Equal Access to Quality in Federally Mandated Tutoring: Preliminary Findings of a Multisite Study of Supplemental Educational Services (SES)

Patricia Burch, University of Southern California
Carolyn Heinrich, University of Wisconsin-Madison
Annalee Good, University of Wisconsin-Madison
Mary Stewart, University of Wisconsin-Madison

Presented at the Sociology of Education Conference
February 18-20, 2011
Monterey, California

Current Research Team Members:
Rodolfo Acosta, Martina Chura, Huiping Cheng,
Annalee Good, Angelica Ware Herrera, Nate Inglis-Steinfield,
Kimberly Jones, Hyun Sik Kim, Nidhi Mehrotra, Robert Meyer
Mary Stewart, Renee Timberlake and Anne Ware

Study web page: http://sesiq2.wceruw.org

We thank the funder of this research, the Institute for Educational Sciences, PR/Award number: R305A090301, Education Policy, Finance, and Systems Research Program, Goal 3. In addition, we also express our most sincere thanks to district staff and officials in the study districts—Austin Independent School District, Chicago Public Schools, Dallas Independent School District, Milwaukee Public Schools and Minneapolis Public Schools—for their time, support and dedication to this project. We also thank the SES providers who have opened their doors to this study, and we are especially grateful for the time and support of staff at the Wisconsin Center for Education Research, including Brie Chapa, Bradley Carl, Penny Clark, Mary Fish, Gwen Goplin, Mary Hegge, Jennifer Madigan, Robert Meyer, Steve Ponisciak, Jean Schneider, Chris Thorn and Sandy Treptow.

Abstract

Under No Child Left Behind (NCLB), schools that have not made adequate yearly progress in increasing student academic achievement for two years or more are required to offer children in low-income families the opportunity to receive extra academic assistance, or supplemental educational services (SES). This paper works with ideas from policy sociology and public management to develop an understanding of SES as shaped by a complex interplay of forces. Seen through a policy sociology lens, our research findings challenge the ideal of SES as increasing access to knowledge among the disadvantaged; or through the lens of public management theories on third-party government, they point to deficiencies of highly devolved or dispersed authority in program implementation. In a multilayered policy that requires various systems and actors to interact in new ways and in varied contextual circumstances, program effects are neither simple nor predictable.

Introduction

Under No Child Left Behind (NCLB), schools that have not made adequate yearly progress in increasing student academic achievement for two years or more are required to offer parents of children in low-income families the opportunity to receive extra academic assistance, or supplemental educational services (SES). A pre-existing, well-established market of after-school study and tutoring programs has been joined by an array of diverse organizations—with widely varying hourly rates, tutor qualifications, tutoring session length, instructional strategies, and curricula—to compete for available SES funds and the opportunity to deliver these services to eligible students. The law specifies that 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) [§1116(e)(12)(B)(i)] and should be based on high-quality research with evidence of their effectiveness in increasing student academic achievement [§1116(e)(12)(C)]. In addition, the law requires states to withdraw approval from SES providers that fail for two years to increase student academic achievement.

Since the passage of NCLB, SES typically has been viewed as a marginal component of the No Child Left Behind Act, both in the sense that it is overshadowed by the law's main emphasis on school day accountability and testing and because of the fact that its existence in NCLB is in part a function of back-door, last-minute political compromise. Republicans were pushing for incorporation of vouchers into NCLB, whereby parents would have had the option of using school funds to select a school of their choice. Democrats resisted this effort, with the political compromise being "mini-vouchers," that is, vouchers for eligible families to use to select after school tutoring or SES (Burch Steinberg, & Donovan, 2007).

While certainly a product of a particular political moment and dynamic, the SES provisions of NCLB also are historically rooted in the original conceptualization of the Elementary and Secondary Education Act (ESEA). ESEA is based on the idea of supplemental instruction as a means of improving the quality of instruction for low-income students. Introduced as part of President Johnson's War on Poverty programs, the Elementary and Secondary Education Act of 1965 helped elevate the education of the poor as the nation's priority. The idea was that poverty and education interacted and could have a great and pervasive impact on both individual and collective economic development, as well as the character of society (Goodlad & Keating, 1990). In order to promote academic learning in the nation, the federal government began providing funds for supplemental instruction to students in high-poverty communities. The law promoted the idea that improving academic learning could not occur in the absence of supplemental education, defined as extra instruction functioning in a subsidiary capacity to eligible children's regular school day curriculum.

As part of ESEA's Title I, SES is, in many regards, an extension of the kinds of policy designs we have seen before in federal categorical programs. Given these similarities, some of the early problems in implementing SES follow, in form, common problems in the implementation of large-scale, federal categorical programs. However, in other important respects, the SES provisions of NCLB constitute a new category of policy design, one in which the legitimacy of the private sector is written into policy text via arguments and rationalizations about the need to make the accountability of educational interventions more like that of the free market that governs other types of services (Burch, 2009). This category follows the theoretical angle of neoliberalism: namely the idea that the market can do everything better and that government should be remade in the market's image. Private property rights, free trade,

consumerism, performance audits and entrepreneurs become the means for improving social welfare (Apple, 2006).

Framing Ideas

These particular aspects of the SES policy and its context point us to the use of policy sociology as one conceptual approach to analysis. Policy sociology, as defined here, is an approach to the study of policy implementation that analyzes the intersection of macro *and* micro systems; as we suggest, such an approach may be useful for understanding the positive and negative effects of these kinds of complex, multi-agency education policies. The macro and micro systems for which we see major influences in SES include, broadly, global policy discourse (in particular, neoliberalism), national policy texts (the wording and rules of policy itself) and the micro-level actions of street-level bureaucrats (Weatherly & Lipsky, 1977), which in the case of SES, include individuals representing not only public but also both not-for-profit and for-profit agencies. The design of SES—even the text itself—reflects the asymmetries and negotiation of power across local, regional, national, and international levels, as well as across the public and private sector.

Another useful, related theoretical lens for this research comes from public management theories addressing the growth of "third-party" government and the increasingly central roles of nongovernmental (private sector) entities in the management of government responsibilities and direct provision of goods and services to the public. Public sector reforms in the last two decades have sought to reorient the role of the public sector through "divesting" of government functions and managerial responsibilities to the private sector, with the objective of promoting more efficient, competitive, results-oriented, and responsive services (Heinrich, 2010). In this theoretical context, the implementation of SES is almost entirely the responsibility of third

parties, both nonfederal and nongovernmental, with the explicit intent to give local and private stakeholders the leeway and flexibility necessary to better meet the educational needs of students in underperforming schools. At the same time, these new arrangements significantly restrict the roles of traditional public actors: state educational agencies have limited program design authority, and the flexibility and capability of local educational agencies in managing the market for SES is highly constrained by limited authority, few resources, and little programmatic guidance or experience (Burch, 2009). Oversight and transparency are correspondingly lacking, and responsibility for ensuring that parents and other local stakeholders have the information essential to drive improvements in services is divided and/or neglected. In this regard, policy sociology and public management theories intersect in their uncovering of power asymmetries across levels of implementation and among stakeholders with varying interests in program execution and outcomes.

In our multi-method evaluation of SES, we highlight early effects of these influences on policy implementation, including, for example, the reduction in the autonomy of both low-income parents and local actors in the implementation of the policy. This example, seen through the lens of policy sociology, challenges the ideal of SES as increasing access to knowledge among the disadvantaged; or through the lens of public management theory on third-party government, points to deficiencies in highly devolved or dispersed authority in program implementation.

We continue by discussing key aspects of the policy sociology framework, relating them where relevant to public management theory. Following this, we discuss preliminary findings from our own cross-site evaluation and other study findings on the implementation and impact of supplemental educational services. We then examine these key findings through these

complementary theoretical lenses, focusing on analysis of qualitative data in this study.

Our analysis reveals three primary themes. First, in order to understand and address the policy challenges in SES, we need to see the policy as driven by a particular theory of action—the idea that public policy needs to be designed around and accommodate the marketplace. Second, while NCLB employs the common rhetoric to equalize educational opportunities for the poor, the design of SES may contribute to power asymmetries that sometimes privilege the financial interests of individual private firms and frustrate collective access to information. Expanding the ability of SES to foster equal opportunity requires addressing the ways in which design and discourse of the policy may deny equal access to knowledge for eligible participants and stakeholders. The other key to improving SES lies in the quality of instruction inside of SES classrooms and the relationships between teachers, parents, students and tutors working with eligible students. In short, any effort to improve SES must address contrary or inadequate incentives or other strategies written into the design of current policy. Furthermore, the solutions for improving policy must be anchored in deep understanding of actual classroom/tutoring conditions and the quality of relationships built there. We conclude with a discussion of the implications of our analysis for the design and implementation of SES.

Theoretical Lenses on Policy Implementation

The idea that policy design and implementation is shaped by broader political, economic, and social forces is not a new idea. It has been a central strand of work employing critical social theory to look at education (Anyon, 2006), as well as education policy implementation studies, (Honig, 2006) and more broadly, public management theories of policy implementation (Salamon, 1989). Critical social theory acknowledges that the settings in which policies are implemented matter but extends this work to include analysis of how power shapes policy

responses. Of particular interest in this paper is the concept of cultural political economy, which focuses on the role of economic dynamics as core influences on both the implementation of policy and on its design. From this perspective, the economy is not a separate sphere in the terrain of social issues. Economic processes and identities are embedded in cultural norms and are also central in the "machinery" of policy implementation.

The idea that policies are nested politically, economically and socially and therefore must be studied in context also has been a central strand of educational policy implementation research (Honig, 2006). Second-wave implementation researchers in education, in congruence with public management researchers who focus on front-line policy implementation, argue that the reasons why social scientists could not find effects with large-scale, federally funded reforms include the fact that policy implementation is an ongoing process and that the meaning of policy is contested, shaped and redefined by those who implement it. Rather than something to be controlled for, variables such as district context, student characteristics, family background, and school culture must be investigated in the study and analysis of policy effects. These variables are all part of the context that changes the meaning of the policy and its relevance as perceived by key stakeholders. We carry these lessons into our study of SES, seeking to understand how the actions of district and state administrators, tutors, and parents mediate the quantity and quality of instruction that students actually receive under SES.

While drawing on ideas from critical social theory, public management and implementation studies, another set of ideas—work done in the emergent field of policy sociology—likewise frames our analysis. Policy sociology represents a subfield of policy studies and sociology, coined as an area of study by Ogza (2000) and developed by scholars working mainly outside of the United States (e.g. Ball, 2005; Lingard et al., 2005). Policy

sociology seeks to understand, among other issues, the intersection of macro *and* micro systems. In particular it not only investigates how broader political, social, and cultural forces act on the design of policies, but also how enactment of policies creates opportunities for people working with policies to challenge policy ideas. For social scientists, policy sociology offers a tool that can help knit the big theories together with close examination of policy effects in specific settings and historical moments.

One important premise of policy sociology is the idea that policy can have first- and second-order effects—that is, effects are multi-layered. First-order effects include effects on practices and structures (often the focus of much implementation research). Second-order effects include less obvious or indirect effects on patterns of distribution and access, including who gets what, who is defined as deserving or an appropriate beneficiary of policy, and what is considered "enough" in terms of a treatment. These second-order effects do not necessarily follow first-order effects; they may precede what is measurable in terms of change but also may be difficult to attribute to a particular policy. We use the lens of policy sociology in part because it makes these second-order effects more visible.

Further, in contrast to much traditional educational policy analysis, a policy sociology framework identifies formal and informal cross-level influences on policy. Governments are key authorities in the design and implementation of policy and always will be. However, they do not act alone—a point which much education research chooses to overlook in the narrow definition of policy actors as levels of government with formal authority in design and implementation of policy (e.g., federal governments, state departments of education, local school districts, school boards, schools). Policy sociology views government as only one of a multitude of actors in the complex landscape of educational reform. From this perspective and in the context of SES,

private firms should be considered central in implementation studies both as implementers and beneficiaries of policies in which their role may be explicitly or implicitly named.

One criticism of policy sociology is that it often underestimates the power of the state in policy processes in education. As Held and McGrew (2002: 123) explain, "The locus of effective political power can no longer be assumed to be simply national governments—effective power is shared and bartered by diverse forces and agencies at national, regional and international levels." Public management scholars have raised related concerns about the implications of looser federal control in the face of lengthening chains of delegation and increasing distance between the origins of legislation and the point of service delivery, along with more widely dispersed or "hollowed-out" state and local management capacity and heavy reliance on private sector providers (Heinrich et al., 2010).

With this basic conceptualization deriving from different but complementary theories, we explore how influence over the implementation and practice of SES is shared by and negotiated among both public and private agencies at the local, state, regional, and national levels. In other words, locating U.S. educational policy in broader contexts does not mean simply subsuming the actions of individuals under the logic of the organization or of the market. The challenge we are raising here is the need to move towards a more heterogeneous view of educational policy work (including policy text production and practice), and one that sees the employees of SES firms as nested within national policy agendas and economic fields but also as able to engage in meaningful ways in framing the local practice of policy, in their everyday decisions in administrative work and in teaching.

Finally, in drawing on these ideas from policy sociology and public management, we also encourage the practice of engaging wider audiences with the knowledge that our research

produces, particularly practitioners and other program stakeholders. To do so, we leverage not only theoretical insights, but also our own and others' empirical insights about the goals and outcomes of out-of-school programming, through data collection that simultaneously engages government and provider staff and investigates their roles in SES. (see the Appendix). We now turn to a review of the empirical research on out-of-school time (OST) programs, including best practice findings and a discussion of early findings from our multi-method study of the implementation and effects of SES.

Existing Research

Out-of-school time best practices

Although relatively little research has been done on best practices specific to SES, prior research on out-of-school time (OST) programs in general tells us that quality programs are characterized by particular elements (Durlak & Weissberg, 2007; Little et al., 2008; Vandell et al., 2007; Beckett et al., 2009; Lauer et al., 2006). First, a quality OST curriculum is content-rich, differentiated to student needs, and connected to students' regular school day. Second, instruction is organized into small grouping patterns (no larger than 10:1), and instructional time is consistent and sustained. Furthermore, instructional strategies are varied (both structured and unstructured, independent and collective, etc.), active (not desk time, worksheets, etc.), focused (program components devoted to developing skills), sequenced (using a sequenced set of activities designed to achieve skill development objectives), and explicit (targeting specific skills).

In addition to elements specific to curriculum and instruction, quality OST programs not only hire and retain tutors with both content and pedagogical knowledge, but also provide instructional staff with continuous support and authentic evaluation from their supervisors. The

research suggests the importance of OST programs actively supporting positive relationships between tutors and students, as well as among students themselves. Lauer et al. (2006) also found that effect sizes were larger for programs that were more than 45 hours in duration, although they became smaller for those longest in duration. Although the duration of a program may not be precisely defined as a best "practice," we can consider it an integral part of an ideal program structure.

SES implementation and effects

Although there has been little systematic research on the nuances of the implementation process and actual instructional landscape of SES, since the start of SES, school districts, as well as states, have been under pressure to comply with federal requirements to assess provider effectiveness in increasing the achievement of participating students. School district accountability and evaluation units have attempted to measure program effectiveness, and in some cases, SES provider efficacy; however, there are numerous challenges to properly evaluating student- and vendor-level SES effects that both district staff and researchers face.

First, participation in SES is voluntary among students eligible for SES. NCLB requires school districts to use the same data to determine eligibility for SES that they use for making within-district Title I allocations (historically, information on eligibility for free school lunch), and school districts are required to notify families of their children's eligibility and the availability of approved SES providers. If more students are expected to sign up for SES than there are funds available to serve them, districts have to establish additional criteria to determine which eligible students get access to services. However, even if students are eligible and given the opportunity to register for SES, not all follow through in attending with a chosen provider, and many stop attending before their total SES dollar allocation is expended. Therefore,

selection into "treatment"—or who gets tutored in SES programs and for how long—is influenced by student characteristics as well as program administration and content.

Our research has also shown that it is important to separately analyze the multiple stages of selection—registration, attendance, and the number of hours attended—as the influence of student characteristics differs across them. For example, we find across multiple sites and years that whites, Hispanics, and Asians are significantly *less* likely to register for or attend SES, but if they attend, they are significantly more likely than African Americans to reach higher attendance thresholds. English language learners (ELL students), alternatively, are more like to register and to attend more hours than non-ELL students. Younger children—specifically, elementary school students—are also more likely to attend SES (after registering for a program) and to attend more hours than middle school or high school students (see Burch et al., 2010; Zimmer et al., 2007; Springer et al., 2009).

As we note above, how SES is implemented emerges from complex interplay of policy mandates (the text of policy), interpretations by public and private actors and power dynamics. For example, the number of hours students attend SES is also influenced mechanically both by the rate per hour charged by SES providers and the dollars allocated per-student by districts for SES. In one school district in our multisite study, approximately $1300 was allocated per student for SES, while over 70% of the participating students received SES from a provider charging $75 or more per hour. At this rate per hour and per-student allocation, the maximum hours of tutoring a SES provider could offer a student was about 18 hours. This level of tutoring is far below the minimum threshold of tutoring hours (40-45 hours) that has been identified in the broader literature on out-of-school tutoring programs as critical to producing measurable effects on students' achievement (Lauer et al., 2006). In our own empirical analysis of SES

effects (Burch et al., 2010), we find 40 hours of tutoring to be a critical threshold; below 40 hours of tutoring, we do not identify any statistically significant effects of SES on students' math and reading gains (as measured by changes in test scores).

In light of research findings that consistently show low student attendance levels in SES, it is perhaps not surprising that there is an emerging consensus in the literature of little to no statistically significant effects of SES on student achievement (Burch, 2007; Burch et al., 2010; Heinrich et al., 2010; Heistad, 2007; Springer et al., 2009; Zimmer et al., 2010). Furthermore, in comparison with other interventions targeted to elementary and middle school students, the sizes of any SES effects identified (primarily in math achievement for elementary and middle school students) are substantively small. Hill et al. (2008) compiled effect sizes of similar, supplemental educational interventions, which indicate that the average effect sizes of SES range from one-tenth to one-third the size of effects of other educational interventions for students with similar characteristics.

If an important challenge in getting more hours of SES to students is the hourly rate charged by SES providers, why not cap the rates that SES providers can charge? As currently structured, state and local educational agencies do not have authority to proscribe or control the hourly rates charged by SES providers (other than district-operated ones). Still, a logical expectation would be that providers charging higher hourly rates would be delivering higher-quality tutoring services. However, in our multisite study as well as in prior research (Heinrich et al., 2010), we see little correlation between provider characteristics—such as student-teacher ratios, total hours offered, student attendance levels, curriculum design, etc.—and hourly rates charged, other than whether a provider is on-line. Part of this lack of relationship may be due to discrepancies we see between reported/advertised and actual practices. For example, three

providers may all report 2:9 tutor to student ratios, but those ratios may look different in actual grouping patterns (e.g. 1:8 and 1:1 versus 1:4 and 1:5) in the classroom. Through the lens of policy sociology, these patterns (or lack thereof) raise serious questions about how the design of policy combined with interpretations of a cross-section of policy actors (governmental and non-governmental, and for profit and not for profit) mediate who benefits (and loses) in SES.

Few studies have incorporated rigorous qualitative fieldwork, especially in tandem with rigorous quantitative research, in investigating the effectiveness of out-of-school-time interventions such as SES. However, the few existing studies have found similar challenges in implementation such as: state and district capacity to monitor SES, communication between various stakeholders (states, districts, schools, providers, families), and equal access to appropriate instruction for all students (Burch et al., 2007; Burch & Good, 2009; CEP, 2007; Gill et al., 2008; Potter et al., 2007). The analysis and findings we turn to now draw on early data from our ongoing, longitudinal, multisite study of the implementation and impact of SES. The central purpose of this study is to understand whether and how providing students with academically focused out-of-school tutoring in reading and mathematics contributes to improvements in their academic performance, specifically in reading and mathematics.

Research Design and Setting

Our multi-method study involves three linked phases of research. *Phase 1* is an in-depth qualitative study designed to define key elements of SES program models and the policy and practice variables that mediate implementation of these models and to also inform the construction of the measures of SES treatment for quantitative analysis. *Phase 2* is a quantitative study investigating selection into SES (i.e., who registers and participates) and SES program impacts, using propensity score matching and fixed-effect methods with nonequivalent (internal)

comparison groups. *Phase 3* is a follow-up qualitative study to examine whether program features identified in Phase 1 continue over time and to further inform our interpretation of the quantitative findings of program impact from Phase 2. We are conducting this research in five urban school districts located in four states and representing different student demographics: Milwaukee, Wisconsin; Minneapolis, Minnesota; Chicago, Illinois; and Austin and Dallas, Texas (see the Appendix). A large part of the work in these three phases is taking place concurrently.

We have already briefly described some of the preliminary findings of the quantitative component of this study above. In the quantitative analysis, we examine whether eligible students participating in the program demonstrate gains in math and reading achievement relative to eligible students not participating in SES, and whether and how impacts differ based on students' grade level, intensity (hours) of service, and selection into different kinds of programs and SES providers. In this paper, we are focusing primarily on the qualitative research component that examines questions about SES implementation and seeks deeper insights into conditions at the classroom, provider, district, and state levels that may contribute to SES effects, as well as a more nuanced interpretation of the complex interactions of policy and systems discussed in our conceptual framework. The core study questions include:

(1) How can school districts increase participation in SES by students who are eligible and most likely to benefit?

(2) What factors influence parent or student choices in selecting (and staying with) SES providers?

(3) What are the key characteristics of different program models of SES tutoring, as enacted by providers and as regulated by districts and states, and how do they influence SES program impacts?

(4) What is the impact of SES on student achievement in reading and mathematics?

(5) What are the policy levers and program administration variables that state and local educational agencies and providers can use to increase SES program effectiveness?

The qualitative data were collected from SES provider records, district and state records, interviews with district and state administrators, interviews with directors of tutoring programs, interviews with tutoring staff, parent focus groups, and observations of tutoring practice (see the Appendix for additional details).

Beyond Effect Sizes: What is in an Hour of SES?

Although test-based accountability and evidence on "what works" are at the core of the NCLB reauthorization, the intent of SES was somewhat different: to facilitate as extensive a choice as possible for students and parents in selecting providers and program types. Under the law, school districts cannot impose program requirements on providers; they only have authority to terminate a provider's contract when the provider violates district policies (e.g., building use) or other such legally binding agreements. District staff responsible for SES contend that their hands are tied in monitoring providers and also point out that most SES tutors do not have to meet "highly qualified" standards or have specific or sufficient training to be academic tutors. District staff (corroborated by provider staff) also feel that some state educational agencies have been lax in evaluating providers and setting minimum standards for tutoring quality and have failed to request essential information on applications for assessing and monitoring quality or to follow through on district complaints about provider incompetence or misconduct. With very few resources for program administration, let alone monitoring and evaluation, district staff have been stretched to find time to observe SES providers and better understand what is taking place in an hour of SES for which districts are invoiced.

A distinguishing feature of our multisite, multi-method study of the implementation and effectiveness of SES is an in-depth qualitative component designed to define key elements of SES program models and to identify how policy and implementation potentially mediate or influence SES impacts. Put differently, what do we see happening in an invoiced hour of SES; how does this vary across different SES provider settings, formats, and approaches to tutoring; and how does it relate to program effectiveness?

Assessing program fidelity in implementation is one method of evaluating the "success" or compliance in program implementation; we begin our qualitative analysis with this type of assessment. While the law intentionally offers SES providers wide-ranging flexibility in the design of their programs, it also specifies information regarding provider contracts with districts and some directives for content focus, location, and the use of research-based practices: SES providers must make reading and mathematics the content focus of instruction, and instruction must be provided outside of the regular school day. Providers are not required to offer services to students with disabilities (SWD) or English language learners (ELL), but if providers do offer these services, the law requires them to be advertised, and districts are responsible for providing these services if no provider is able or willing to do so. In addition, the law states that SES tutoring should be "high-quality, research-based, and specifically designed to increase student academic achievement" [Section 1116(e)(12)(C)]. In our study, we interpret "research-based" practices as the best practices identified as making a measurable impact on student achievement by the out-of-school-time (OST) literature. The following analysis evaluates whether SES in practice is faithful to these basic requirements.

Best practices for out-of-school-time tutoring and their use in SES

Although there is little research on best practices specific to SES, prior research on OST programs generally tells us that high-quality programs are characterized by: (1) consistent and sustained instructional time; (2) small grouping patterns (no larger than 10:1, but smaller is better); (3) curriculum that is content-rich, differentiated to student needs and connected to students' regular school-day learning; (4) instruction (or content delivery) that is varied (e.g., structured and unstructured, independent and collective, etc.), active (not desk time, worksheets, etc.), focused on skills development, sequenced to achieve skill development objectives, and explicit in its targeting of specific skills; (5) positive relationships between tutors, students and peers; and (6) teachers/tutors with both content and pedagogical knowledge and continuous support, as well as constructive evaluation, from their administrators. To identify these best practices in SES sessions, Burch, Good, and colleagues designed a standardized observation instrument to systematically collect information on teaching methods and instructional materials in use and to identify the correlation of different formats, resources (curriculum materials, staffing, etc.), and instructional methods on students' observed levels of engagement.

In year one (2009-2010), field researchers conducted observations of 56 tutoring sessions in five, large urban school districts in four states across a range of provider characteristics—including on-line, in-home, in-school, and community-based tutoring; for-profit, not-for-profit, district-provided, and faith-based organizations; providers with large market share (in terms of students served), two or more years of SES provision, and with higher than average levels of student attendance; and providers advertising services to SWD and ELL populations.

In general, the model of tutoring we observed tended to take the form of traditional academic learning environments, with students being tutored in tested subjects—mathematics and reading—and typically in the form of whole group instruction (one focal activity). While this

focus on the "basics" is faithful to the explicit SES policy, our observation data does not provide evidence of predominantly innovative, active learning, nor activities complementary to the regular school day. Instead, SES was largely based on traditional forms of teacher-directed instruction and isolated from the students' school-day instruction. Thus, regarding content and instructional models, observation data offers an initial, conflicting picture of "quality": fidelity to the policy but lacking in characteristics related to student achievement in out-of-school time.

Further, research on OST argues, without qualification, for differentiated programming that responds to students' different learning styles or needs. In our observations, students attending SES who might learn best via project-based learning, arts integration, or links to community-oriented activities encountered few opportunities of this sort. Perhaps most troubling, however, was the fact that we encountered very few tutors with training or experience in instruction differentiated to ELL or SWD, and with very few exceptions, neither curriculum nor instruction were tailored to the unique needs of these students. This is particularly disturbing given that as suggested by our quantitative data, ELL students are most likely to register and attend sessions. It appears that high demand for ELL-appropriate services in the SES market is not being adequately met.

Access to Learning Opportunities

The observation instrument also allowed for assessment of continuity of practice across an entire session and how much *instructional* time students actually received. Irrespective of the format, students received less instructional time than what was advertised by providers, although the magnitude of these differences varied by format and by district. In more than half of all observations with two or more students (primarily off-line, school-based settings), students that started a session were observed arriving late, leaving for part of the session, or leaving the

tutoring session altogether (what we call "attendance flux"). Through interviews with tutors and provider administrators, we confirmed that school-based SES programs often compete with other after school programs (e.g., athletics, clubs) for students' time, and classrooms with multiple students required coordination and set-up that cut into instructional time.

Regardless of the reasons for attendance flux or other barriers to instructional time, the quantitative and qualitative findings of our multisite study together suggest that students are not getting enough hours of high-quality, differentiated SES instruction to produce significant gains in their learning, and given that invoiced hours may not equal instructional time, this is not a problem that will be resolved *only* by setting minimum hours standards for SES providers. We elaborate on this recommendation in our concluding remarks.

What is Advertised Versus What is Offered

Advertised time does not always equal instructional time, and sometimes even invoiced time differs based on our preliminary analysis. These differences tend to vary by format. In on-line organizations, instruction started the minute the student came on-line and ended promptly at the end of the (typically) hour-long advertised time. The tight focus on instructional time also was observed in home-based settings. Instruction started immediately after the tutor arrived and ended on average five or six minutes early, leaving time for collection of materials, record-keeping, and departures. In school and community settings, instructional time was bookended by classroom management activities or logistics such as transportation (25 minutes in the case of school-based tutoring and approximately 45 minutes in the case of community based tutoring). However, some providers, for example, advertise their flexibility as a selling point to parents with special logistical requirements, and we may see this as one reason for discrepancies in advertised versus instructional time. On the other hand, some observations in the community-

based setting did reveal less organization and classroom management. Overall, the range of grouping patterns, instructional approaches, and other details of community-based observations were too broad to make conclusions regarding instructional versus advertised time in this setting.

Further, advertised services do not always equal access to those services. According to providers' advertised services, 14 out of 20 providers in our sample advertised that they could serve ELL students, at least in a limited way or for limited languages. Thirteen out of 20 (though not necessarily overlapping the 14 mentioned above) providers advertised that they could serve students with disabilities (SWD), at least in a limited way or for limited special needs. Here we also see a major implementation and evaluation obstacle for providers, tutors, and researchers: identifying students with documented ELL or special education needs. The majority of tutors we observed and interviewed did not have access to IEPs or district data on ELL identification. If they did, it typically was because they happened to be a teacher at the school site where tutoring took place. Therefore, our observations of instruction for such subgroups may have limited ability to project patterns in the appropriateness of implementation (i.e., identification of students to providers and tutors) or instruction. However, in a few cases there was evidence of tutor knowledge of ELL and special needs students. The following collection of vignettes illustrates the ways in which students in these subgroups were and were not served.

Example: A composite vignette of providers serving ELL students in one site:

> Providers and tutors discuss difficulty in communicating with students and/or their parents who speak less common foreign languages (e.g., Somali, Vietnamese). Several Spanish-speaking tutors reported using Spanish and English to instruct and clarify; with non-Spanish-speaking families, tutors used a variety of strategies. Some bilingual (Spanish-English) tutors used their informal knowledge of language learning to check

vocabulary (e.g., Tutor: "What is this word? 'Gen-er-a-tion.' What that means is, you have a son, [starts drawing a diagram/picture] a father, a grandfather, a great-grandfather…" "Grandmother," adds the student). Another strategy was asking the student for definitions (e.g.: The student reads the sentence with the next vocabulary word in it. "I don't know what this means." Tutor: "Deduce, use context. What's a 'hit' in baseball?" Student: "Like you hit someone with the ball?" Tutor: "Well…kind of close but not quite. Let's keep thinking." Student: "Hit means in baseball to use your bat and make an inning." Tutor: "We need to have something like 'to connect with the ball.'"). Since nearly all tutors had only informal language training, if any, accurate knowledge of students' fluency levels was sparse and checks for understanding were sporadic.

Particularly regarding students with disabilities, providers did not have sufficient information to appropriately identify these students. Most providers only knew of students with disabilities because their tutors also were teachers in the regular school or parents notified them. From the providers' perspective in our sample of districts, the district administrators did not have a systematic process to provide this information. This could be the result of legal issues with sharing IEP information. With a few exceptions, tutors did not have specific training or certification in working with students with disabilities. It should be noted that we observed many sessions with certified teachers as tutors. Most of these tutors would have had training related to special education as part of their certification process and in many cases considerable experience working with students with disabilities in their regular classrooms. However, we did not find any examples of curriculum specifically formatted to accommodate the particular needs of students with disabilities. Existing curriculum was sometimes "slowed down" for these students, or a lower grade level curriculum was used.

Example: A composite vignette of providers serving students with disabilities in one site:

> Regarding students with disabilities, students with ADD/ADHD seemed appropriately served in one-on-one sessions, as discussed with tutors and observed in on-line sessions. However, easily distracted students (whether or not identified as ADD/ADHD) did not tend to receive appropriate attention or behavior management during group sessions (e.g.: "Tutors were actively helping the two groups of girls, who were working relatively quietly, carefully, and intently, but no one was sitting with the group of 3 boys. Two of the boys were messing around on other computers, which the lead tutor of the lesson finally put a stop to about halfway through SES time. . . By this point most of the other groups were nearing completion of the task"). Disabilities other than ADD/ADHD were not observed, perhaps because of the fact that many tutors did not know for sure whether their students had disabilities. However, licensed special educators and educators with special education experiences (not necessarily licensed) described using their knowledge and experience of differentiated instruction in sessions with *all* of their students.

Some Evidence of Good Teaching

Despite the lack of appropriateness and quality in ELL and special needs instruction, we did find evidence of tutors actively employing practices identified by OST research as supporting learning gains that one could identify as minimal characteristics of "good" teaching. In particular, tutors used materials towards the goal of instruction in mathematics and reading language arts. This indicator, described in our instrument manual as "Staff use available materials for the purposes of instruction on stated skills/goals," was rated highly (>.8) across all districts and formats. Tutors were also observed engaging positively with students (>.75) across all districts and formats; as described in our manual, engaging positively means that:

> Staff have positive interactions with students. These interactions are constructive and supportive. Staff use affirming words and tone of voice, speaking in a manner that indicates respect, appreciation, and belief in the value and potential of students. Staff initiate informal conversations with students and respond to students' efforts to talk to them by showing interest and extending the conversation. Staff make an effort to build relationships with the students through a variety of means. Staff also move around to student workspaces, instead of staying in one place (i.e., their desk) the entire session.

Overall, tutors were observed as engaging with students in a predominantly positive way (that is, tutoring sessions had high ratings on a variety of related indicators such as "Provide constructive criticism"; "Encourage participation from disengaged students"; and "Listen actively and attentively to students").

Further, while much of SES tutoring reflected traditional schooling practices, there were some important differences. One major difference between tutoring and the regular classroom was grouping patterns. Structured academic support in a one-on-one setting is typically impossible in regular classrooms, unless the student is pulled out of the classroom. In our observations of SES tutoring, one-to-one (29/56 sessions) or one or two-to-three (21/56) instructional support was available in the majority of sessions. No tutoring sessions had a group of students larger than nine.

Limited Information for Parents

SES is based on the assumption that improving public schooling hinges in part on giving low-income parents the opportunity and choice of a vendor of OST supplemental instruction. However, based on preliminary analysis of focus group data, some SES vendors, like large school bureaucracies, can be hard for parents to access when they are seeking information or

have complaints. In focus groups, parents reported that they lacked a clear understanding of their options under SES. Across districts, parents participating in the session were surprised to learn that they had options when it came to choice of provider. They were eager to obtain tutoring for their child and went with the first vendor that contacted them, unaware that there were other vendors that might be a better fit for their child in terms of scheduling, focus, and format. This was the case in Austin, where parents voiced that they knew little of and about SES. In Milwaukee, one of the parents mentioned that she had liked the services offered to her child but found out half way into the program that math services were also available. She reported that if she had known about math services, she would have ensured that her child had participated in them as well.

Parents' comments also reveal how the challenges of being a recipient of government services can be exacerbated when multiple private firms and the state are involved in the design and delivery of those services. There is a lot of information sent to parents at once; it is hard to process or know where it is coming from. A parent in Chicago stated, "They could have actually broke it down and gave more detail about why they chose this one, that one, and that one for your child. You know, instead of just having to decide on your own." Parents' experiences corroborate the trend noted by district, school, provider, and tutoring staff that lack of communication among stakeholders is a major barrier to successful implementation.

Further, across districts, parents did not feel that they had appreciable control over the quality of the structure of the tutoring program beyond removing their children from the program or confronting individual SES instructors. For example, in Minneapolis, some of the parents opted to remove their children from services because providers were inflexible with scheduling or because they witnessed poor professional decorum on the part of providers.

Overall, parents felt that the tutoring services had the potential to be a positive influence on their children's educational experiences. In contrast to most district administrators, parents saw and reported direct and positive experiences for their children as a result of the program and pointed to both tangible (improvements in grades, increases in test scores) and less tangible (engagement in school, students sense of self efficacy as learners) effects. As a whole, they felt that additional instruction outside of the school could only benefit their children and welcomed an intervention with that purpose. On the other hand, many parents felt that there were a number of problems that needed to be worked out before the programs could have their desired effect. For example, one suggestion voiced throughout the focus groups was to improve communication between vendors and school personnel in the hopes of helping to prioritize and select students who would benefit from SES the most. As a whole, parents felt that their voiced concerns (some of which might be turned into interventions) could only be made on the part of their individual children and would not have an appreciable effect on the policy design and program overall.

Discussion: Equal Opportunity to Quality After-School Programming

Early Lessons from the Implementation and Impact of Supplemental Educational Services

As we have seen, the intensity (hours) of the SES intervention is directly tied to the hourly rate charged for service provision. The more providers charge, the less likely are students to receive the level of tutoring (in terms of hours) that is a key predictor of program effectiveness. However, even among those receiving a level of tutoring necessary to generate effects, the magnitude of these effects is modest as gauged by effect sizes for similar kinds of interventions. Preliminary findings from the qualitative analysis suggest that the lack of effects or minimal effects may stem from the actual amount of tutoring time used towards instruction as well as critical omissions in the quality and character of instructional programming. Among these

critical omissions is programming to address the needs of ELL and SWD students—a problem made more troubling by evidence that at least in the case of ELL students, these students are signing up and attending SES at higher rates than other students.

However, with these significant limitations in mind, under certain conditions we see some positive outcomes for some participating students as measured by changes in their mathematics and reading test scores (particularly for elementary school students). In addition, in observations of practice, we saw preliminary evidence of structures and practices (i.e., format of tutoring, grouping patterns, clustering of instructional practices) that appear to be linked to student engagement.

That said, if there is one agreed-upon and consistent pattern in educational policy research, it is that "we have learned that there are few slam bang effects" (McLaughlin, 1991, p. 15). Lessons from over three decades of policy implementation research argue that this is the case for at least two reasons. First, there is a mismatch between the design of policy and implementers' own incentives and understanding of policy. The disconnect partly stems from the top-down nature of much policymaking, and the idea that "policymakers should develop policies for implementers to carry out and monitor implementers' compliance" (Honig, 2006, p. 3). Second, federal education policies tend to be very explicit in specifying the end goal, but intentionally vague in specifying the "how," and also short on resource commitments (financial and administrative) in supporting implementation.

We see continued evidence of this problem in the design and implementation of supplemental educational services. Key implementers in the case of supplemental educational services include district and state administrators *and* directors and employees of not-for-profit and for-profit firms contracting with the district to provide services. NCLB and SES offer only a

few specific regulations around standards for tutoring. While providers are offering instruction in reading and mathematics, they also are clearly implementing programs based on their own perspectives on what is needed and/or based on the firm's interest as a business. They are charging more or less, tailoring curriculum to meet the needs of particular populations or not, and making decisions about what program characteristics are most important (e.g., tutor qualifications, setting for tutoring, length of session) in ways that reflect the mutual adaptation dynamics emphasized as a core element of implementation research.

In these ways, SES represents some classic problems of policy implementation. The problem of SES is born both of the messy political nature of implementation *and* the design and language of the policy. However, other implementation problems are harder to explain using the classic lens of public (that is, government-based) policy implementation, leading us to draw from policy sociology in understanding and accounting for these challenges.

For example, federal law prohibits districts or states from prescribing or encouraging particular curricular approaches among private firms, consistent with the neoliberalism vision of a minimal role for the state. The design of SES frames the role of the government as the central barrier to addressing the achievement gap—a claim that has helped set the terms of the debate — leading many local administrators to view the policy as an aggressive move that they must resist. This aspect of SES is in line with Stephen Ball's ideas that much of the power of policy lies in its ability to create a set of circumstances and dialogue that limit or expand the range of options for implementation. Put differently, the design of policy establishes rules of the game under which the power of some is enhanced and the power of others reduced.

In addition, SES is much more specific than typical education policy about what local and state administrators can and cannot do. Local government entities are constructed as

secondary implementers who, *by design,* are given very few resources and little discretion in either the type of instruction offered in their districts or any real authority when or if providers in their district fail to comply with requirements. Thus, under SES, local and state administrators face different policy challenges than they might with other kinds of policy. As in the case of other policies, they are still managing multiple demands and operating under conditions of limited capacity and information. However, under SES, their power to implement policy in ways consistent with goals of student improvement is curtailed by the explicit design of the policy. For example, as noted above, the law and its regulations forbid states from using accountability strategies that have been core policy instruments in federal programs including NCLB, such as standard setting and high stakes testing. Specifically, districts and states cannot require providers to utilize a common standardized test as a measure of their own effectiveness in order to discern which providers or methods work best in increasing student achievement.

This aspect of the design points to a central contradiction in the law. Although the policy explicitly attempts to equalize the playing field for low-income students to obtain supplemental academic help, in reality the policy does not address unequal power dynamics regarding access to information and political leverage. The rhetoric of NCLB is that all eligible students, particularly those who have been underserved in the past, should have equal access to programming. Knitted to the idea of equal opportunity in NCLB, however, is the ideology that access is enabled in large part through freedom of choice. Extending educational opportunity in the area of out-of-school time means giving poor urban families in eligible schools the opportunity to participate in OST programming free of charge and the right to choose a vendor for their child, as opposed to having the district choose for them.

As our research indicates, families are taking advantage of this opportunity. However, while generating more support for the idea that poor families have the right to a better education, the law ignores or perhaps even reinforces a power dynamic that works against families taking full advantage of this right—that is, poor families' inadequate access to information and their lack of effective levers for holding providers *and* districts *and* states accountable for their responsibilities. As in the case of other education policies, and we would argue particularly in the case of SES, "some parties have more power than others in determining the course of implementation" (Dumas and Anyon, 2006, p. 165). These parties are those with resources of time, organization, and legal and financial knowledge, which are characteristics of providers and school districts much more often than of low-income parents.

These complexities and contradictions within the design of SES exemplify the traits of a new kind of education policy, one in which political power can no longer assumed to be national, state or local governments but includes various private actors as well. Under SES, however, the role and authority of government does not disappear. Instead, local governments are expected (as in a marketplace) to barter and compete with private firms at the local, national, regional, and international level. In our own mixed-method study, Good et al. (2011) document vividly, through the voices of tutors and teachers, how the costs of some of these conflicts are born most heavily by both teachers and families—the least powerful and the most vulnerable groups involved in the policy. For example, under conditions where parents have limited recourse and authority around activities of providers, providers can cancel services mid-year if and when they deem it is in their financial interest. In our above findings, we describe how some providers are using the flexibility written into the law to charge high rates, perhaps at the expense of quality instructional characteristics, such as a minimum number of hours.

These patterns require that we take account of how policy designs create power asymmetries not only between new categories of policy actors (public sector and private sector) but also *within* these new categories of actors. Within the implementation of SES, the production of policy involves actors at different levels of the firm, with different kinds of power. In our work, we pay close attention to the actions of street level *private sector* implementers. As demonstrated by our qualitative work, it is an oversimplification to say that profit drives the decisions of private sector firms in SES. In particular, in many instances, the decision and perspectives of tutors in for profit companies reflect the logics of best practice from the out-of-school literature. Under difficult conditions, these tutors and directors are creating learning settings that, based on existing evidence from OST, further the interests of students. As well, small local providers are at a disadvantage with differential access to resources that allow them to build market share in SES. They also have differential access to opportunities to demonstrate effects: they cannot hire outside firms or depend on an internal research department, and they usually do not have the number of students to create a reliable sample size.

While the policy is designed to effect changes in district practice, preliminary analysis points to intermediary effects on eligible students' access to resources, defined here in terms of access to highly qualified tutors, differentiated curriculum for English language learners and students with disabilities, and intense interventions (the threshold of hours that existing research on out-of-school time defines as a basic condition for effectiveness). As argued above, the original ESEA was designed to equalize opportunities in education via supplemental funds. However, based on our preliminary analysis, an interlocking set of policy arrangements (including limited capacity on the part of district and state, implementation exceptions within the policy—e.g., no requirements for providers to serve ELL or SWD—and conditions both inside

schools and on the part of providers) create barriers to students' access to the learning that are unintended in SES. As noted above, it is encouraging that ELL students are more likely than non-ELL students to sign up and register for SES. However, based on our qualitative analysis, the quality of tutoring that they receive during sessions is clearly inadequate. Even among well-meaning and compassionate staff, and among providers stating under contract that they could and would serve students with special needs, huge barriers to learning for these subgroups existed. Seldom did local government officials examine tutoring programs for evidence of equalized access to quality tutoring opportunity for these groups or for instruction aligned with a student's IEP, as is their responsibility [Section 1116(e); 34 C.F.R. §200.46(a)(4); 34 C.F.R. §200.46(b)(3)].

Implications

Leveraging the empirical findings and theoretical insights noted above, what most deserves our attention? In this last section, we offer a few recommendations to direct the time and energy of researchers, practitioners, and policymakers and to better allocate our limited financial resources.

An immediate logistical issue related to equity and program effectiveness is the problem of prioritizing students when SES programs are oversubscribed. We need to explore the implications of districts' various strategies for targeting SES to their eligible students—such as prioritizing those with very low academic performance in addition to those with low incomes—and consider how these changes influence equity in access, programming, hours attended, as well as the effectiveness of SES.

We should also reconsider the rationale behind and the serious implications of allowing providers to fully determine their hourly rates and instructional strategies. The evaluation of

effects in this study and others clearly point to a minimum threshold of tutoring hours after which tutoring appears to produce measurable effects on student achievement. Federal lawmakers should reconsider allowing states and districts to cap the hourly rates providers can charge. In addition, the interactions and relationships among instructional variables are fundamental to intervention quality, including student grouping patterns, location, time spent on instruction, attendance flux during sessions, and student engagement and patterns of OST best practices. The level of instructional differentiation towards students with special needs (i.e. ELL and students with disabilities) warrants particular attention. This is a critical, and in our examination, neglected piece of tutoring program quality.

Finally, researchers need to focus significant energy on the processes of, and state capacity for, accountability mechanisms (approval and monitoring). Many of the variables in our study—curriculum, instructional strategies, tutor qualifications, attendance—are included in specific sections of state applications; however, we have little evidence that these variables are part of the monitoring process. If we are to accurately evaluate the design and implementation of SES, we must have a better idea of where the weak (and strong) links lie in the accountability system (as currently designed).

References

Apple, M. (2009). Series editor's introduction. In P. Burch, *Hidden markets: The new education privatization* (ix). New York: Routledge.

Ball, S.J. (2006). Policy sociology and critical policy research. In S.J. Ball (Ed.) *Education policy and social class: The selected works of Stephen J. Ball*. New York: Routledge

Beckett, M., Borman, G., Capizzano, J., Parsley, D., Ross, S., Schirm, A., & Taylor, J. (2009). *Structuring out-of-school time to improve academic achievement: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

Burch, P., Heinrich, C., H. Cheng, A. Good, M. Stewart and R. Timberlake. 2010. Research Briefing: Preliminary Findings of a Multisite Study of the Implementation and Effects of Supplemental Educational Services (SES). Wisconsin Center for Education Research.

Burch, P. (2009). *Hidden markets: The new education privatization*. New York: Routledge.

Burch, P. & 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., Steinberg, M., & Donovan, J. (2007). Supplemental educational services and NCLB: Policy assumptions, market practices, emerging issues. *Educational Evaluation and Policy Analysis*, 29(2), 115-133.

Center on Education Policy. (2007). *Behind the numbers: Interviews in 22 states about achievement data and No Child Left Behind Act policies*. Washington, D.C.: Center on Education Policy.

Debray, E. (2006). *Politics, ideology, and education: Federal policy during the Clinton and Bush administrations*. New York: Teachers College Press.

Dumas, M. & Anyon, J. (2006). Toward a critical approach to education policy implementation: Implications for the battlefield. In M. Honig (Ed.), *New directions in education policy implementation* (149-169). Albany, NY: SUNY.

Durlak, R., & Weissberg, R. (2007). *The impact of after-school programs that promote personal and social skills*. Chicago: CASEL.

Gill, B., McCombs, J., Naftel, S., Ross, K., Song, M., Harmon, J., Vernez, G., Birman, B., Garet, M. & J. O'Day. (2008). Title I, school choice and supplemental educational services under No Child Left Behind: Progress towards implementation. Santa Monica, CA: RAND Corporation.

Good, A., Burch, P., Stewart, M., Jones, K., Mehrotra, N., & R. Timberlake. (2011). *Instruction matters: Lessons from a mixed method evaluation of supplemental educational services under No Child Left Behind*. Working paper presented at the 2011 American Educational Research Association Conference, New Orleans, LA.

Goodlad, J.I. & Keating, P.M. (1990). (Eds.). *Access to Knowledge: An Agenda for Our Nation's Schools*. New York: The College Board.

Heinrich, C. J. (2010). Third-party governance under No Child Left Behind: Accountability and performance management challenges in supplemental educational services provision. *Journal of Public Administration Research and Theory,* 20(1): 59-80.

Heinrich, C.J., Lynn, L.E. & H.B. Milward. (2010). A State of agents? Sharpening the debate and evidence over the extent and impact of the transformation of governance. *Journal of Public Administration Research and Theory,* 20(1): 3-19.

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

Heistad, D. (2007). *Evaluation of supplemental education services in Minneapolis Public Schools: An application of matched sample statistical design*. Minneapolis: Office of Research, Evaluation and Assessment, Minneapolis Public Schools.

Henig, J. R. (2007). The political economy of supplemental education services. In F. M. Hess & C. E. Finn, Jr. (Eds.), *No remedy left behind: Lessons from a half-decade of NCLB* (66-95). Washington, DC: American Enterprise Institute Press.

Hill, C.J., Bloom, H.S., Black, A.R. & M.W. Lipsey. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives,* 2 (3), 172-177.

Honig, M. (2006). Complexity and policy implementation. In M. Honig (Ed.). *New directions in education policy implementation* (1-23). Albany, NY: SUNY).

Lauer, P.A., Akiba, M., Wilkerson, S.B., Athorp, H.S., Snow, D. & 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.

Lingard, B, Rawolle, S. & Taylor, S. (2005). Globalizing policy sociology in education: Working with Bourdieu. *Journal of Education Policy,* 20(6), 759-777.

Little, P., Wimer, C., & Weiss, H. (2008). *Programs in the 21st century: Their potential and what it takes to achieve it*. Cambridge: Harvard Family Research Project.

McLaughlin, M.W. (1990) The RAND change agent study revisited: Macro perspectives and micro realities. *Educational Researcher,* 19(9), 11-16.

Ogza, J. (2000). *Policy research in educational settings: Contested terrain*. Berkshire, UK: Open University Press.

Potter, A., Ross, S., Paek, J., McKay, D., Ashton, J. & A. Sanders. (2007). *Supplemental educational services in the state of Tennessee: 2005-2006.* Memphis: Center for Research in Educational Policy.

Salamon, L. (1989). *Beyond privatization: The tools of government action*. Washington, DC: The Urban Institute Press.

Springer, M.G., Pepper, M.J. & B. Ghosh-Dastidar. (2009). *Supplemental educational services and student test score gains: Evidence from a large, urban school district.* Working paper. Nashville: Vanderbilt University.

Vandell, D., Reisner, E., & Pierce, K. (2007). *Outcomes linked to high-quality afterschool programs: Longitudinal findings from the study of promising practices*. Irvine, CA: University of California; and Washington, DC: Policy Studies Associates.

Weatherly, R & Lipsky, M. (1977). Street level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Education Review*, 47(2), 171-197.

Zimmer, R., Gill, B., Razquin, K. Booker, K. & 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.

Zimmer, R., Hamilton, R., & R. Christina. (2010). After-school tutoring in the context of no Child Left Behind: Effectiveness of two programs in the Pittsburgh Public Schools. *Economics of Education Review,* 29, 18-28.

Appendix: Research Design and Methodology

Quantitative Methodology

We estimate effects of SES as changes in students' standardized math and reading scores, by the different levels of SES attended, and by district and elementary, middle and high schools. The sample of districts and students used in this analysis is shown in the table below. We use propensity score matching methods to account for selective differences between those who register for SES but do not attend and those who attend lower vs. higher levels of SES. Specifically, we estimate the effects of SES attendance at common peaks (20, 40 and 60 hours attended), separately matching students with attendance levels above and below each of these points and only reporting results where balance and adequate support/comparability for the matches is achieved.

Quantitative Sample Summary

Number of Students Eligible, Registered and Attending SES by School Year and District

2007-08 School year			
	SES-eligible	*Registered*	*Attended SES*
Milwaukee	8284	3704	2194
Minneapolis	9217	1345	1124
Chicago	166386	46856	37095
Dallas	24031	6179	4632
Austin	6297	489	212
2008-09 School year			
	SES-eligible	*Registered*	*Attended SES*
Milwaukee	11992	4267	2689
Minneapolis	10618	2567	1412
Chicago	212504	69073	56921
Dallas	36770	8619	4911
Austin	7330	2761	1929
2009-10 School year			
	SES-eligible	*Registered*	*Attended SES*
Milwaukee	26798	6933	4998
Minneapolis	16484	4910	3533
Chicago	135418	65531	33465
Dallas	30916	10950	10637
Austin	4470	2986	1306

Qualitative Research Design

Implementation findings are based on data collected from provider records, district and state records, interviews with district and state administrators, interviews with directors of tutoring programs, interviews with tutoring staff, parent focus groups, and observations of tutoring practice. Sample sizes and further details on the data sources are below:

- *Observations of tutoring sessions* (n=56) using the a classroom observation instrument (described below) designed to capture key features of instructional settings
- *Interviews with provider administrators* (n=45) about structure of instructional program, choice of curricula and assessments, challenges in implementation, and choices in staffing
- *Interviews with tutoring staff* (n=64) about instructional formats, curriculum, adaptations for special student needs, staff professional background and training
- *Interviews with district and state administrators* (n=19) involved in program implementation
- *Parent focus groups* (n= 174) with parents of students who were eligible to receive SES, most with children currently receiving SES; two focus groups of approximately 1.5 hours each were conducted in each site and translation was offered in Spanish, Hmong and Somali
- *Document analysis:* formal curriculum materials from providers, diagnostic, formative, or final assessments used, policy documents on federal, state or district policies concerning the implementation of SES

These sample sizes (n) are cross-site and for the 2009-10 research year, upon which most of the analysis in this paper is based.

We developed a standardized observation instrument in order to more accurately capture the nature of the SES intervention. The instrument has the capability of not only providing descriptive information on instructional materials and teaching methods in use but also detecting the impact of different kinds of format, resources (curriculum materials, staffing, etc.), and instructional methods on students' observed levels of engagement. The observation instrument includes indicator ratings at two 10-15 minute observation points as well as materials collection, a rich description in the form of a vignette, and follow-up information provided by the tutor(s).

Sample Selection for Qualitative Study

Limitations include reluctance on the part of providers; low numbers of providers with more than one year of service in smaller urban districts (i.e., districts that only recently had to start offering SES); and limited numbers of providers that target ELL students and students with disabilities.

In addition, while the number of on-line providers represents one-fifth of our sample (4/20), two of these four providers were unobservable because of their software-based tutoring (as opposed to live, on-line tutoring). Thus, the number of observations we have for on-line providers is not representative of the proportion of students they serve. However, it should also be noted that on-line tutoring, based on existing observations, interviews with administrators and tutors, and the

nature of a software-based interface, tends to be more standard from session to session than in-person tutoring. Therefore, we feel confident in making some statements about on-line services despite the low number of observations (6/56) in our data set. In addition, in all of our discussions of patterns we include the number of observations and specific limitations, when relevant, of these numbers.