measured with these variables; any vendor with at least 10 students is included in the model, and one additional "other provider" variable was created to combine the smaller vendors (and students' total time in SES with these providers). In other words, we are investigating the effects of students' total number of hours with particular providers to determine if some providers are, on average, more (or less) effective than others in increasing students' reading and math achievement (compared to the counterfactual of no SES participation).

Line 8 in Table 4 presents these results for each of the four subgroups of students; the sample sizes are the same as those reported in line 1. The results suggest that just six providers were effective in producing gain in students' math and/or reading achievement; they are labeled as vendors A-F in Table 4 to protect the identity of the providers while the observational component of the study is ongoing. Although the effect sizes appear to be fairly different across

the groups and years, they are congruent with expectations based on the differences in the tests discussed earlier; specifically, parameter estimates are larger for 2004-2005 than 2005-2006 by approximately a factor of 10, and parameter estimates for high school students are about double the size for those of middle school students. In addition, their size and substantive significance are generally comparable, that is, there is no one provider that stands out as being appreciably more effective than the others in increasing student achievement. Thus, we will not interpret each of these effect estimates, but as an example, we consider the effect of vendor D on high school students' reading gains in 2004-05. Specifically, for each additional hour of SES attended with NESI, student reading scores increased by 3.2 test units, or approximately 4 percent of a standard deviation reading gain for that test. Consistent with the finding discussed above for total hours attended, the estimated SES effects are quite small relative to the variability of gains in student test scores.

Nonetheless, we were interested in knowing if there are particular characteristics of these six SES providers that might explain why they were somewhat more effective than others in increasing student test scores. Table 6 presents some summary information about the providers for the 2004-05 school year, the only year in which all of these providers were active in MPS. Three of the providers are larger, for-profit national providers of SES (vendors D, E and F), and the others are nonprofit, locally-based providers. The information in this table suggests little in the way of consistent attributes or logical relationships among them, (i.e., hourly rates charged for tutoring, reported student-teacher ratios, number of sessions offered, or student attendance rates). For example, the provider charging the highest rate per hour (Vendor A) has the same student-teacher ratio and offers approximately the same number of sessions as one of the lowest-cost providers (Vendor C), and the less costly of these two providers is also more effective at

getting registered students to attend SES. In addition, the rankings for the percentage of students attending 20 or more hours of SES shows that only one of these providers (Vendor C) is among the top 10 in this category. Also contrary to what one might expect, the second highest-cost provider has the highest reported student-teacher ratio (as well as the second largest market share of all providers in 2004-05). In the current school year (2007-08), only two of these six providers are still offering SES in Milwaukee, suggesting there may also be little relationship between providers' effectiveness and their student market share.

### **Discussion and Conclusions**

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, and 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—i.e., to identify with confidence the effectiveness of alternative 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. First, while 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 low and 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 surveys of MPS students who had registered for SES, we inquired about what would encourage them to attend (or to go more frequently). The students clearly indicated that they are responding to incentives and prizes (i.e., computers, gift cards, fun and games, food and candy). Students also reported that they are spending close to one-third of their typical time in an SES session socializing or engaged in other non-academic activities. The focus group findings suggested that parents had very different concerns in choosing participation or a SES provider for their children, i.e., 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, Donovan and Steinberg (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](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. Other reasons reported by students for not attending SES included conflicts with other after school plans or activities, communication and transportation problems, and family issues. However, perhaps most importantly, a majority of students did not view attending SES as helpful to their performance in school. Less than 30 percent of MPS students surveyed in the spring of 2007 reported that the SES sessions had been helpful to them in improving their academic performance, and about a quarter of the students reported that they didn’t learn anything new in their SES sessions.

Based on these student reports and the empirical analyses in this study that produced little evidence of the effectiveness of SES in increasing student achievement, one might 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 identify any statistically significant *average* impacts 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 2005-06 (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. The analyses of the effectiveness of specific SES vendors, in which the treatment measure took into account the number of hours students attended SES with each vendor, did identify six providers that were effective in increasing students' math and/or reading achievement. However, these analyses were performed using OLS, a comparatively less rigorous approach to controlling for student selection, and thus, they should be interpreted and used with caution. In addition, the magnitude of these estimated effects was again substantively small, and there were no clear patterns among the characteristics of these vendors' programs that might suggest for parents what they should be looking for in choosing an effective SES provider.

Our dismal conclusion does not necessarily imply that SES programs should be discontinued in Milwaukee or elsewhere. Other studies discussed in this paper find some effects of SES on elementary school students' achievement. In ongoing research with colleagues at the Wisconsin Center for Research, we are attempting to get "inside the black box" to better understand why SES programs are not currently more effective, 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. In addition, we faced important limitations in this research. We do not have complete test data for measuring student achievement in each school year, and it is still possible that we may not be adequately controlling for student selection into SES registration and attendance. Furthermore, for one of the school years, we know that more disadvantaged students were less likely to take the reading

and math achievement tests, which also limits the generalizability of our results. 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 probably 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.

## References

- Abadie, Alberto and Guido W. Imbens. 2006. "On the failure of the bootstrap for matching estimators." NBER Technical Working Paper 325.
- Anderson, Leslie M. and Katrina G. Laguarda. 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.
- Burch, Patricia. 2007. "Supplemental Educational Services under NCLB: Emerging Evidence and Policy Issues." Boulder: Educational Policy Research Unit, University of Colorado.
- Burch, Patricia, Matthew Steinberg and Joseph Donovan. 2007. "Supplemental Educational Services and NCLB: Policy Assumptions, Market Practices, Emerging Issues." *Educational Evaluation and Policy Analysis* 29(2):115-133.
- Casserly, Michael. 2004. "Choice and Supplemental Services in America's Great City Schools." Chapter 9 in Hess, Frederick M. and Chester E. Finn (eds.), *Leaving No Child Behind?: Options for Kids in Failing Schools*. New York: Palgrave Macmillan, pp. 191-212.
- Chatterji, Madhabi, Young Ae Kwon and Clarice Sng. 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). Retrieved September 17, 2007 from <http://epaa.asu.edu/epaa/v14n12/>.
- Chicago Public Schools. 2007. "The 2007 Supplemental Educational Services Program: Year 4 Summative Evaluation." Chicago Public Schools Office of Extended Learning Opportunities, Research, Evaluation and Accountability.
- Dynarski, Mark, Susanne James-Burdumy, Mary Moore, Linda Rosenberg, John Deke and Wendy Mansfield. 2004. "When Schools Stay Open Late: The National Evaluation of the 21<sup>st</sup> Century Community Learning Centers Program: New Findings." Washington, D.C.: U.S. Department of Education.
- Elbaum, Batya, Sharon Vaughn, Marie T. Hughes, and Sally W. Moody. 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 – 19.
- Farkas, George and Rachel Durham. 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." Washington, DC: GAO Report 06-758.

Hanushek, Eric A., John F. Kain, Steven G. Rivkin and Gregory F. Branch. 2007. "Charter school quality and parental decision making with school choice." *Journal of Public Economics* 91(5-6): 823-848.

Heinrich, Carolyn J. and Gregory W. Whitten. 2006. "Supplemental Education Services (SES) in Milwaukee Public Schools (MPS): Focus Group Report." Madison, WI: Wisconsin Center for Education Research.

Lauer, Patricia A., Motoko Akiba, Stephanie B. Wilkerson, Helen S. Athorp, David Snow and Mya L. Martin-Glenn. 2006. "Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students." *Review of Educational Research* 76(2): 275-313.

Lockwood, J.R., and McCaffrey, Daniel 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., and d'Apollonia, S. 1996. "Within-Class Grouping: A Meta-Analysis." *Review of Educational Research* 66(4): 423-458.

McCaffrey, Daniel F., Bing Han and J.R. Lockwood. 2008. "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, February 28-29.

Meyer, Robert H. and Michael Christian. 2008. "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, February 28-29.

Padilla, Christine and Alejandra Lopez-Torkos. 2006. "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, April 7-11, San Francisco, California.

Peterson, Paul. 2005. "Making Up the Rules as You Play the Game." *Education Next* 5(4): 42-49.

Rickles, Jordan H. and Melissa K. Barnhart. 2007. "The Impact of Supplemental Educational Services Participation on Student Achievement: 2005-06." Report of the Los Angeles Unified School District Program Evaluation and Research Branch, Planning, Assessment and Research Division Publication No. 352.

Ryan, Susan, and Serah Fatani. 2005. "SES Tutoring Programs: An Evaluation of the Second Year—Part One of a Two Part Report." Policy report, Office of Research, Evaluation and Accountability, Chicago Public Schools, Chicago.

Steinberg, Matthew S. 2006. "Private Educational Services: Whom Does the Market Leave Behind?" *PolicyMatters* 4(1): 17-22.

Sunderman, Gail M. and Jimmy Kim. 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, Deborah L., Elizabeth R. Reisner, B. Bradford Brown, Kimberly Dadisman, Kim M. Pierce, Dale Lee and Ellen M. Pechman. 2005. "The Study of Promising After-School Programs: Examination of Intermediate Outcomes in Year 2." Madison, WI: Wisconsin Center for Education Research.

Vegari, Sandra. 2007. "Federalism and Market-Based Education Policy: The Supplemental Educational Services Mandate." *American Journal of Education*, 113: 311–339

Wilde, Elizabeth T. and Robinson Hollister. 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, Ron, Brian Gill, Paula Razquin, Kevin Booker and J.R. Lockwood 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." Report to the U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

**Table 1: Student SES eligibility, registration and attendance in Milwaukee Public Schools**

<b>Academic Year</b>	<b>Eligible (Middle and High School)</b>	<b>Number Registered (% of eligible)</b>	<b>Number Attended (% of registered)</b>
2003-2004	6508	3707 (57%)	3333 (90%)
2004-2005	9433	3826 (41%)	2610 (68%)
2005-2006	7351	3996 (54%)	2543 (64%)
2006-2007	8119	3897 (48%)	1315 (34%)

**Table 2: Factors influencing registrations among students eligible for SES**

Each number reported is a statistically significant effect (at  $\alpha < 0.05$ )—the percent 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.

Variables	Middle School			High School		
	2004-05	2005-06	2006-07	2004-05	2005-06	2006-07
female	+22%		+50%		+32%	+31%
white (v. black)				-65%		-57%
Hispanic (v. black)				-63%		-60%
Asian (v. black)	-76%	-81%	-72%	-61%	-66%	-72%
Indian (v. black)						
other race (v. 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 (v. 8)				n.a.	n.a.	n.a.
grade year 7 (v. 8)				n.a.	n.a.	n.a.
grade year 10 (v. 9)	n.a.	n.a.	n.a.			
grade year 11 (v. 9)	n.a.	n.a.	n.a.			
grade year 12 (v. 9)	n.a.	n.a.	n.a.		-31%	-46%
<i>number of observations</i>	N=2178	N=1683	N=1374	N=7225	N=5207	N=6635

**Table 3: Estimated effects of attending any SES, 2004-05 and 2005-06 school years**

(Standard errors in parentheses; results statistically significant at  $\alpha \leq 0.05$  shown in bold)

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
<b>2004-05 school year</b>				
Attended any SES				
1. unmatched	-2.486 (4.562)	-3.368 (5.232)	-10.486 (6.243)	<b>-14.420 (7.139)</b>
2. matching	2.024 (5.557)	3.038 (5.916)	-5.427 (8.107)	-4.565 (8.860)
<i>Number of observations</i>	N=1562	N=1571	N=1224	N=1262
<b>2005-06 school year</b>				
Attended any SES				
1. unmatched	-0.529 (0.413)	0.708 (1.202)	0.235 (0.297)	<b>2.846 (1.132)</b>
2. matching	-0.232 (0.427)	0.323 (1.099)	-0.372 (0.357)	1.397 (1.099)
<i>Number of observations</i>	N=1075	N=1016	N=2215	N=2200

**Table 4: Estimated effects of total hours attended SES, 2004-05 and 2005-06 school years**

(Standard errors in parentheses; results statistically significant at  $\alpha \leq 0.05$  shown in bold)

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
<b>2004-05 school year</b>				
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. less<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. less<20)	<b>23.093 (10.201)</b> n=431	16.596 (13.442) n=431	5.888 (13.426) n=185	6.050 (16.675) n=190
6. >30 hours (vs. less<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. less<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
8. # of SES hours w/specific vendors (OLS)	Vendor A.: <b>1.481 (0.623)</b> Vendor B: <b>1.128 (0.444)</b> Vendor C: <b>0.968 (0.362)</b>	Vendor A <b>1.406 (0.621)</b> Vendor B.: <b>1.082 (0.442)</b>	None	Vendor D: <b>3.180 (1.380)</b> Vendor B: <b>3.666 (1.731)</b>
<b>2005-06 school year</b>				
1. # SES hours attended (OLS)	-0.005 (0.013) n=1075	-0.010 (0.040) n=1016	0.007 (0.011) n=2215	<b>0.087 (0.042)</b> 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. less<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. less<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. less<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. less<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
8. # of SES hours w/specific vendors (OLS)	Vendor D: <b>0.136 (0.064)</b>	none	none	Vendor E: <b>0.245 (0.074)</b> Vendor F: <b>0.193 (0.078)</b>

**Table 5: Fixed Effects Model Results**

**Mean, Variance, and Reliability of Student Achievement  
in Grades 8, 9, and 10, 2004-2006**

Subject/Grade	Mean	Standard Deviation	Variance	Average Measurement Error Variance	Reliability	Variance Stabilizing Multipliers
<b>Reading</b>						
8	664.89	36.87	1359.11	184.16	0.86	1.00
9	664.57	49.65	2465.08	182.38	0.93	1.35
10	490.42	63.78	4067.62	386.68	0.90	1.73
<b>Math</b>						
8	670.77	40.85	1668.44	228.05	0.86	1.00
9	666.55	54.53	2973.48	288.04	0.90	1.33
10	517.00	49.00	2401.20	327.15	0.86	1.20

**Fixed Effects Estimates of the Effect of SES in 8<sup>th</sup> and 9<sup>th</sup> Grades, 2004-2006**

	Achievement grade 9 – 8	Achievement grade 10 – 9	Fixed Effects (Double-difference)
<b>Reading</b>			
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			
<b>Math</b>			
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			

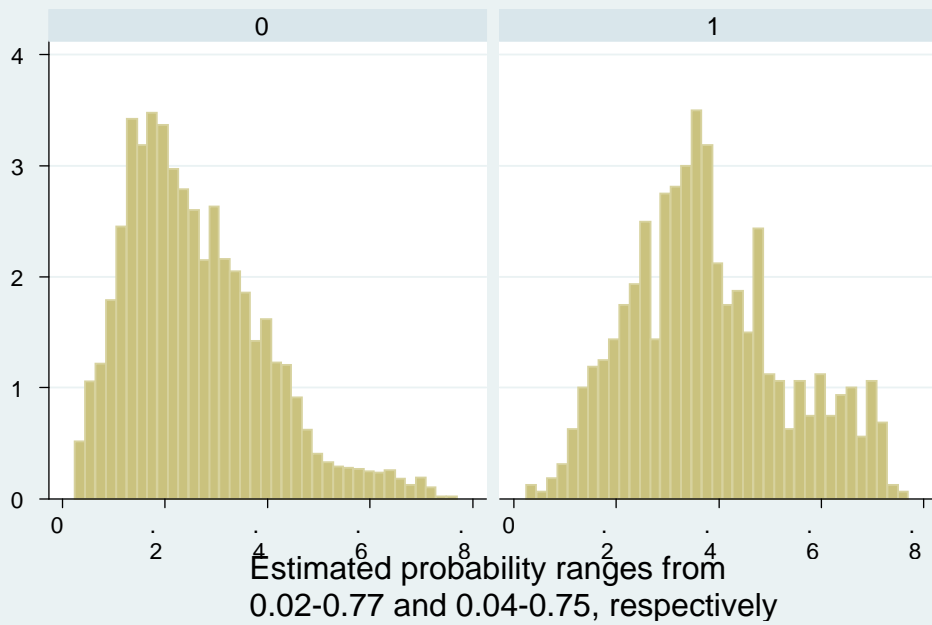
**Table 6: Characteristics of SES Providers Identified as Effective in Increasing Student Test Scores**

<b>Provider characteristics 2004-05 school year</b>	Vendor A	Vendor B	Vendor C	Vendor D	Vendor E	Vendor F
Legal status	Nonprofit	Nonprofit	Nonprofit	For-profit	For-profit	For-profit
Hourly rate charged*	\$80	\$40	\$40	\$40	\$65	\$63
Reported student-teacher ratio	5:1	n.a.	5:1	7:1	10:1	3:1
# of sessions offered	19-23	64	21	46	23	n.a. (at home, on-line)
% of registered students attending any hours (and rank among providers)	44.6 (20)	75.2 (4)	70.0 (8)	43.7 (21)	72.6 (7)	64.8 (11)
% of students attending 20 or more hours (and rank among providers)	33.9 (16)	40.3 (12)	44.4 (9)	17.2 (21)	30.1 (18)	29.3 (19)
Share of registered students	1.5	3.4	2.4	2.3	12.7	7.2

\* Hourly rates charged by approved providers in Milwaukee in 2005 ranged from \$20 per hour to \$100 per hour.

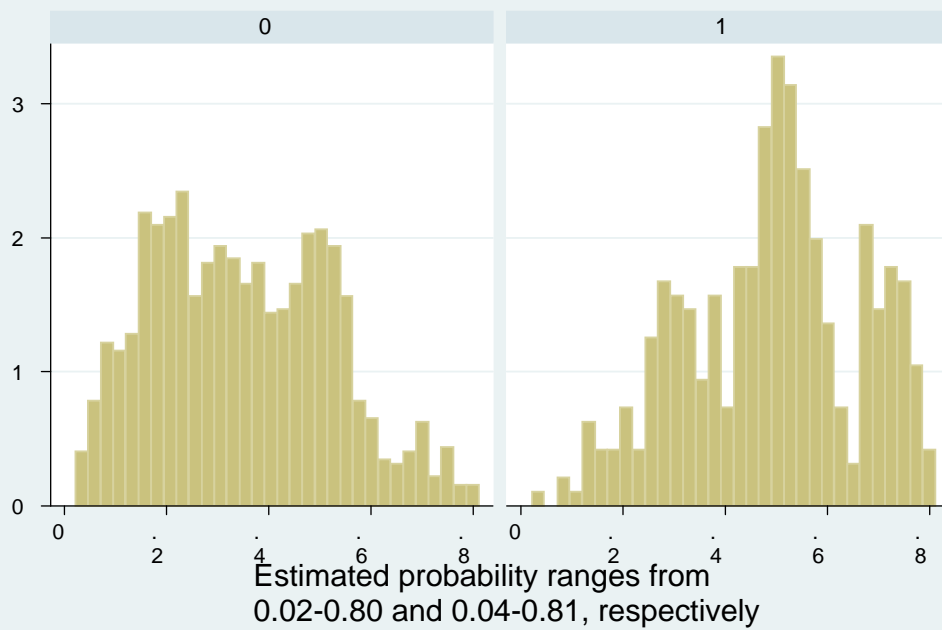
### Figure 1: Predicted probability of SES registration

By attended at least 1 SES session, high school students, 2005-06

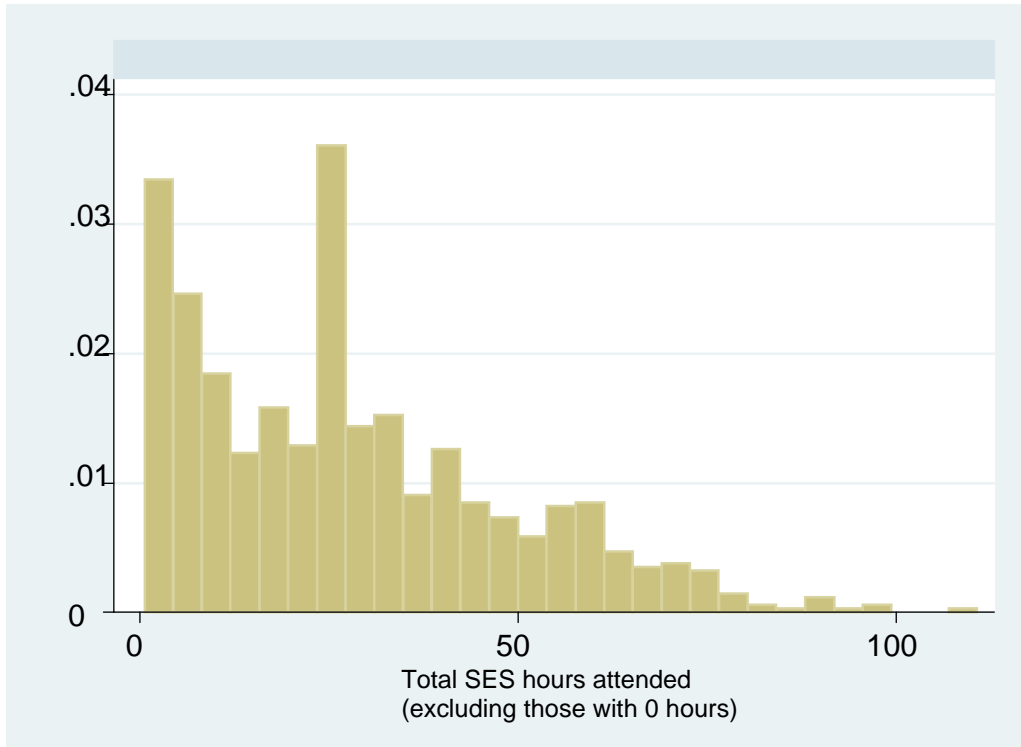


### Figure 2: Predicted probability of SES registration

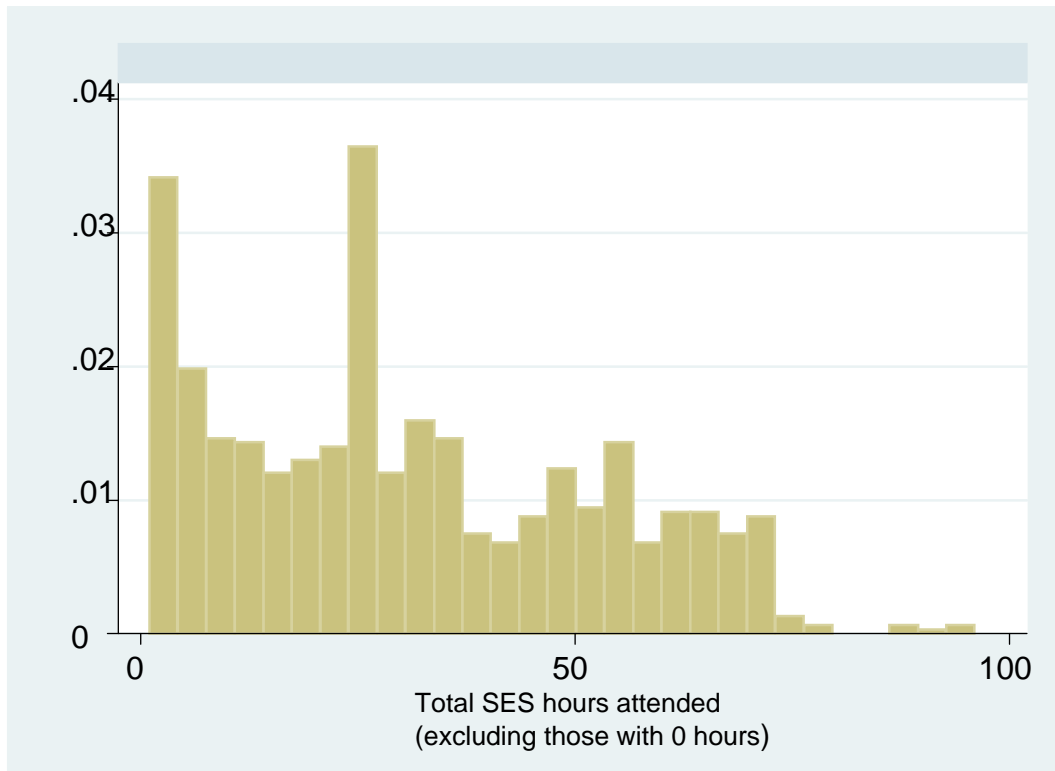
By attended at least 1 SES session, middle school students, 2005-06



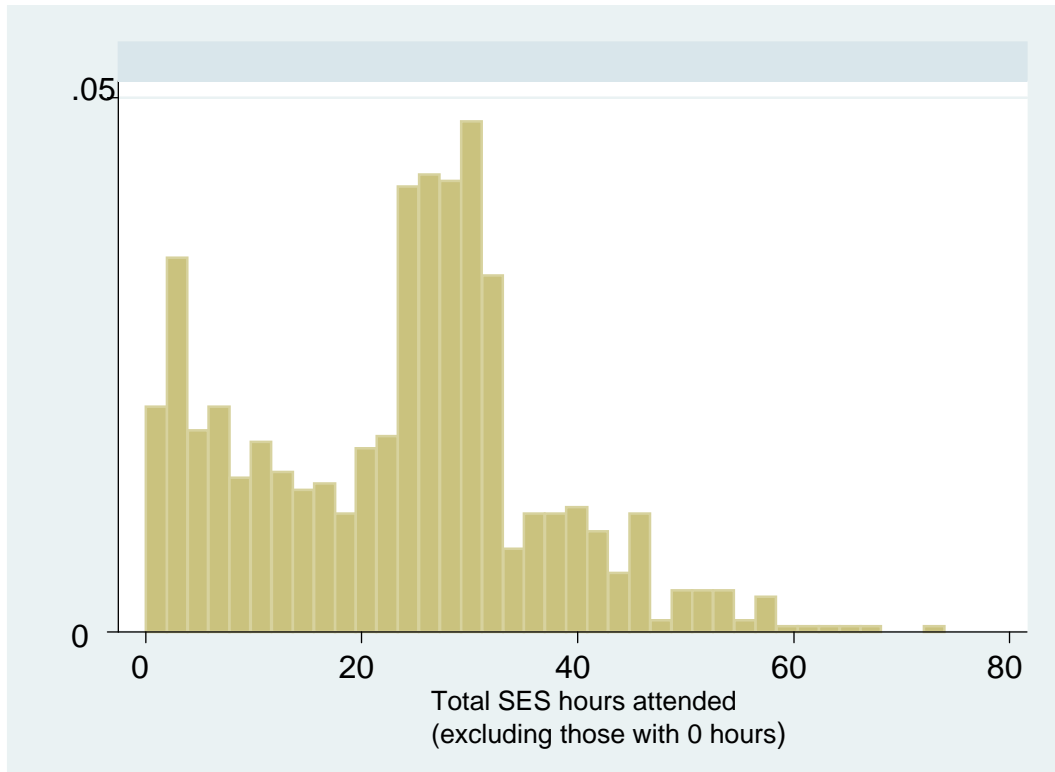
**Figure 3a: Distribution of hours attended, high school students, 2004-05**



**Figure 3b: Distribution of hours attended, middle school students, 2004-05**



**Figure 3c: Distribution of hours attended, high school students, 2005-06**



**Figure 3d: Distribution of hours attended, middle school students, 2005-06**

