

Contract No.: ED-01-CO-0038-0009
MPR Reference No.: 6413-140

MATHEMATICA
Policy Research, Inc.

**Feasibility and Conduct
of an Impact Evaluation
of Title I Supplemental
Education Services**

*Part B: Supporting Statement
for Request for OMB Approval
of Collection of Information
Employing Statistical Methods*

August 20, 2008

Submitted to:

Institute of Education Sciences
U.S. Department of Education
80 F Street, NW
Room 308D
Washington, DC 20208

Project Officer:

Audrey Pendleton

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director:

Brian Gill, Ph.D.

CONTENTS

	Page
B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS.....	1
1. Respondent Universe and Sampling Methods.....	1
2. Statistical Methods for Sample Selection and Degree of Accuracy Needed.....	3
3. Methods to Maximize Response Rates and Deal with Nonresponse.....	10
4. Tests of Procedures and Methods to be Undertaken.....	11
5. Individuals Consulted on Statistical Aspects of the Design.....	12
REFERENCES.....	14

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

Collection of information is needed to support a rigorous evaluation of supplemental educational services (SES) for the U.S. Department of Education (ED). The No Child Left Behind Act (NCLB) requires school districts to offer SES to students who attend schools that have failed to make adequate yearly progress (AYP) for three consecutive years. SES are tutoring or other academic support services offered outside the regular school day by state-approved providers, free of charge to eligible students. Parents can choose the specific SES provider from among a list of providers approved to serve their area. This evaluation is authorized under the No Child Left Behind Act of 2001, Section 1501 (PL No 107-110).

Mathematica Policy Research (MPR) is working with ED to design and conduct a rigorous evaluation of SES based on a regression discontinuity (RD) design in up to 9 districts. The primary research questions to be answered by the evaluation are: (1) what is the effect of SES on student achievement? and (2) how does the effect of SES vary by student characteristics and provider type? MPR will assess the impact of SES by comparing a treatment and control group of students, where the treatment and control group are formed purposefully based on a measure of prior achievement (such as a test score or grade point average). Valid estimates of the effect of SES can be determined by comparing the average reading and math scores of students who were accepted into SES to the average scores of students who were not accepted into SES, after regression adjusting for the measure of prior achievement used to determine acceptance (this is the definition of an RD design). MPR will assess how impacts vary by provider type by calculating provider-specific impacts and then relating those impacts to provider type, as measured using a survey of SES providers.

We are requesting OMB approval for the regression discontinuity design and for baseline data collection activities. The total evaluation consists of three phases of work, including recruitment, baseline data collection, and outcome data collection. The recruitment phase includes assessing feasibility and recruitment of up to 12 districts. The baseline data collection phase includes collecting parents' choice of providers from districts. We have learned that districts typically ask for parents' provider preferences on the SES parent application form. OMB approval for the outcome data collection phase, which will take place in spring 2009, will be requested in the fall of 2008. This includes collection of information from SES providers and collection of school records.

1. Respondent Universe and Sampling Methods

School districts, students, and SES providers are the primary units of data collection and analysis for the full evaluation. ED's Office of Innovation and Improvement (OII) has compiled a list of all school districts nationwide that are oversubscribed for SES during the 2007-2008 school year and likely to be oversubscribed during the 2008-2009 school year. These districts are listed in Table B.1. As part of the first year design and feasibility study, we will assess the feasibility of conducting the evaluation through informal conversations with district officials in 9 of these districts. The feasibility will be based on the likelihood of oversubscription and whether or not districts would be willing to accept student applications for SES based on a continuous measure of prior achievement.

TABLE B.1

SCHOOL DISTRICTS FROM WHICH FEASIBILITY WILL BE ASSESSED
AND SAMPLE WILL BE RECRUITED

District Name	State
Albuquerque Public Schools	New Mexico
Baltimore City Public Schools	Maryland
Bay County School District	Florida
Brevard County School District	Florida
Bridgeport	Connecticut
City of Chicago SD 299	Illinois
Collier County School District	Florida
Dade County School District	Florida
Denver County 1	Colorado
Flagler County	Florida
Gadsden County	Florida
Hillsborough County School District	Florida
Indianapolis Public Schools	Indiana
Lee County School District	Florida
Leon County School District	Florida
Little Rock School District	Arkansas
Los Angeles Unified	California
Oakland Unified	California
Osceola County School District	Florida
Palm Beach County School District	Florida
Pinellas County School District	Florida
Polk County School District	Florida
Sacramento City Unified	California
San Francisco Unified	California

If more districts are eligible and willing to participate in the study than needed, we will randomly sample districts from geographic strata until sample size targets are met (about 52,600 applicant students). From each district we will include *all* eligible SES applicants in the impact analysis and will include the 25 largest SES providers in the provider survey (we expect fewer than 25 providers in most districts) (see Table B.2). Based on prior studies, we anticipate a response rate of at least 80 percent to the SES provider survey, and we anticipate that test score data from administrative records will be available for at least 95 percent of students, for a total sample of 50,000 students.

Determining Eligibility. School districts are eligible for the evaluation if (1) they have more applicants for SES than can be served with available funds and (2) they accept applicants based on a quantifiable, continuous measure of prior achievement (such as test score or grade point average). The districts identified by OII are believed to have a high probability of meeting the first eligibility requirement. We will assess eligibility and interest in participating in the study through brief telephone conversations with nine districts. In the late summer/early fall of 2008, after OMB clearance, we will follow up with up to 12 districts that we expect to be eligible

and willing to participate in the study in order to begin coordinating their participation in the study.

TABLE B.2
SAMPLE SUMMARY

Unit of Data Collection/Analysis	Number Provided by OII	Number Expected for Study	Anticipated Response Rate
School district	24	12	Not Applicable
Student (SES applicants)	Unknown ^a	50,000	> 80%
SES provider	Unknown ^a	300 (at most)	> 80%

^a We will not know the number of SES applicants and SES providers in each district until we begin recruiting districts for the study in late summer/early fall 2008, after OMB clearance. However, a GAO study (number 06-758) found that over half of the more than 400,000 SES participants in 2004–2005 were concentrated in the largest 21 districts. Because our list of 24 districts includes many of those large districts, and because SES enrollment has likely risen since 2004–2005 as more schools fall short of making AYP, we believe that including 50,000 students in the evaluation is a realistic goal.

2. Statistical Methods for Sample Selection and Degree of Accuracy Needed

Our goal is to include 12 districts and 50,000 students in the impact evaluation. This sample should allow us to detect impacts of 0.20 standard deviations with high probability for specific subgroups of SES providers. Power calculations and other details are shown below.

a. Statistical Methodology for Stratification and Sample Selection

We will not draw a random sample of service providers or students. We will only randomly sample school districts if the number of districts eligible and willing to participate in the study is more than we need to achieve our sample size target. If we need to draw a random sample, we will divide districts into geographic strata and sample within those strata. The strata definition will depend on which districts are eligible and willing to participate in the study. It is possible that we will take all districts in some strata and randomly sample districts in other strata.

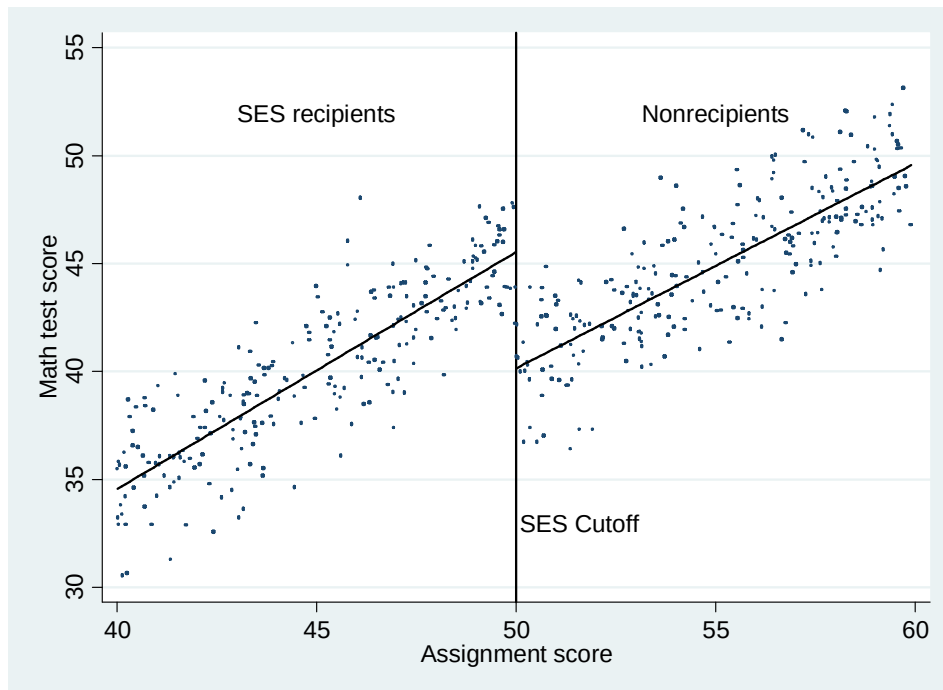
b. Estimation Procedures

A randomized experimental evaluation of SES is precluded by NCLB, which requires that all eligible students who request services receive them, as long as resources are available. Although a randomized design is precluded by statute, NCLB’s rules about the allocation of services when resources are constrained create the opportunity for an RD analysis that will allow causal inferences with rigor approaching that of a randomized experiment.

Impacts will be estimated in a manner consistent with the study’s RD design. Using an RD design, valid estimates of the effect of SES can be determined by comparing the average reading and math scores of students who were accepted into SES to the average scores of students who were not accepted into SES, after regression adjusting for the measure of prior achievement used to determine acceptance. Figure B.1 illustrates the RD design graphically, using a hypothetical example in a hypothetical district. Measures such as prior test scores or grade point average could be used for assignment to treatment or control groups. In this example, students with an

assignment score of 50 or less receive SES (the treatment group), and students with a score over 50 do not (the control group). This figure plots student math test scores against assignment scores. It also displays the fitted regression line for the treatment and comparison groups. The estimated impact on math test scores is the vertical distance between the two regression lines at the cutoff value of 50. An important consideration in calculating impacts using an RD design is the functional form used to regression adjust for prior achievement. In Figure B.1, the functional form is linear. In practice we will also calculate impacts using non-parametric regression techniques that allow for a more flexible functional form.

FIGURE B.1
HYPOTHETICAL EXAMPLE OF THE RD METHOD



the Because

assignment score will be defined differently across districts (we anticipate that in most cases it will be based on a prior year's test score) and because each district will use a different cutoff for allocating services, we will estimate separate impacts for each district in the sample and then compute a weighted average of these estimates to obtain an overall estimate of the impact of SES among the districts in our sample.¹ We will weight district-specific estimates according to the number of eligible students in each district, which will provide an estimate of the impact of SES on the *average student* under study.

We will be seeking OMB clearance (in an addendum to be submitted in fall 2008) to collect data from SES providers in spring 2009 and will use the RD design to explore the relationship

¹ In some districts, assignment score and/or cutoff might also differ by grade, in which case we will estimate district/grade-specific impacts.

between SES provider characteristics and effectiveness. As part of the district SES application materials, parents are typically asked to identify up to three preferred SES providers. Because the preferred providers will be identified prior to determining the RD cutoff, we will be able to calculate provider-specific impacts.² We can then correlate impacts with provider characteristics and practices. Dimensions along which interventions might vary include substantive focus (for example, math or reading), intensity (for example, frequency of student attendance), and method of delivery (for example, small group activities, one-on-one tutoring, or the use of computer technology).

One additional consideration is that some students offered SES might not receive the services, and some students whose assignment score exceeds the cutoff might nonetheless manage to receive SES.³ If this is the case, the impact estimates will represent the impact of *offering* students SES rather than the effect of *receiving* SES. We will be collecting data on whether students received SES from the SES provider survey and from district administrative records. If many students who were offered SES chose not to receive them, or if students who should not have received SES according to their assignment score do in fact receive them, we can compute an additional estimate reflecting the impact on students of receiving SES, using what is known as a “fuzzy” RD design (Trochim 1984; Hahn et al. 2001). This approach is similar to calculating the impact of treatment on the treated in a randomized control trial using a Bloom (1984) adjustment, essentially using the discontinuity in SES receipt at the assignment score cutoff as an instrumental variable for SES receipt, holding constant a function of the assignment score.

c. Degree of Accuracy Needed

An important distinction between an evaluation of SES and many other evaluations of education interventions is that SES are not a single intervention. Instead, they provide the parents of low-income, low-achieving students with the opportunity to enroll their children in an intervention of their choosing among a variety of programs within a range constrained by NCLB and the state approval process. Consequently, the effects of specific types of interventions funded by SES may be of as much interest to parents and policymakers as the overall effect of SES.

Our goal is for the study to be able to detect, with high probability, an impact of 0.20 standard deviations for subgroups of students corresponding to specific types of services. Dimensions along which interventions might vary include substantive focus (for example, math or reading), intensity (for example, frequency of student attendance), and method of delivery (for example, small group activities, one-on-one tutoring, or the use of computer technology). Many

² Provider specific impacts will be calculated by comparing the outcomes of students in the treatment group who identified a given provider as their preferred provider to the outcomes of students in the control group who identified the same provider as their preferred provider. By identifying the preferred provider during the application process, we will be able to know which students to include in the control group for this analysis (otherwise, we would not know who to compare the treatment group to).

³ This second concern is known as comparison group “crossover,” which might occur if the district erroneously provides the student SES or does not have a systematic approach for allocating available services from a waiting list when students initially offered SES decline them.

evaluations of education interventions are designed to detect effects on academic achievement of at least 0.20 standard deviations with high probability. For example, the contractor (MPR)'s evaluation of math curricula and the evaluation of reading comprehension interventions are designed to detect effects in the range of 0.20 and 0.25 standard deviations. By designing the study to detect moderate effects of specific intervention types, we will also be able to detect very small effects of SES overall. The RAND study of SES (Zimmer et al. 2007) found overall average effects in some school districts of less than 0.10. Our study will most likely be able to detect effects smaller than 0.10 for the full sample, as we show below.

The MDE for an RD Design

The minimum detectable effect (MDE) for an RD design is different from the MDE for a random assignment design because of the correlation between treatment status and the score used to define the cutoff, which in turn determines treatment status. We can obtain the impact estimate on an outcome, y , by using the following equation:

$$(1) \quad y = \beta_0 + \beta_1 T + \beta_2 \text{Score} + u,$$

where T is the treatment indicator variable, Score is the score used to assign the units to the treatment or comparison groups, and u is a random error term. In this equation, the impact is identified by assuming that the relationship between y and Score (that is, β_2) is the same for the treatment and comparison groups and that the functional form specifying this relationship is linear. The intercepts of the fitted lines, however, are allowed to differ by research status. Thus, the impact estimate is β_1 and represents the difference between the intercepts of the fitted lines for the treatment and comparison groups. Stated differently, the impact is the difference between the two fitted lines at the point of “discontinuity” (that is, at the threshold score value in the y - Score plane).

To calculate an MDE, we need to know the variance of β_1 . Because of the high correlation between β_1 and β_2 , the variance of β_1 is greater than in the case of a randomized controlled trial (RCT). The ratio of the variance of β_1 in an RD design to the variance of β_1 in an RCT is the RD “design effect,” represented mathematically as

$$(2) \quad \text{Design Effect} = \frac{(1 - R_1^2)}{(1 - R_0^2)} * \frac{1}{(1 - R_{T|\text{Score}}^2)},$$

where R_1^2 is the regression R^2 value from equation (2), R_0^2 is the regression R^2 value under an experimental design, and $R_{T|\text{Score}}^2$ is the R^2 value when T is regressed on Score (and an intercept). The first ratio in the design effect is essentially 1, because the same explanatory variables would be used in either an RD or a random assignment design. The second ratio is what drives the design effect. This ratio could also be expressed in terms of the correlation between T and Score —the greater that correlation, the greater the design effect.

The correlation between T and $Score$ depends on two things: (1) the relative proportion of individuals in the treatment and control groups and (2) the distribution of the $Score$ variable. In Table B.3, we examine how the design effect varies with respect to these two factors. The design effects are calculated using computer simulations.

TABLE B.3
REGRESSION DISCONTINUITY DESIGN EFFECTS

Proportion of Students in Treatment/Control Groups	Probability Distribution of Score	
	Normal	Uniform
90:10	1.5	1.4
80:20	1.9	1.9
70:30	2.4	2.8
50:50	2.8	4.0

The MDE of an RD design is found by multiplying the MDE of an RCT (with the same sample size) by the square root of the RD design effect. For example, if the MDE for an RCT is 0.20 and the RD design effect is 1.5, then the MDE for an RD study with the same sample size as the RCT will be 0.24.

Relationship Between the MDE and Key Design Parameters

The MDE depends on the following key design parameters:

- ***The Number of Students in the Study.*** As with RCTs, including more students in the study increases the precision of impact estimates and reduces the study's MDE.
- ***The Distribution of the Cutoff Score.*** As described in the previous section, the distribution of the cutoff score influences the correlation between the score and the treatment variable. The greater that correlation, the greater the RD design effect. Because the cutoff score is likely to be a measure of prior achievement, we assume in all calculations presented below that the cutoff score follows the normal distribution.
- ***The Proportion of Students That Falls Below the Cutoff to Receive Services.*** As with RCTs, an RD design will have greater statistical power if the number of students assigned to treatment is the same as the number assigned to the control condition. Because demand for SES has only recently begun to outpace funding, we anticipate that the proportion of students that falls below the cutoff to receive services will be high, which means that there will be many more in the treatment than in the control condition. We assume that the size of the control group will not exceed 20 percent of the total sample.

- ***The Difference Between the Treatment and Control Groups in the Proportion of Students That Actually Participates in SES.*** Some students who are offered SES (those below the RD cutoff) might not actually enroll, which means that the proportion of students that participates in SES in the treatment group may be less than 1. At the same time, we anticipate that districts will allow some students above the cutoff to participate in SES in order to “fill the slots” left open by those who were offered SES but declined the offer. This is conceptually similar to the problem of noncompliance in RCTs, and the same correction can be used in an RD design as is used in an RCT. We assume that the difference in participation rates could be as low as 50 percent (for example, an 80 percent participation rate below the cutoff and a 30 percent participation rate above it).

The clustering of students within districts is not listed as a design parameter because we will treat district effects as fixed, not random. That is, districts are not the unit of assignment and they are not a unit of random sampling. Therefore, they do not contribute variance to the impact estimate. Clustering of students within schools or classrooms is also not an issue, because we are not sampling schools or classrooms (we are including all students who apply to SES in the study districts). In Table B.4, we show MDEs for a range of sample sizes, proportions of students in the treatment and control groups, and differences in the participation rates between the treatment and control groups, holding the regression R^2 fixed at 0.40.⁴ This table shows that for a subgroup of 5,000 students,⁵ the study will have an MDE of less than 0.20 as long as the proportion of students in the control group is 10 percent or higher and the difference in participation rates between the treatment and control groups is at least 65 percent. If the proportion of students in the control group falls to 5 percent, then a difference in participation rates of 80 percent would be needed in order to attain an MDE of 0.20 for a subgroup of 5,000 students. With a subgroup of 2,500 students, the study would need a difference in participation rates of nearly 80 percent and 20 percent of students in the control group to attain an MDE of 0.20 standard deviations.

TABLE B.4

VARIATION IN MINIMUM DETECTABLE EFFECT SIZES WITH RESPECT TO TAKE-UP RATES AND THE PROPORTION OF STUDENTS IN THE TREATMENT/CONTROL GROUPS

SES Applicants	Control Group Size	Difference in the Participation Rate Between the Treatment and Control Group		
		80 Percent	65 Percent	50 Percent
5 Percent of Students in the Control Group				
50,000	2,500	0.06	0.08	0.10

⁴ We typically assume a regression R^2 of 0.50 in cases where a baseline test score is available as a covariate. In this study, a baseline test score will be available in most cases and used as the basis for the RD design. However, because we anticipate that some students in our sample will lack a baseline test score, we assume an overall regression R^2 of 0.40 instead of 0.50. These students can still be included in the study if another measure of prior achievement, such as grade point average, is available. But those other measures may not be as highly correlated with the follow-up test score as a baseline test score would have been, hence the lower R^2 assumption.

⁵ With a total sample of 50,000 students, we would be able to analyze 10 subgroups of this size that would correspond to different types of services.

SES Applicants	Control Group Size	Difference in the Participation Rate Between the Treatment and Control Group		
		80 Percent	65 Percent	50 Percent
25,000	1,250	0.09	0.11	0.14
10,000	500	0.14	0.17	0.23
5,000	250	0.20	0.25	0.32
2,500	125	0.28	0.35	0.45
10 Percent of Students in the Control Group				
50,000	5,000	0.05	0.06	0.08
25,000	2,500	0.07	0.09	0.11
10,000	1,000	0.11	0.14	0.18
5,000	500	0.16	0.19	0.25
2,500	250	0.22	0.27	0.36
20 Percent of Students in the Control Group				
50,000	10,000	0.04	0.05	0.07
25,000	5,000	0.06	0.07	0.10
10,000	2,000	0.09	0.12	0.15
5,000	1,000	0.13	0.17	0.21
2,500	500	0.19	0.23	0.30

Note: The MDEs are expressed in effect size units and were calculated assuming (1) a 2-tailed test; (2) a 5 percent significance level α ; (3) an 80 percent level of power β ; (4) a reduction in variance of 40 percent owing to the use of regression models to estimate impacts, R^2 ; and (5) an RD score variable that follows the normal distribution. The figures were calculated using the following formula:

where fct is the sum of two critical values (corresponding to α and β) from the T-distribution with df degrees of freedom, RD is the regression discontinuity design effect, ΔPR is the difference in participation rates between students below and above the RD cutoff, PSC is the proportion of students in the control group, PDF is the probability density function of the score used to determine participation in the RD design (assumed normal), N_T is the number of students in the treatment group, and N_C is the number of students in the control group.

$$\frac{fct(\alpha, \beta, df)}{\Delta PR} * \sqrt{RD(PSC, PDF) * (1 - R_1^2) \left(\frac{1}{N_T} + \frac{1}{N_C} \right)}$$

Based on these calculations, our goal is to recruit enough districts into the study to provide a student sample of about 50,000. If we are able to include only districts that appear likely to have an oversubscription rate of at least 10 percent, then it is likely that we will be able to detect an MDE of 0.20 for a subgroup of 5,000 students. If we are unable to meet that sample target, we can still calculate meaningful effects of SES overall and for larger subgroups. But our preference is to be able to detect moderate effects for smaller subgroups in order to provide a more refined understanding of the impacts of specific provider types.

d. Unusual Problems Requiring Specialized Sampling Procedures

We do not anticipate any unusual problems that require specialized sampling procedures.

e. Use of Periodic Data Collection Cycles to Reduce Burden

The data collection plan calls for the minimum amount of data needed to measure differences in student achievement based on SES provider. The collection of SES application data and parental consent for the student achievement test will be one-time collections.

3. Methods to Maximize Response Rates and Deal with Nonresponse

Parents of students in the study will already be asked to fill out the SES application form by the school district. The district application forms typically ask parents to provide the names of their first three choices of SES providers. Compared with undertaking an independent effort to acquire information on parents’ preferred providers, doing so as part of the application process has major advantages in terms of cost and response rate. We anticipate that 50,000 parents will complete the district SES application and provide information about preferred providers (95% response rate from approximately 52,600 total parents who fill out SES applications).

4. Tests of Procedures and Methods to be Undertaken

To help ED address the study research questions, the contractor will collect and analyze data from several sources. Clearance is currently being requested for collection of SES application data.

ED will request OMB clearance to collect outcome data in an addendum to the current OMB package, including: (1) an SES provider survey (which will allow the contractor to assess provider characteristics that can then be linked to impacts) and (2) the collection of student records/district test scores (the main outcome for the evaluation). Table B.5 shows the schedule of these data collection activities.

Collection of SES Application Data

The contractor will gather information from school districts on parents’ preferred providers listed as part of the SES application process in fall of 2008. The district application form typically asks parents to provide the names of their first, second, and third choice of SES providers. We will ask districts to record this data from SES applications and submit information on parents’ preferred providers in an electronic file to the contractor.

TABLE B.5

DATA COLLECTION SCHEDULE

Activity	Respondent	Clearance Requested in	Clearance To Be Requested in Addendum
----------	------------	------------------------	---------------------------------------

		Current Package
Baseline Data Collection, Fall 2008:		
Collect SES application data (50,000 records from 9 districts)	Parent/guardian via school districts	X
Outcome Data Collection, Spring 2009:		
SES provider survey (225 providers at most)	SES provider	X
Outcome Data Collection, Summer/Fall 2009:		
Obtain student records/district test scores (50,000)	District/School staff	X

5. Individuals Consulted on Statistical Aspects of the Design

This study is being conducted by contractor, Mathematica Policy Research, Inc. (MPR), under contract to the U.S. Department of Education. The project director is Dr. Brian Gill, the principal investigator is Dr. John Deke, and the survey director is Ms. Laura Kalb—all MPR employees. The project team consulted with Dr. Peter Schochet, senior researcher at MPR, about the statistical aspects of the study design. Contact information is provided below.

Brian Gill, Mathematica Policy Research, Inc.	617-301-8962
John Deke, Mathematica Policy Research, Inc.	609-275-2230
Laura Kalb, Mathematica Policy Research, Inc.	617-301-8989
Peter Schochet, Mathematica Policy Research, Inc.	609-936-2783

REFERENCES

- Bloom, H.S. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review*, 8, pp. 225-246, 1984.
- Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, vol. 69, no. 1, January 2001.
- Shadish, W.R., Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin, 2002.
- Trochim, W. *Research Design for Program Evaluation: the Regression-Discontinuity Approach*. Beverly Hills: Sage Publications, 1984.
- Zimmer, Ron, Brian Gill, Paula Razquin, Kevin Booker, and J.R. Lockwood III. *State and Local Implementation of the No Child Left Behind Act. Volume I: Title I School Choice, Supplemental Educational Services, and Student Achievement*. Washington, DC: U.S. Department of Education, 2007.