Supplemental Education Services Under No Child Left Behind:

Who Signs Up, and What Do They Gain?

Carolyn J. Heinrich
Robert H. Meyer
Gregory W. Whitten

Conference Paper
August 2009

Please visit **www.performanceincentives.org** to learn more about our program of research and recent publications.

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

GREGORY W. WHITTEN
Wisconsin Center for Education Research

Abstract

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

"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.[1]

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.[2] These services typically include tutoring and remediation in reading and mathematics and are provided outside of the regular school day by public or private (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. According to the law, the content and educational practices of SES should be aligned with the state's academic content standards (and applicable federal, state, and local health, safety, and civil rights laws) *[Section 1116(e)(12)(B)(i)]* and should be based on high-quality research evidence of their effectiveness in increasing

student academic achievement *[Section 1116(e)(12)(C)].* In fact, the law requires states to withdraw approval from SES providers that fail for two years to increase student academic achievement.

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

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

We begin now with a brief review of studies on SES implementation and effectiveness and related literature on after school programs and the public and private educational services

market. We next describe the study data, samples and methods of analysis, followed by a presentation of the data analyses and findings. We conclude with a discussion of the findings and their implications for state educational agencies, school districts, and students who are the intended beneficiaries of SES. In general, we find no average impacts of SES attendance on student achievement gains and just a few small effects of the number of SES hours attended.

What (little) we know about SES

In his comments at Woodridge Elementary and Middle Campus in Washington, 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 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 with the law, however, which requires them "to promote maximum participation by providers to ensure, to the extent practicable, that parents have as many choices as possible" and to notify parents of the availability of SES and allow them to select "any approved provider that they feel will best meet their child's needs" *[Section 1116(e)(4)]*.

State and local educational agencies have, in effect, been charged with the major responsibilities for initiating SES and ensuring compliance with NCLB specifications, including: establishing processes for identifying and approving providers and encouraging their participation to expand choice for students; developing, implementing, and publicly reporting on standards and methods for monitoring SES quality and provider effectiveness; managing contracts and financial systems for paying providers for services, and designing procedures for

withdrawing approval from unsuccessful providers. For many state and local educational agencies, the development of these procedures for coordinating and evaluating SES, with no new federal funding, is still a work in progress. The Government Accountability Office (GAO, 2006) reported that three-fourths of states 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.

In addition, participation in SES among eligible students has generally been low nationwide. The onset of SES was initially slowed by the NCLB stipulation that only schools 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, 19 percent in 2004-2005 and an estimated 23 percent (3 million) in 2006-2007 (Peterson, 2005; Government Accountability Office, 2006; Education Industry Association, 2007). Still, 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 program administrators indicated that it was not only difficult to get students registered, but their attendance at SES sessions also declined significantly as the school year progressed (GAO, 2006; Peterson, 2005). Recognizing this problem, some providers

offer students incentives to sign up for services and/or to encourage student attendance, ranging from computers to school supplies and gift certificates.

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

For some of these same reasons that state and local educational agencies have been challenged in their efforts to acquire knowledge of SES content and effectiveness, researchers have also been limited in their ability to conduct rigorous evaluations of SES. Early research on SES was largely descriptive and exploratory, focusing mostly on the challenges of implementing SES in a rapidly evolving market with limited capacity and asymmetric information on both the demand and supply sides (Burch, Steinberg, and Donovan, 2007; Vegari, 2007; Farkas and Durham, 2006; Padilla and Lopez-Torkos, 2006; Steinberg, 2006; Anderson and Laguarda, 2005; Casserly, 2004; Sunderman and Kim, 2004). These researchers 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, 2009; Burch, Steinberg, and Donovan, 2007;

Farkas and Durham, 2006). They described providers' widely varying hourly charges for SES

and attempted, 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). 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.

Empirical studies of SES effectiveness

A growing number of studies have sought to empirically estimate the effects of SES. For

example, Chatterji, Kwon and Sng (2006: 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.[4] The researchers estimated generalized linear 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. The authors did not discuss, however, whether any sensitivity testing was conducted to assess the validity or robustness of their model to alternative assumptions in model specification.

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 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.

Likewise, two studies in Minnesota that used matched samples of students who took the

Northwest Achievement Levels Tests (NALT, a national norm-referenced test) to compare

students who did not participate in SES to those receiving SES did not find positive effects of

SES participation or significant differences among SES providers as assessed by changes in

students' annual NALT reading scores (Burch, 2007).

A RAND study (Zimmer 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.

There is at least one other study in progress—a feasibility study by Mathematica Policy Research, Inc (MPR) and COSMOS Corp. initiated in September of 2007—that is investigating whether there are a sufficient number of school districts where there are more SES applicants than can be served with the available Title I funding to facilitate a regression discontinuity approach to evaluating the impact of SES. Such a study would also require school districts to use quantitative data to establish a cutoff point for students' eligibility for SES and to use this cutoff to determine who participates in SES. This design would limit researchers to estimating the effects of SES for students who apply and are allowed the opportunity to participate compared to those below (excluded by) the cutoff point within a specified bandwidth of the cutoff, and prior research suggests that not all applicants above the cutoff would actually follow through and attend SES. The potential of such a research design to produce a relatively accurate and generalizable impact estimate of SES is yet to be determined.

Insights from other after school and tutoring program studies

Of course, other after school study and tutoring (or "out-of-school-time") programs have long been in operation, including federally funded programs, and there is a larger literature on their implementation and effects. Lauer et al. (2006) conducted a recent synthesis of the research evidence on out-of-school-time programs specifically in response to the new NCLB requirements to offer SES. They acknowledged up front that although evaluations of after

school tutoring and related interventions are profuse, relatively few are rigorous in their research design and methodological approach. Selecting only those studies (35) published in peer-reviewed journals and using control or comparison groups to estimate effect sizes (e.g., gain scores), they used meta-analysis techniques to explore the relationship of program focus, duration, timeframe, student grouping and grade level to program outcomes. They find, based on their review, that out-of-school-time programs can have a positive effect on student achievement (in relation to at-risk students who do not participate), although the effects are not likely to be large enough to close the achievement gap between at-risk students and those who are more advantaged. In addition, effect sizes were larger for programs that were more than 45 hours in duration but became smaller for those longest in duration. In a random assignment study of a national 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; Burch, 2009). In their study of parental decision making and charter school choice 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, student surveys, interviews with provider and district staff and information from observations of SES providers, and secondary student-level demographics, transcript and achievement test data from MPS. The focus groups with parents of students enrolled in eligible SES schools for 2005-2006 were conducted by Heinrich (lead author) with the support of three doctoral students in August 2006 in a formative stage of the research intended to explore basic issues about how parents learned of SES (or *if* they knew of the program), how they chose a provider for their child, their opinion of the tutoring's effects on their child's academic performance, and any difficulties that they encountered in arranging services. A random, stratified sample of 320 parents/students was first drawn from MPS administrative data to ensure that we included both parents (or guardians) whose children had registered for SES in the 2005-06 school year (approximately 60%) and

those who chose not to participate in SES. Letters were mailed to parents by MPS to invite them to participate in the focus groups, which took place in public libraries easily accessible by public transportation and in zip codes with high concentrations of SES schools. Although the total number of focus group participants (n=17) was low as a 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 focus group protocol is shown in Appendix I.

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 students about how they were spending time in their tutoring sessions and how these sessions were affecting their performance in school. That is, the collection of these data were important in obtaining another window on what Burch and Good (2009) describe as the "instructional core" of SES. The surveys also asked students about how they chose their SES provider and what would encourage them to attend more tutoring sessions. Questions included closed-ended items such as "In a typical or average SES session, how many <u>minutes</u> do (or did) you spend working *one-on-one* with a tutor in your SES sessions?" as well as open-ended items such as "What have you learned in your SES sessions this school year?" The survey questions are shown in Appendix II. In establishing the sample frame for these surveys, eight SIFI schools with the largest SES enrollments were selected. Students in seven of these eight schools completed the survey a second time within one month (in late March or early April) if they were still participating in SES to assess the consistency of (or changes in) responses. A total of 1441 students responded to the first survey, and 874 students participated in the second survey.

The field research that included in-depth observation of SES sessions and interviews with

tutors and other MPS district staff was conducted by colleagues at the Wisconsin Center for Education Research, Patricia Burch and Annalee Good. The information that we draw into this study is based on the first phase (January-May 2007) of their companion qualitative research, involving four SES providers and 22 hours of classroom observation, 15 interviews with provider staff and six interviews with district-level staff. The four providers that they studied were selected to allow for diversity in tutoring formats, administrative structures, settings (e.g., home, community, school) and curricular approaches. Additional details on the qualitative study design are provided in Burch et al. (2007) and Burch and Good (2009). We also conducted some follow-up interviews with the MPS SES program coordinators over the course of this study.

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

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

data sets are reported in Table 8.

Sources of Test Score Data by Grade and Year				
Grades	2004-2005	2005-2006	2006-2007	2007-2008
5 - 8	Old WKCE or TerraNova	New WKCE AGS	New WKCE	New WKCE
9	TerraNova	TerraNova AGS	none	ThinkLink
10	n.a.	New WKCE AGS	New WKCE	n.a.

To evaluate the effect of SES during the 2004-2005 school year, we used reading and mathematics scores from three different sources. Achievement prior to participation in SES was measured in November 2004 using the state assessment, the Wisconsin Knowledge and Concepts Examination (WKCE), in grade 8, and the Terra Nova in grades 6, 7, and 9. Both tests were scored on what we refer to as the "old" WKCE scale. (A new WKCE scale was introduced in 2005.) Achievement after participation in SES was measured in November 2005 using the new Wisconsin Knowledge and Concepts Examination, administered in grade 3-8 and 10, and the TerraNova in grade 9.[5]

The state assessments in 2004 (and prior years) and in 2005 (and later years) were scored on different scales: scores at a given grade were typically lower by several 100 points on the new scale, but the variances in scores were very similar. As a result, it is reasonable to measure growth in student achievement from 2004 to 2005 as the gain in student achievement *relative* to the average gain in achievement.[6] Including grade indicators in the model (the strategy used in this research) addresses the issue of differences in the location (means) of the state test scales. The key requirement is that the variances of the two scales are similar at each grade (although possibly different at different grades).[7]

In 2005-2006, AGS[8] assessments were administered to all students enrolled in schools required to offer SES to their students, with tests in both math and reading administered twice in

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

It is important to make clear the major differences between the assessment data available for evaluating the impact of SES in school years 2004-2005, 2005-2006, and 2006-2007, given their implications for the interpretation of SES effect estimates. First, the data for 2004-2005 covered SES participation only up through ninth grade; the 10th grade test score (administered in November 2005) serves as a posttest outcome for 9th grade participation. Second, the data for 2006-2007 covered SES participation only up to 8th grade as a pretest was not administered in 2006 in 9th grade. In contrast, the AGS assessment was administered in all middle and high school grades in 2005-2006, although a substantial number of middle and high school students did not take the assessment (see Table 8). In contrast, there was very little missing test information for the other assessments since all students were required by the state or MPS to take the state assessment or the TerraNova assessment. Our statistical analysis of the patterns of missing test data in 2005-06 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.[9]

Another important difference between the AGS assessment and the state assessment (and related assessments) is that the growth period is approximately 7 months (October though April) for the AGS assessments (2005-2006) and 12 months (November to November) for the latter assessments. 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. We 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 (standard deviations) of the AGS test scale and the state assessment (and related assessments) differ substantially, as indicated in Table 8. This information implies the following: (1) Parameter estimates should generally be larger for analyses based on the state data than the AGS data (2005-2006 only). (2) In the AGS data (2005-2006 only), parameter estimates based on reading scores should generally be larger (by a factor of three) than estimates based on math scores.

Estimation methods

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

17

registration rates that do not beget rationing of services.

In the absence of random assignment, it is important to understand and model the process by which those who participate in SES choose to receive services. First, we know that NCLB requires school districts to use the same data to determine eligibility for SES that they use for making within-district Title I allocations, which has historically consisted of information on free school lunch eligibility. School districts are required to notify families of their children's eligibility for SES and to cooperate with approved providers in disseminating information to students and parents about available services. Only if more families request SES than there are funds available to serve them are districts obligated establish priorities or criteria to determine which eligible students get access to services.[10] At the same time, we know that applying for SES typically takes some initiative on the part of parents and students in following through the steps of registration and choosing a provider. Thus, although we alleviate some selection concerns by using an internal comparison group of students who are also eligible but do not sign up for SES to compare with those who register to receive services (Heckman et al. 1999), there may still be some selective differences between these two groups of students that we need to adjust for in our analysis.

In this study, we investigated selection into SES using multiple methods. As data are not systematically collected on how parents and students decide to participate in SES (or not) after receiving information about the services, we conducted focus groups with parents and some students to probe and gain insight into their decision making processes (see Appendix I and Heinrich and Whitten, 2006 for further details). Student responses to the surveys administered to eligible students in the 2006-2007 school year were also informative about selection into SES. In addition, we used transcript and administrative data for SES-eligible students to predict their

probability of registering for SES and the probability of attending SES (conditional on registration), using logistic regression. The predicted probabilities generated by the logistic regression models subsequently serve an important purpose—this information is used to remove possible bias associated with pre-intervention differences between students who participated in SES and those who did not register, allowing for more accurate identification of SES effects.

Propensity score matching

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

In attempting to identify the effects of SES in increasing students' reading and math achievement, we compare eligible students who registered for SES with eligible students who did not register; and among students registered for SES, we also compare the effects of differing levels of hours of SES attended. We use propensity score matching methods—matching on the predicted probabilities from the logistic regressions noted above—to account for observed differences between the groups, including differences in student characteristics and in the public schools they attended.[11] We also recognize, however, that this approach will be inadequate in the face of any important, unobserved factors that influence both selection into SES and student

outcomes. Indeed, Wilde and Hollister (2007) recently compared experimental estimates of the

Project STAR class-size reduction experiment on student achievement test scores with estimates

produced using propensity score matching and concluded that the nonexperimental results were

insufficiently close to experimental impacts to place confidence in them, which they attributed to

the role of unobservable characteristics. It is also worth noting, however, as these authors

acknowledged, that only a limited number of covariates were available for use in matching, and

one half of their treatment subgroup sizes were very small (i.e., five or fewer cases.)

Because our dependent variables (measuring changes in student achievement) are defined

as the difference between pre- and post-treatment outcomes (test scores) for each student, we are

able to use a "difference-in-differences" version of the matching estimator that allows for a time-

invariant (unobserved) difference in outcomes between participants and comparison group

members and often performs better than its cross-sectional counterpart, in part because it allows

for this unobserved heterogeneity (Mueser et al., 2007). The specific matching technique we

apply is radius caliper matching, with the caliper set at 0.01. In other words, students who did

not register (or attend) SES are matched with students who participated if their predicted

probabilities of registration are within 0.01 of each other.[12] We also impose a "common

support" requirement, so that if there is not a match between students in the two groups, the

student is dropped from the analysis; no more than two cases were dropped (among both high

school and middle school students) due to a lack of a common support. After students are

matched, the differences in their changes in achievement test scores—the change in test scores

from the fall before SES to spring or the following fall after SES—are calculated, along with

bootstrapped standard errors and bias corrections. In light of recent work suggesting that

bootstrapping may not be appropriate with nearest neighbor matching methods (Abadie and

Imbens, 2006), we also applied alternative techniques, including local linear matching methods and "nnmatch" in STATA, which yielded comparable results.

Fixed effect model

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

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

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

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

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

Equation (3) yields separate, unbiased estimates of the effect of any SES participation in years

one and two given the maintained model assumptions, although note that the coefficient on

SES_{1i} is the negative of the year one SES effect.

More efficient estimates of the impact of SES can be obtained by aggregating estimates

across years one and two and across grade levels. Since SES providers often serve students at all

grade levels, it is quite reasonable to pool information across grade levels (as in the estimates

based on match methods). Below, we report estimates of SES effects for students in middle

school and students in high school. Both sets of estimates are based on the most recently

available data from Milwaukee Public Schools.

One important requirement of the fixed effects model presented above is that the test

scores in different years and grades must be measured on the same scale. Since the data set used

to estimate the SES impact for high school students measures achievement on different scales,

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. Since the data set used to estimate the SES impact for middle school students in

school years 2005-2006 and 2006-2007 measured achievement on the same scale in all years (the

new WKCE scale), there was no need to rescale these data.

Who signs up for SES?

As the discussion above suggests, there are multiple stages of student/parent selection

into SES, including: *awareness* through dissemination of information to students and parents by

schools about eligibility and availability of services; *registration* of students with a specific SES

provider that is encouraged by providers through marketing (i.e., sending out brochures, inviting

parents to presentations, and offering incentives to students to register with them); and following

registration with a provider, student *attendance* at SES sessions, for which providers invoice the

school district to request payment for the number of hours of service provided. Table 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). Approximately one half of eligible students registered for

SES over the study period, but the number attending any sessions declined substantially over

time (from 90% in 2003-04 to only one-third in 2006-07). Clearly, understanding why more

eligible students do not progress through registration and attend at least one session is important

not only for school districts and providers trying to comply with NCLB and improve student

outcomes, but also for research efforts to evaluate the effectiveness of SES.

Although the focus groups with MPS parents and students were primarily exploratory

with a small number of attendees (n=17),[13] an important finding was that despite the

considerable information made available by schools and providers to inform parents of SES

options for their children, not all parents receive or understand this information. Fifteen of the

17 parents had a hard time distinguishing SES from other school-based tutoring or after-school

programs and identifying which options were available free-of-cost to their children, and they

indicated that they received little or no assistance in making an informed choice for their child.

Many of these parents did not receive the booklet on SES options prepared by the school district,

and some reported being skeptical of information received through postcards and other direct

mailings from SES providers: "*I think a lot of it is smoke and mirrors, you know… what they

really have to offer as far as 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, these same parents had a clear idea of what factors they *should* be considering in making these choices. They most frequently responded that they would like to know how much one-on-one tutoring their child would receive (time per session and total number of hours), what the student-teacher ratio would be in group SES sessions, and specific information on the tutor qualifications and academic content of SES sessions. Some of this information was available in the district SES booklet, but only two parents reported seeing or using it. Other parents described important logistical concerns about their children's attendance (e.g., transportation) or their ability to accommodate an in-home provider or 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). Indeed, approximately 18 percent of eligible students who responded to the 2007 surveys reported that they missed attending SES sessions because of problems getting to and/or from the location of services.

Table 2 presents basic descriptive information on the middle and high school students who were eligible for SES in the 2004-2005, 2006-2006 and 2006-2007 school years by whether they registered for SES or not. It is apparent that females and black students in middle and high school are more likely to register for SES, whereas white and Asian students and Hispanic students in high school are less likely to register for SES. Over time, schools appear to be more effective in targeting those receiving free lunch for services. Other than these associations with demographic characteristics, however, most other statistically significant associations shown in Table 2 suggest that the SES-eligible students who register to receive services are more advantaged than those who do not sign up: they have a lower number of student absences and are less likely to be retained, and among high school students, they have higher grade point averages

and are more likely to be English proficient. In addition, those who have attended SES in a prior year are more likely to register again.

We also explored the factors influencing SES registration and attendance decisions in logistic regression models that were used to estimate the propensity scores for matching, where the dependent (or outcome) variable was a binary variable equal to one if a student registered for SES and zero if the student did not sign up. In addition to the student demographic, school performance and attendance variables shown in Table 2, these models also included controls for students' grade year and school attended. Bearing in mind the focus group findings, we recognize that relevant information about possible parental influences and family constraints on the decision to register (e.g., parental education levels and employment status, location of residence, etc.) is not included in these models, and to the extent that they differ between registrants and non-registrants and relate to SES outcomes, they may bias the results.

The results of these analyses are presented in summary form in Table 3, and like the simple descriptive analysis, they show some consistent relationships between student characteristics and school experiences and registration for SES among eligible students. In this table, a blank cell indicates that a particular variable did not (statistically) significantly influence student registrations in that year.[14] If the variable was a statistically significant predictor, the increase in odds of registering for SES associated with that variable is reported. For example, female SES-eligible students are significantly more likely to register than males (with 22% to 50% higher odds of registering), and Asian students have approximately 60-80% *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). As expected, free

lunch eligibility becomes a stronger predictor over time, with up to 161% higher odds for middle school students eligible for free lunch, as does prior SES attendance, with up to 181% higher odds for high school students in the 2005-2006 school year. As also seen in the descriptive analysis, student absences from school (measured in the school year prior to the SES enrollment year) lower the odds of registering for SES.

Since the factors that influence student *attendance* at SES sessions (following registration) differed little from those affecting registration, we more briefly discuss the statistically significant findings of these models (available from the authors). Among the registered students, females and those who attended SES in a prior year were significantly more likely to attend any SES sessions, and student absences from school in the prior year were again negative and statistically significant in all models. The new and substantively important findings were that among middle school students, registered Hispanic students (and those who were not English proficient) were significantly more likely to attend. In addition, contrary to the observed relationship between free lunch eligibility and registration, high school students eligible for free lunch were less likely to attend any SES (although this relationship was statistically significant only in 2004-05). Registered high school students who had been retained were also significantly less likely to attend any SES sessions. Although not reported in Table 3, most of the school indicators in both the registration and attendance models were statistically significant, suggesting that school-based factors (e.g., teacher roles in encouraging attendance, school location, etc.) also influence these decisions. An analysis using Chow tests for differences in coefficients across schools in the models predicting SES attendance suggested that the factors influencing SES attendance likely differed across schools (although no specific factors were identified).[15]

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

The effects of SES on student achievement

Effects of any SES attendance

We begin with a simple propensity score matching estimation of the effects of *any SES attendance* on student achievement (our "black box" estimation), where the SES "treatment" is measured using a binary variable that indicates any time spent in SES. In each of the models (for middle school and high school students in the 2004-05, 2005-06 and 2006-07 school years), there was strong overlap in the distribution of propensity scores for registrants and non-registrants, implying that it was relatively easy to find matches between students who registered for SES and

similar students who did not register.[17] The results of balancing tests used to assess the

performance of the propensity scores in balancing the distribution of observed characteristics

between the treatment and comparison groups (available from the authors) showed that after

matching, there were no statistically significant differences in their mean characteristics, with

substantial percent reductions in bias for most variables (as high as 99.4%).

Table 4 presents the estimated effects of attending any SES on changes in reading and math

test scores in the 2004-05, 2005-06 and 2006-07 school years;[18] 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 13 hours for high school students in 2006-07 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). As 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

We examined the distributions of hours attended for middle and high school students in 2004-05, 2005-06 and 2006-07 (if they attended any hours), based on provider reports of actual student hours attended that are required to receive payment for services.[19] The patterns were very similar across these four groups, with the highest peaks around 25 hours, but also with nearly as high spikes in the distributions close to 0 hours (left skewness). This is consistent with both student and vendor reports in our study, as well as in other research, suggesting that getting students to attend (to show up regularly for SES) is an ongoing challenge in the implementation of SES (GAO, 2006). In the literature discussed above, one study of SES (Ryan and Fatani, 2005) and a meta-analysis of after-school programs (Lauer et al., 2006) suggested that effects of tutoring programs were larger for programs that were more than 40-45 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 6 percent of high school students, and in 2006-07, just 0.07 percent of middle school students and 0.05 percent of high school students attended SES for 40 or more hours.

In analyzing the effects of total hours attended on changes in students' math and reading test scores, we first estimated an ordinary least squares (OLS) regression model with the same set of controls for student characteristics and school attended (as shown in Table 2) and with a continuous measure of total hours of SES attended. The results, reported in line 1 of Table 5 for each of the student subgroups, show only one statistically significant, positive effect of total 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 at the stage of 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, and in 2006-07, about one half of registered middle school students and just over one-third of registered high school students attended at least one session. Lines 2 and 3 in Table 5 show the results of models that compare changes in test scores for students with 20 or more hours of SES who are matched with registered students who did not

participate, and registered students with 40 or more hours of SES matched to registered students with no hours attended. Across all of these estimated effects, there are no statistically significant effects of attending more than 20 hours or more than 40 hours of SES.

In an additional set of analyses, we further restricted the samples to include only students who attended some hours of SES, and then we matched students with differing levels of SES: those with greater than 10 hours of SES 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 (as all are registrants) or at that which determines whether or not students attend any SES; however, there may still be selection into these varying levels of attendance, which the propensity score matching approach will address to the extent that selection is on observable characteristics. The results of these matching analyses are summarized in lines 4-7 of Table 5. Amongst all of these estimates, there is only one statistically significant finding for the group of middle school students in 2004-05 receiving at least 20 hours of SES (compared to less than 20 hours). A final set of matching analyses conducted separately by grade for grades 8-12 (but not reported here) likewise did not change the predominant finding of no statistically significant effects of SES attendance on students' math and reading achievement.

Recall that the fixed effects or "double difference" model (from equation 3 above) produces separate, unbiased estimates of the effects of any SES participation on math and reading achievement (given the assumptions of the model). The results of these models are shown in Tables 6 (2004-2006) and 7 (2005-2007). The top part of Table 8 shows the mean, variance, and reliability of student achievement in grades 8, 9, and 10 for the period 2004-2006 and the multipliers used to produce equal test variances for all three tests. Table 6 presents the

parameter estimates and standard errors of primary interest in this study, that is, the coefficients of the SES indicators in the fixed effects models for grades 8 and 9. As indicated in the table, the SES estimates are not statistically different from zero. The middle part of Table 8 shows the means and standard deviations for reading and math achievement for three grades for each of the three middle school cohorts for the period 2005-2007. Table 7 presents parameter estimates and standard errors of the coefficients of the SES indicators in the fixed effects model for these cohorts, along with a single pooled estimate that optimally combines the information from the separate cohorts. These results provide no evidence of statistically significant SES effects. Clearly, these results are consistent with those of the propensity score matching models, showing no statistically significant effects of SES participation on students' math or reading achievement at any grade level. 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.

Qualitative study findings on SES effectiveness

The qualitative components of this study—student surveys, observations of SES providers, and interviews with SES provider and school district staff—yielded valuable information and insights for understanding the empirical findings of no SES effects on student achievement. One obvious problem is that students are not attending a sufficient number of SES hours, with less than one percent of middle and high school students attending for at least 40

hours in the 2006-07 school year. Perhaps of greater concern, though, is what students are doing in the SES sessions that they attend. Data collected in the MPS student surveys showed that they attend an average of 2.4 SES sessions per week of an average length of just over 60 minutes. In a set of five questions that asked students how they spent their time in SES sessions—the number of minutes working one-on-one with a tutor, working on their own (e.g., on homework or in other self-directed activities), in group activities led by a tutor, in group activities without a tutor, and socializing or other non-academic uses of time—they reported spending the most time in self-directed activities (30 minutes on average). In their observations of instruction in SES sessions, Burch et al. (2007) and Burch and Good (2009) saw that many of these activities take the form of "more school," that is, they consist of a lot of desk time and worksheets, with few opportunities for richer activity-based programming and other such activities. In one example reported by Burch and Good (2009: p. 8), they describe how "students exclusively sat at desks in every tutoring session observed over two years, and the only exceptions to packet or worksheet-based activities were occasional class discussions, a Jeopardy-type review game, and an activity where students did math questions related to a documentary."

Another problem identified by Burch and Good (2009) through their observational study is that even in SES sessions where the level of student engagement is high, the work in which they are engaged tends to be disconnected from the school curriculum, contributing to a lack of continuity in their day-time and after-school learning environments. This finding was consistent with MPS student responses to a set of survey questions that inquired about subject areas where they needed help and whether the content and activities of SES sessions helped them to improve in these areas. For example, just over a quarter of students who reported needing help on reading worked on their reading, and less than 10 percent of students reported writing an essay or

paragraph during a SES session. Overall, less than 30 percent of these MPS students responded that the SES sessions had been very helpful to them in improving their performance in school. Burch et al. (2007) add that planned curriculum frequently does not match curriculum in use, and that direct connections to the school day curriculum are more likely to happen by chance, such as in the case where a tutor is also a teacher in the child's school. As one school principal expressed in an interview (Burch et al., 2007, p. 16): "This is my problem with the whole SES program…how can they expect to move children academically if you don't look at the curriculum that they're being taught during the day and connect that with what they're being taught at nighttime during tutoring? It's gonna be…a worksheet heaven."

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

SES providers, for their part, may have a legitimate grievance in stating that inconsistent or waning student attendance makes it difficult for them to help students advance and to improve the effectiveness of instruction. At the same time, more than a quarter of students reported in the

MPS survey that they were not learning anything in the sessions, and thus, it is possible that if

SES instruction were more effective, better attendance would follow. Student responses to the

survey question asking what would encourage them to attend more sessions were mixed, though,

and the most common response suggested more food, prizes, games or rewards, rather than

improvements in academic content.

Discussion and 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, while

relying heavily on the private sector to supply the core services and present adequate competition

and choice. This demanded the support of school staff and teachers, parents, and community

members at a time when the opportunity costs of Title I funds were high. In addition, NCLB

requires state and local educational agencies to assess SES providers' effectiveness in increasing

student achievement and to disseminate this information to parents of children who are eligible

to receive services, and our study shows just how difficult this is to do—i.e., to identify with

confidence the effectiveness of SES vendors.

Six years since NCLB mandated the provision of SES, what do we know about its

implementation and effectiveness? Although this study focuses on a single, urban school

district, our findings are generally consistent with the growing body of field research and related

studies that are investigating SES programs in medium- to large-size urban districts. First, 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 declining over time, particularly among older students. And although it is positive to find that MPS students eligible for free lunch (i.e., from lower income families) are significantly more likely to sign up (among those eligible for SES), the empirical evidence suggesting that free lunch eligible students are significantly less likely to attend after registering is discouraging. The empirical analyses also suggested that students with more absences or who had been retained were less likely to register and/or attend SES. In effect, students who are more likely to have higher levels of academic need for SES may be missing or declining the opportunity to receive the extra tutoring and individual help that NCLB intended to provide.

In inquiring about what would encourage students to attend more SES, the students indicated that they are primarily responding to incentives and prizes (i.e., computers, gift cards, fun and games, food and candy). The focus group findings suggested that parents had very different concerns in choosing participation or a SES provider for their children, 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). Although we are not able to establish a causal relationship, we suggest that it is highly likely that there is some link between the dramatic drop-off in student attendance at SES programs in the 2006-2007 school year (down to 34% from 64% in 2005-2006, as shown in Table 1) and these new restrictions on incentives. Alternatively, based on these student reports and the empirical analyses in this study that produced little evidence of the effectiveness of SES in increasing student achievement, one might also speculate that parents and students are, in fact, choosing rationally in not registering for or attending SES. Using propensity score matching techniques and fixed effects models to adjust for student selection into SES, we failed to 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.

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. With colleagues at the Wisconsin Center for Research, we are also continuing with research that focuses on getting "inside the black box"

to better understand why SES programs are not currently more effective, and state and local educational agencies are also eager to increase their understanding of how these programs' effectiveness might be improved through policy and market governance changes, such as the use of performance-based contracting. In addition, we faced important limitations in this research. We did 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. And of course, this study is based on research from a single urban school district, and although this poorer and predominantly minority population of SES-eligible students is very comparable to the larger national population of eligible students, the cross-state and -district variation in SES provider markets and program administration that is relevant to program outcomes might very well limit the wider applicability of these study findings.

Notes

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

[2] Students attending Title I schools identified for improvement are given the option to transfer to another public school or to receive supplemental educational services, depending on the eligibility of the student and the status of the school. These options continue until the school has made 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.

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

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

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

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

[7] 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.

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

[9] 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.

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

[11] 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.

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

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

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

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

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

[17] For example, for high school students in 2005-06, predicted probabilities of SES registration ranged from 0.02-0.77 for students who did not attend any SES and 0.04-0.75 for those who attended; for middle school students, these same probability ranges were 0.02-0.80 and 0.04-0.81, respectively.

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

[19] 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.

References

Abadie, A. and G.W. Imbens. 2006. "On the failure of the bootstrap for matching estimators." NBER Technical Working Paper 325.

Anderson, L.M. and K.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, P. 2009. *Hidden Markets: The New Educational Privatization.* New York: Routledge. The Critical Social Thought Series

Burch, P. 2007. "Supplemental Educational Services under NCLB: Emerging Evidence and Policy Issues." Boulder: Educational Policy Research Unit, University of Colorado.

Burch, P. and A. Good. 2009. "Getting to the Core: The Role of Instructional Setting in Federally Mandated Tutoring." Working paper presented at the 2009 American Educational Research Association Conference, San Diego, CA.

Burch, P., A. Good, C. J. Heinrich, R.H. Meyer and G. Whitten. 2007. Summary of Preliminary Findings: Milwaukee Public Schools Supplemental Education Services Evaluation. Report to Milwaukee Public Schools, Wisconsin Center for Education Research, August.

Burch, P., M. Steinberg and J. Donovan. 2007. "Supplemental Educational Services and NCLB: Policy Assumptions, Market Practices, Emerging Issues." *Educational Evaluation and Policy Analysis* 29(2):115-133.

Casserly, M. 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, M., Y.A. Kwon and C. 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): 1-44.

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, M., S. James-Burdumy, M. Moore, L. Rosenberg, J. Deke and W. Mansfield. 2004. "When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: New Findings." Washington, D.C.: U.S. Department of Education.

Education Industry Association. 2007. "Supplemental Education Services (SES) and the 2007-2008 School Year: The 'No Child Left Behind' Program Can Make a Difference For More than 3 Million Students Nationwide." Fact Sheet, August 24.

Elbaum, B., S. Vaughn, M.T. Hughes, and S.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, G. and R. 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, E.A., J.F. Kain, S.G. Rivkin and G.F. Branch. 2007. "Charter school quality and parental decision making with school choice." *Journal of Public Economics* 91(5-6): 823-848.

Heckman, J. J., R. J. Lalonde, and J. A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs" in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, NY: Elsevier, Chap. 31: 1865-2097.

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

Lauer, P.A., M. Akiba, S.B. Wilkerson, H.S. Athorp, D. Snow and M.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, D.F. 2007. "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement." *Electronic Journal of Statistics* 1(1), 223*C*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, D.F., B. 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, R.H. and M. 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.

Mueser, P. R., Troske, K. R. and Gorislavsky, A. (2007). Using State Administrative Data to Measure Program Performance." *Review of Economic and Statistics,* 89(4), 761-783.

Padilla, C. and A. 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, P. 2005. "Making Up the Rules as You Play the Game." *Education Next* 5(4): 42-49.

Rickles, J.H. and M.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, S., and S. 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, M.S. 2006. "Private Educational Services: Whom Does the Market Leave Behind?" *PolicyMatters* 4(1): 17-22.

Sunderman, G.M. and J. 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, D.L., E.R. Reisner, B.B. Brown, K. Dadisman, K.M. Pierce, D. Lee and E.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, S. 2007. "Federalism and Market-Based Education Policy: The Supplemental Educational Services Mandate." *American Journal of Education*, 113: 311–339

Wilde, E.T. and R. 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, R., B. Gill, P. Razquin, K. 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

Academic Year	Eligible (Middle and High School)	Number Registered (% of eligible)	Number Attended (% of registered)
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: Characteristics of SES registrants vs. eligible non-registrants (middle school and high school students), 2004-2007

Variables	Registered for SES? - Middle school						Registered for SES? – High school					
	2004-05		2005-06		2006-07		2004-05		2005-06		2006-07	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
female	0.45	0.45	0.51*	0.46	0.50*	0.43	0.45	0.45	0.52*	0.43	0.50*	0.43
white	0.03*	0.09	0.03*	0.09	0.04*	0.12	0.04*	0.11	0.07*	0.11	0.05*	0.11
black	0.74*	0.64	0.70*	0.61	0.78*	0.70	0.86*	0.55	0.67*	0.48	0.80*	0.62
Hispanic	0.20	0.20	0.23	0.23	0.16*	0.11	0.07*	0.27	0.23*	0.34	0.11*	0.20
Asian	0.05	0.05	0.01*	0.05	0.01*	0.05	0.02*	0.05	0.02*	0.04	0.02*	0.05
Indian	0.004	0.006	0.003	0.006	0.003	0.006	0.004	0.009	0.005	0.01	0.005	0.01
other race	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.02	0.01	0.02	0.01
G.P.A.	2.00	2.02	1.92	1.92	1.97	1.98	1.91*	1.82	2.00*	1.78	1.92*	1.76
took foreign language	0.40	0.43	0.35*	0.31	0.22	0.22	0.25	0.24	0.26	0.26	0.36*	0.29
English proficient	0.86	0.86	0.84*	0.88	0.91*	0.95	0.97*	0.87	0.90*	0.87	0.97*	0.95
free lunch recipient	0.84*	0.79	0.93*	0.86	0.95*	0.91	0.66	0.65	0.81*	0.78	0.87*	0.82
special education	0.21	0.20	0.19	0.21	0.22	0.23	0.19	0.21	0.18	0.20	0.20	0.20
prior year absences	16.7*	19.8	15.7*	21.1	17.7*	20.9	21.8*	30.5	21.8*	31.8	23.8*	31.5
retained	0.46*	0.49	0.53*	0.45	0.38	0.39	0.27*	0.31	0.29*	0.35	0.25*	0.31
prior SES attendance	0.15*	0.08	0.20*	0.10	0.19*	0.10	0.09*	0.03	0.16*	0.05	0.12*	0.07
number of observations	1244	2577	1271	1748	993	1132	1367	7052	1723	4431	2255	5136

* Indicates the difference between SES registrants and non-registrants (in middle or high school in a given year) is statistically significant at α<0.05.

Table 3: 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 (vs. black)				-65%		-57%
Hispanic (vs. black)				-63%		-60%
Asian (vs. black)	-76%	-81%	-72%	-61%	-66%	-72%
Indian (vs. black)						
other race (vs. black)						
grade point average						
grade point average-squared						
foreign language course						+16%
English proficient						
free lunch recipient	+31%	+161%	+54%			+43%
special education student						
total absences (prior year)	-0.8%	-0.7%	-0.9%	-1.1%	-0.8%	-0.7%
retained						
attended SES prior year	+52%	+147%	+92%	+79%	+181%	+88%
grade year 6 (vs. 8)				n.a.	n.a.	n.a.
grade year 7 (vs. 8)				n.a.	n.a.	n.a.
grade year 10 (vs. 9)	n.a.	n.a.	n.a.			
grade year 11 (vs. 9)	n.a.	n.a.	n.a.			
grade year 12 (vs. 9)	n.a.	n.a.	n.a.		-31%	-46%
Pseudo R^2	5.0%	11.3%	10.0%	15.9%	8.9%	7.0%
number of observations	N=2178	N=1683	N=1374	N=7225	N=5207	N=6635

Table 4: 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
2004-05 school year				
Attended any SES 1. unmatched 2. matching	-2.486 (4.562) 2.024 (5.557)	-3.368 (5.232) 3.038 (5.916)	-10.486 (6.243) -5.427 (8.107)	**-14.420 (7.139)** -4.565 (8.860)
Number of observations	N=1562	N=1571	N=1224	N=1262
2005-06 school year				
Attended any SES 1. unmatched 2. matching	-0.529 (0.413) -0.232 (0.427)	0.708 (1.202) 0.323 (1.099)	0.235 (0.297) -0.372 (0.357)	**2.846 (1.132)** 1.397 (1.099)
Number of observations	N=1075	N=1016	N=2215	N=2200
2006-07 school year				
Attended any SES 1. unmatched 2. matching	-0.112 (3.993) 0.595 (4.343)	5.798 (4.566) 4.022 (5.771)	n.a.	n.a.
Number of observations	N=462	N=464	Only 10th graders were tested (N=80), and only 7 of them registered for SES[†]	

[†] In light of the very small number of high school students (of the 2,255 registered for SES in the 2006-2007 school year) who were tested, we do not report SES effects for high school students in this year.

Table 5: Estimated effects of total hours attended SES, 2004-05, 2005-06 and 2006-07 school years

(Standard errors in parentheses; results statistically significant at α ≤ 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
2004-05 school year				
1. # SES hours attended (OLS)	0.046 (0.068) n=1562	-0.017 (0.068) n=1571	-0.127 (0.158) n=1224	-0.148 (0.178) n=1262
SES hours attended (matching)				
2. at least 20 hours (vs. none)	7.727 (5.921) n=1419	5.256 (6.493) n=1428	-24.240 (14.367) n=202	-7.191 (16.262) n=207
3. at least 40 hours (vs. none)	12.757 (7.774) n=1273	11.648 (8.790) n=1282	-26.676 (14.662) n=140	-24.679 (19.241) n=145
If attended > 0 hours:				
4. >10 hours (vs. 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)	**23.093 (10.201)** n=431	16.596 (13.442) n=431	5.888 (13.426) n=185	6.050 (16.675) n=190
6. >30 hours (vs. 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
2005-06 school year				
1. # SES hours attended (OLS)	-0.005 (0.013) n=1075	-0.010 (0.040) n=1016	0.007 (0.011) n=2215	**0.087 (0.042)** n=2200
SES hours attended (matching)				
2. at least 20 hours (vs. none)	-0.055 (0.441) n=366	0.117 (1.672) n=343	0.246 (0.504) n=626	0.328 (1.992) n=595
3. at least 40 hours (vs. none)	0.175 (0.814) n=216	0.547 (2.694) n=200	-1.832 (1.094) n=346	-5.220 (3.546) n=333
If attended > 0 hours:				
4. >10 hours (vs. 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
2006-07 school year				
1. # SES hours attended (OLS)	0.023 (0.190) n=462	-0.045 (0.222) n=464	n.a.	n.a.
SES hours attended (matching)				
2. at least 20 hours (vs. none)	0.310 (7.381) n=370	-4.178 (6.356) n=379	n.a.	n.a.
3. at least 40 hours (vs. none)	-11.199 (22. 639) n=285	-5.820 (11.812) n=289		
If attended > 0 hours:				
4. >10 hours (vs. less<10)	2.359 (8.719) n=462	-6.606 (6.876) n=463		
5. >20 hours (vs. less<20)	-2.501 (7.072) n=462	-5.539 (6.893) n=463		
6. >30 hours (vs. less<30)	0.991 (12.573) n=462	-4.974 (9.849) n=464		
7. >40 hours (vs. less<40)	0.032 (35.163) n=462	10.368 (22.727) n=464		

Table 6: Fixed Effects Model Results for Grade 8-9-10 Cohort, 2004-2006

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

Table 7: Pooled Fixed Effects Model Results for Middle School Cohorts, 2005-2007

| | | Reading | | | Math | |
| | | Standard | Sample | | Standard | Sample |
Grade Cohort	Estimate	Error	Size	Estimate	Error	Size
5 & 6	5.10	(7.93)	4679	-1.12	(6.64)	4680
6 & 7	-1.02	(4.38)	4892	-0.47	(3.89)	4883
7 & 8	-1.80	(4.99)	3848	-5.66	(4.40)	3864
All Middle School Grades	-0.41	(2.19)		-2.49	(1.94)	

Table 8: Means and Standard Deviations of Test Scores by Data Set, Grade, and Year

Grade 8-9-10 Cohort, 2004-2005

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

[a] "Old" WKCE, 2004-2005
[b] TerraNova, 2005-2006
[c] "New" WKCE, 2006-2007

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

Grade	Reading					
	Mean			Standard Deviation		
	2005-06	2006-07	2007-2008	2005-06	2006-07	2007-2008
5	452.9			48.7		
6	464.3	467.5		52.4	52.6	
7	477.0	476.7	477.5	47.8	49.5	50.8
8		494.4	489.6		51.4	54.9
9			492.5			58.3
Grade	**Math**					
	Mean			Standard Deviation		
	2004-05	2005-06	2006-07	2004-05	2005-06	2006-07
5	450.8			42.7		
6	471.6	474.0		43.9	43.2	
7	488.6	497.8	492.7	46.7	43.1	43.4
8		504.8	497.1		46.3	54.8
9			507.1			49.3

AGS Assessment, 2005-2006

Grade	Statistic	Reading			Math		
		Fall	Spring	Gain	Fall	Spring	Gain
6	Mean	429.47	443.41	13.45	490.87	494.77	2.96
	Std. Dev.	28.07	30.49	23.58	10.03	7.48	6.79
	N	607	484	484	678	530	530
7	Mean	441.36	452.13	9.62	493.81	498.05	3.65
	Std. Dev.	25.76	26.64	18.51	8.19	6.41	5.72
	N	808	671	671	898	732	732
8	Mean	452.24	460.33	5.98	495.53	500.08	3.08
	Std. Dev.	26.68	26.99	18.21	10.10	7.45	6.49
	N	1017	826	825	1152	872	872
9	Mean	454.59	465.96	7.57	497.72	500.77	1.93
	Std. Dev.	27.36	27.60	21.47	7.18	7.01	5.82
	N	1950	1112	1112	1795	1105	1105
10	Mean	461.42	471.33	6.73	500.03	502.01	0.95
	Std. Dev.	28.40	28.88	24.59	8.07	7.13	6.18
	N	1443	862	862	1392	891	891
11	Mean	469.59	478.21	4.33	501.78	503.49	0.28
	Std. Dev.	30.15	31.47	24.56	8.95	8.51	6.88
	N	1221	748	748	1275	798	798
12	Mean	479.15	483.29	2.05	503.74	504.23	-0.78
	Std. Dev.	29.27	29.50	20.91	8.59	9.66	8.30
	N	831	419	419	907	395	395

Appendix I

Supplemental Education Services Evaluation Focus Group Protocol

1. <u>Opportunity to Participate in SES Services</u>

1.1   When did you first find out your child was eligible to receive SES services?

1.2   Who notified you of your child's eligibility to receive SES services?

1.3   Did you accept or pursue the opportunity for your child to receive these services?

1.4   If so (or not), what influenced your decision (you and/or your child) to accept (or not to accept) the offer to participate?

2. <u>Choice of SES Provider</u>

2.1   What information was available to you when you and your child were choosing a SES provider?

2.2   Did you request information about any particular SES programs or providers?

2.3   Was the information you received adequate and helpful to you in making a decision about a SES provider?

2.4   Did anyone offer you advice or assistance in making your decision about an SES provider?

2.5   If so, who helped you to make this choice?

2.6   Is there other information that you think would have been useful to you in making your choice, but which was not available at the time?

2.7   If so, what type of information would you have liked to have about SES providers, and why?

2.8   What SES provider did you choose this past year, and why?

3. Experiences and Satisfaction with SES Services

3.1   Did your child stay with the SES provider you or he/she chose for the entire time that he/she received SES services this past year?

3.2   If not, did you go to a different SES provider, or did your child just stop participating?

3.3   How did you make these decisions to keep or change SES providers or to discontinue receipt of services?

3.4   Were you and your child satisfied with the SES services that were provided?

3.5   Please tell me about your experiences with SES services your child received in the Milwaukee Public School district, including what was good and what was disappointing.

3.6   What would you change about the content or delivery of the SES services provided to your child, if anything?

3.7   Is there anything you would change about how the SES services are offered to students or the way the program is managed in the Milwaukee Public School district?

4. Conclusion

4.1   Overall, do you think the SES services were beneficial for your child?

4.2   If yes, in what ways were they beneficial? If not, why not?

4.3   If given the opportunity, would you enroll your child in a SES program again? Why or why not?

4.4   Are there any questions or comments that anyone would like to make?

Appendix II

MILWAUKEE PUBLIC SCHOOLS
2006-2007 Supplemental Education Services
Student Survey

Your school name:

_____

Your grade in school: _____   **Your gender (circle one):   male   female**

Directions:

The following questions ask you about your participation in Supplemental Education Services (SES) during the current (2006-2007) school year. Please answer each question as completely as possible. If you do not know or are unsure of how to respond, you may write "don't know" or "unsure" next to the question.

1. What is (or was) the name of your SES provider in the school year that started September 2006?

   _____

2. How many SES sessions do (or did) you attend in a typical week? _____ sessions

3. How long (in minutes) does (or did) an average SES session last? _____ minutes

Now we will ask you about how you spent your time during the SES sessions you attended. In answering these questions, please think about your activities in a typical or average SES session.

In a typical or average SES session, how many <u>minutes</u> do (or did) you spend:

4. Working *one-on-one* with a tutor in your SES sessions? _____ minutes

5. Working on your own (on your homework or in another self-directed activity)? _____ minutes

6. Working in a group with other students in an activity led by a tutor? _____ minutes

7. Working in a group with other students in an activity that did not involve a tutor? _____ minutes

8. Socializing with other persons or in other activities that did not involve schoolwork or learning? _____ minutes

Now I'm going to ask you some questions about the types of activities you have participated in during your SES sessions.

9. Do (or did) you typically receive help with your schoolwork or homework?
   Circle one: **yes   no**

10. On average, about how much time do (or did) you spend working on your homework in SES sessions? Choose one of the following:
    [1] No time spent on my homework
    [2] Less than an hour
    [3] Less than half of the total time I spent in a typical SES session
    [4] Most of the time I spent in a typical SES session
    [5] All of the time I spent in a typical SES session

11. What subjects do (or did) you work on in your SES sessions?
    [1] Math
    [2] Reading
    [3] Science
    [4] English
    [5] Other _____

12. What school subject do you feel you need the *most* help with in your studies?
    [1] Math
    [2] Reading
    [3] Science
    [4] English
    [5] Other _____ [write subject name; for example, Social Studies]

13. Have the SES sessions been helpful to you in improving your work in the subject you need the most help with
    [identified above in Q. 12]?
    [1] Not helpful at all
    [2] Somewhat helpful
    [3] Very helpful

14. Which of the following things have you done during a typical SES session? (Check all that you have done)
    [1] Taken a test or practice test
    [2] Completed worksheets or assignments given to you by the tutor
    [3] Written an essay or paragraph
    [4] Worked on math problems

15. Do (or did) you do work on a computer in a typical SES session?
    Circle one: **yes**   **no**

    15a. If yes, about how many minutes (on average) do (or did) you spend working at a
        computer in your SES sessions?

        _____ minutes

16. Do (or did) you have specific goals that your tutor set for you to meet in your SES
    sessions?   *Circle one:*   **yes**   **no**

    16a. If yes, have you met your goals? _____

17. Have you received any rewards or prizes from your tutor?
    Circle one: **yes**   **no**

    17a. If yes, what were the rewards, and why did you receive them?
    _____
    _____

18. Have you learned new things in your SES sessions? Choose from the
    following:
        [1] I haven't learned anything new
        [2] I have learned a few new things
        [3] Most of what I have learned was new to me

19. How helpful have the SES sessions been to you in improving your performance or grades in school? Choose from the following:
    [1] Not helpful at all
    [2] Somewhat helpful
    [3] Very helpful

20. If you have not attended one or more of your scheduled SES sessions this school year, what was the main reason you missed these sessions? Choose one or more from the following reasons:
    [1] I had problems getting to or from the location of my SES program
    [2] I had other school activities (sports or other) that were more important
    [3] I wanted to spend time with my friends
    [4] I didn't like the activities we were working on this week
    [5] I didn't think it was worth my time
    [6] Other (please write in): _____

21. Why did you choose this particular SES program or provider to attend this school year?
    _____

22. Did you receive any help in making this choice? *Circle one:*   **yes**   **no**

    22a. If yes, who helped you to make this decision?
    _____

23. What have you learned in your SES sessions this school year?
    _____

24. Is SES helping you in school? *Circle one:*   **yes**   **no**

    24a. If yes, how? For example, has your attendance improved and/or have your grades improved?
    _____

25. What would encourage you to come to the SES sessions more often?
    _____

    26. Was there anything that you disliked about the SES sessions?
    _____

Please feel free to add any other comments in the space below.

Faculty and Research Affiliates

Matthew G. Springer
Director
National Center on Performance Incentives

Assistant Professor of Public Policy
   and Education
Vanderbilt University's Peabody College

Dale Ballou
Associate Professor of Public Policy
   and Education
Vanderbilt University's Peabody College

Leonard Bradley
Lecturer in Education
Vanderbilt University's Peabody College

Timothy C. Caboni
Associate Dean for Professional Education
   and External Relations
Associate Professor of the Practice in
   Public Policy and Higher Education
Vanderbilt University's Peabody College

Mark Ehlert
Research Assistant Professor
University of Missouri – Columbia

Bonnie Ghosh-Dastidar
Statistician
The RAND Corporation

Timothy J. Gronberg
Professor of Economics
Texas A&M University

James W. Guthrie
Senior Fellow
George W. Bush Institute

Professor
Southern Methodist University

Laura Hamilton
Senior Behavioral Scientist
RAND Corporation

Janet S. Hansen
Vice President and Director of
   Education Studies
Committee for Economic Development

Chris Hulleman
Assistant Professor
James Madison University

Brian A. Jacob
Walter H. Annenberg Professor of
   Education Policy
*Gerald R. Ford School of Public Policy
   University of Michigan*

Dennis W. Jansen
Professor of Economics
Texas A&M University

Cory Koedel
Assistant Professor of Economics
University of Missouri-Columbia

Vi-Nhuan Le
Behavioral Scientist
RAND Corporation

Jessica L. Lewis
Research Associate
National Center on Performance Incentives

J.R. Lockwood
Senior Statistician
RAND Corporation

Daniel F. McCaffrey
Senior Statistician
PNC Chair in Policy Analysis
RAND Corporation

Patrick J. McEwan
Associate Professor of Economics
Whitehead Associate Professor
   of Critical Thought
Wellesley College

Shawn Ni
Professor of Economics and Adjunct
   Professor of Statistics
University of Missouri-Columbia

Michael J. Podgursky
Professor of Economics
University of Missouri-Columbia

Brian M. Stecher
Senior Social Scientist
RAND Corporation

Lori L. Taylor
Associate Professor
Texas A&M University

NATIONAL CENTER ON
Performance Incentives

**EXAMINING PERFORMANCE INCENTIVES
IN EDUCATION**

National Center on Performance Incentives
Vanderbilt University Peabody College

Peabody #43
230 Appleton Place
Nashville, TN 37203

(615) 322-5538
www.performanceincentives.org

V VANDERBILT
PEABODY COLLEGE