

**A STUDY OF THE EFFECTIVENESS
OF A SCHOOL IMPROVEMENT INTERVENTION/
DATA COLLECTION
(Study 2.1d)**

OMB Clearance Package Supporting Statement

***Part B: Collection of Information
Employing Statistical Methods***

Regional Educational Laboratory
for the
Central Region

Contract #ED-06-CO-0023

Elisabeth A. Palmer, Ph.D.
ASPEN Associates, Inc.

Submitted to:

Institute of Education Sciences
U.S. Department of Education
555 New Jersey Ave., N.W.
Washington, DC 20208

Submitted by:

REL Central at
Mid-continent Research for Education and Learning
4601 DTC Blvd., #500
Denver, CO 80237
Phone: 303-337-0990
Fax: 303-337-3005

Project Officer

Project Director:

Sandra Garcia, Ph.D.

Louis F. Cicchinelli, Ph.D.

Deliverable #2007-2.3/2.4

July 22, 2007

MREL
© 2007

This report was prepared for the Institute of Education Sciences under Contract #ED-06-CO-0023 by Regional Educational Laboratory Central Region, administered by Mid-continent Research for Education and Learning. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

TABLE OF CONTENTS

B. Collection of Information Employing Statistical Methods.....	1
1. Respondent Universe and Sampling Methods.....	1
Universe.....	1
Recruitment.....	1
Selection Criteria.....	3
Random Assignment.....	3
Sample.....	3
2. Procedures for Data Collection.....	4
Stratification and Sample Selection.....	4
Estimation Procedures.....	5
Degree of Accuracy Needed.....	5
Data Quality	5
Power Analysis	6
Unusual Problems Requiring Specialized Sampling Plans.....	9
Use of Periodic Data Collection Cycles to Reduce Burden.....	9
3. Methods to Maximize Response Rates and Deal with Non-response.....	9
General Factors Influencing Feasibility of the Study.....	9
Response Rates and Non-Response.....	10
Ongoing Communication	11
Minimal Data Collection.....	11
Minimal Burden.....	11
4. Test of Procedures or Methods.....	11
5. Individuals Consulted on Statistical Aspects of the Design.....	12
References.....	12

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. RESPONDENT UNIVERSE AND SAMPLING METHODS

Success in Sight is a comprehensive approach to school improvement. As a result, the proposed study employs a hierarchical design, with schools as the unit of assignment. Student-level data are therefore nested within school-level clusters. To address the research questions, a randomized-control (experimental) design will be employed. After confirming eligibility, elementary schools will be randomly assigned to the treatment or control condition using the “random sample” procedure in the SPSS software (SPSS, 2003). Treatment schools will participate in professional development and onsite support provided by *Success in Sight* mentors to build school capacity to plan for, manage, and engage in whole-school improvement activities. Control schools will engage in their regular school reform practices.

Universe. The study will target elementary schools with low to moderate performance for participation in the study, as these schools are most likely to need and benefit from a school-wide intervention for improving student performance. Elementary schools with this level of performance will likely contain at least average percentages of low SES, minority, and ELL students. Districts that are large enough to provide approximately 10 or more eligible elementary schools will be recruited to participate, preferably within a single state; randomization will occur within a district. Power analyses indicate that a total of 52 schools (half assigned to the intervention condition and half assigned to the control condition) are needed to detect significant but potentially small differences between the treatment and control groups.

This study is based on a purposive sample, which consists of schools volunteering to participate in the study. Although the sample is not statistically representative of any larger group of schools (e.g., all low- to moderate-performing schools in the state), it is intended to represent schools from the full range of low- to moderate-performing schools found in larger school districts with economically and ethnically diverse student populations. It was not feasible to obtain a probability sample of schools for this experimental setting study for a variety of reasons, including the need to conduct the study within a single state in order to increase internal validity (i.e., using the same state assessment for all schools), avoid overlapping the sample for this study with the sample for other IES studies, and work within a reasonable budget.

Recruitment. Recruiting efforts will target the district level. District recruiting affords a number of advantages, including support of district administration for the study, a reduction in the number of district-level approvals required, and the facilitation of access to student-level achievement data. Site recruitment will occur primarily through outreach at professional education conferences and through other professional networks, including contacts at state departments of education and school districts. Once potential sites have been identified, the study team will work closely with the districts to garner the participation of eligible elementary schools

within their district. In identifying eligible schools, the research team will work closely with the district to ensure that potential schools are not slated to be closed or restructured during the study period.

Recruitment efforts to date have been focused on identifying districts actively seeking comprehensive school improvement interventions. Two potential districts in Minnesota — Minneapolis and Saint Paul — have already been identified as both interested and eligible to participate; preliminary discussions are currently underway with these districts. Additional schools within Minnesota will also be targeted, notably within the seven-county metropolitan area, which includes these two school districts and two other larger school districts in the state. The sampling frame presented here represents all eligible elementary schools in major metropolitan area school districts by level of performance (see Table 6). Additional schools from outside of this area will be recruited as needed.

Table 6. Sampling Frame

	Low Performing¹ Schools	Moderate Performing² Schools	Additional Schools³	Total
Target Districts				
Minneapolis	15	0	0	15
Saint Paul	16	14	5	35
Other Districts Within the Metropolitan Area	7	6	15	28
Totals	38	20	20	78

¹ Low-performing elementary schools have already been designated as Not Making Annual Yearly Progress at some point during the three years prior to the study in either reading or mathematics.

² Moderate-performing elementary schools have already been designated as Making Annual Yearly Progress in each of the three years prior to the study, but just barely achieved proficiency on the state assessment in either reading or mathematics.

³ Additional elementary schools include schools that are likely to be designated as low- to moderate-performing during the recruitment phase of the study.

Minnesota was chosen for this study for the following four reasons:

- The intervention, *Success in Sight*, was developed and tested within the Central Region states, making it difficult to find a set of schools within the region that had not already been exposed to the intervention.
- The two potential districts are large, urban school districts with student populations that are ethnically and economically diverse. This setting provides the opportunity to study the impact of *Success in Sight* on student populations that are of wide interest to other schools across the nation.

- Given that the research firm contracted to conduct this study is located in Minnesota, it will be most cost effective to implement the study in the same region.
- Using achievement data from a single state will increase the internal validity of the study by allowing for a more direct comparison of mean student achievement levels between the intervention and control groups.

Selection Criteria. In order to strengthen the validity of the study, selection criteria have been developed to identify schools eligible for participation. Districts or individual schools that wish to participate in the study will need to meet several criteria. Prior to random assignment, a pool of eligible schools will be screened for the following target characteristics: (1) low or moderate student performance; (2) at least three classrooms each in grades 3, 4, and 5; (3) average student demographics (e.g., percent SES, ELL, minority); (4) not already engaging in a comprehensive school reform model that includes an emphasis on shared leadership and collective efficacy (two unique features of *Success In Sight*) and have no plans to do so; and (5) available and ready to complete all the study requirements, including random assignment. Schools excluded from the study include charter schools and alternative schools. Solicitation of schools will be based on school performance and school size.

Random Assignment. The final sample of 52 eligible schools will be randomly assigned with districts to treatment or control group (26 schools in each) using a simple random sample. Schools randomly assigned to the intervention will participate in *Success in Sight* for two years. The control schools will continue with their usual practice. Given that many if not all of the eligible low- to moderate-performing schools will have been identified as not making adequate yearly progress under NCLB, it is expected that all of the eligible schools will be engaged in or have plans to engage in some form of school improvement in line with their adequate yearly progress status. Schools assigned to the control group may continue with these efforts, as *Success in Sight* is not designed to replace existing reform efforts but rather engages schools in a change process that incorporates existing and new school improvement practices. At the end of the study, the control schools may elect to participate in the intervention at their own discretion and expense.

Sample. The final sample for this study will include 52 low- to moderate- performing elementary schools. The 52 schools will be drawn from population of 78 elementary schools that are low- to moderate- performing, have enrollments in the target grade levels that meet the eligibility requirements, and have expressed an interest in participating in the study. As such, this is not a probability sample.

Each school is assumed to have at least 20 teachers across the grade levels served, taking into consideration teacher mobility (i.e., that should a teacher leave the school during the study and not be replaced, that the school would still have a minimum of 20 teachers on staff).¹ All teachers

¹ The intervention, *Success in Sight*, is designed to address teacher mobility as part of its focus on school reform.

will participate in the teacher survey.

Two student samples will be used to examine the main effects of *Success in Sight* on student achievement. The student samples are based on the initial cohort of students in grades 3, 4, and 5. This means that baseline data will be collected from students in these grade levels. The first follow-up data collection then will occur when these students are in grades 4, 5, and 6, with the final data collection occurring in grade 5, 6, and 7. Each eligible school must have at least three classrooms at each of these grade levels to reach the required sample sizes. Should a school have more than this number of students enrolled in these grade levels, all students will be included in the sample. The following student samples account for 30% student attrition over the course of the study.

- Sample A: Longitudinal School Cohort of 110 “stayers,” students who remained in the same school throughout the study, and
- Sample B: Typical School Cohort 150 “stayers plus in-movers,” students who remained plus students who moved into the school after the baseline year.

The site visits will include data collection with the school principal, school leadership teams, and a cross-section of teachers. All school principals and leadership teams will participate in interviews. Focus groups will be conducted with cross-section of teachers representing all subject areas and grade span levels within the school.

2. PROCEDURES FOR DATA COLLECTION

The plans for data collection are presented in *Part A-2 — Purposes and Use of Data*.

Stratification and Sample Selection. The student sample will be representative of all students enrolled in the participating schools during the initial data collection in spring 2008 who remain throughout the study (“stayers”) and students who enrolled in subsequent years of the study (“stayers plus in-movers”). Teachers will represent all teachers employed in the participating schools during the study. We do not plan to use any stratification within a school site.

The sampling frame shown in Table 6 represents the number of elementary schools that serve the required grade levels and are of sufficient size to meet eligibility criteria. As such, the number of schools, teachers, and students in the final sample represents approximately two-thirds of the target population (see Table 7).

Table 7. Comparison of Final Sample to Sampling Frame

Specifically, the intervention includes procedures for socializing teachers new to the school into the reform process.

Participants	Final Sample	Target Population
Elementary schools 26 treatment / 26 control schools	52	78
Teachers, school administrators, school staff 20 per school ¹	1,040	1,560
Students starting with baseline cohort in grades 3, 4, and 5 110 “stayers” per school ¹ 150 “stayers + in-movers” per school ¹	5,720 “stayers” 7,800 “stayers + in-movers”	8,580 “stayers” 11,7000 “stayers + in-movers”

¹ Represents minimum number of staff and students. All teachers in the school will be surveyed and all students in the cohort grades will be tested.

Estimation Procedures. The plans for the statistical analysis of the data are presented in *Part A-16 — Tabulation, Analysis and Publication Plans and Schedule*.

Degree of Accuracy Needed.

Data Quality. Several protocols will be followed to ensure data quality. In addition to a formal quality assurance review for all deliverables and key project documents, the study includes quality assurance procedures for data collection, data management, and data analysis.

Quality assurance with regard to data collection is addressed in a number of ways. First, all data collection includes well-tested questions that are not only reliable and valid measures of key constructs, but clearly written to be understandable to respondents. Second, all data collection has been kept to a reasonable length to ensure that respondents do not become fatigued and lose interest. Third, the specific protocols for data collection have been shown to result in quality data (e.g., online self-administered teacher survey, online student assessment administered by trained proctors), particularly the use of online methods which remove the additional step of data entry. Fourth, the researchers will work with the site coordinators to schedule all data collection activities and will provide training to teams of researchers conducting the site visits. Fifth, with the exception of students, all persons participating in data collection will be partially compensated for their time spent outside of their normal duties; likewise, all agencies engaged in providing access to the data will be compensated for their time.

Quality assurance with regard to data management includes specific protocols for cleaning and checking data, maintaining and accessing secure data files, naming files to designate data source and iteration, and preparing well-documented public data files. Quality assurance with regard to data analysis includes training in the coding and analysis of qualitative data from site visits and clear procedures for conducting and documenting all analyses, from the pre-intervention analysis to the final analyses of main effects.

Power Analysis. Three separate power analyses, two for the main effect of achievement and one for the proximal outcomes of school improvement practices, were conducted in order to determine the sample size necessary to detect the school-level effects of assignment to the intervention on these outcomes. All power analyses were conducted using Optimal Design software (Raudenbush, Spybrook, Liu, Congdon, & Martinez, 2006) specifically made for power analyses for hierarchical cluster randomized designs. Schools are the unit of randomization and thus the unit of analysis. School-level student achievement data is the main dependent variable (i.e., students nested within schools). Proximal outcomes to be examined are school-level improvement practices (i.e., teachers nested within schools).

Sample and cluster size were chosen to achieve a high level of power, $>.80$. Conservative values for parameter estimates were chosen in order to maintain a high level of power to detect small but practically significant effects. Rationales for the estimates for effect size, intraclass correlation, and the covariate are described below. Power analyses were conducted for fixed effects. According to Schochet (2005), the loss of degrees of freedom associated with blocking by district is negligible and does not require adjustments to the power analyses. Power analyses were adjusted to reflect the inclusion of one covariate. A summary of the parameters used in the power analyses for the main effects and proximal outcomes examined in this study are presented in Table 8.

Table 8: Parameter Estimates for Power Analyses

	Effect Size	ICC	R ²	Min. Power	n	Cluster Size
Main Effects: Student Achievement						
<i>Sample A: Longitudinal School Cohort of “Stayers”</i>	.20	.10	.50	.80	110	48
<i>Sample B: Typical School Cohort of “Stayers Plus In-Movers”</i>	.20	.10	.50	.80	150	47
Proximal Outcomes: School Improvement Practices	.30	.10	.30	.80	20	42
Final Sample ¹						52

¹ Includes over sample to account for school attrition

Main Effects: Student Achievement — As shown in Table 8, the assumed minimum detectable effect for the main effect of student achievement is .20. This value is a conservative estimate based on the literature on the effects of whole-school reform on student achievement. No empirical evidence is available from field trials of the intervention itself. However, estimates of effect sizes are available from other studies of whole-school reform. These estimates vary according to the type of intervention and the outcome measure. In their meta-analysis, Borman,

Hewes, Overman, and Brown (2003) report that the effects of comprehensive school reform (CSR) on student achievement typically range from .09 for third-party studies using comparison groups to .15 for all evaluations of the achievement effects of CSR. When using all available studies, the effects of four CSR models most closely aligned with McREL's *Success in Sight* were .09 for Accelerated Schools, .13 for the Center for Effective Schools, .15 for the School Development Program, and .25 for Onward to Excellence (Borman, et. al., 2003). Based upon the documented size of the effect on student achievement for *Onward to Excellence*, the model most closely aligned with *Success in Sight*, a minimum detectable effect of .20 is reasonable and reflects the effects of *Success in Sight* when well-implemented over a period of two years by the highly trained McREL mentors.

A conservative value of .10 was selected for the intraclass coefficient (ICC) based on the following sources. Liu, et al. (2006) cite typical intraclass correlation coefficients for educational achievement to be between .05 and .15. Schochet (2005) states that ICC for standardized test scores often range between .10 and 0.20. Schochet also found ICCs in grade 3 and 4 reading and math ranging from .06 to .08 (adjusted for district effects) across 71 Title I schools in 18 school districts engaged in whole-school reform.

Prior achievement was selected as a cluster-level covariate, and the proportion of postintervention variance explained by preintervention test scores of .50 was deemed an appropriately conservative estimate based on prior research. Schochet (2005) concludes that the proportion of variance explained by pretest measures is at least .50 when student-level data are used. Bloom, Bos, and Lee (1999) found similar values. Bloom, Richburg-Hayes, and Black (2005) found values ranging from .33 to .81 across five districts for school level pretests.

Power analysis for the outcome of student achievement was conducted using the above parameter values as well as the following very conservative estimated end-of-study sample sizes:

- Sample A: Longitudinal School Cohort of 110 “stayers,” students who remained in the same school throughout the study, and
- Sample B: Typical School Cohort 150 “stayers plus in-movers,” students who remained plus students who moved into the school after the baseline year.

Both sample sizes assume students to be nested within each school and 30% student mobility each year. School-level cluster sizes assumes three classrooms per grade level in grades 3, 4, and 5 with approximately 25 students per classroom (common in the urban school districts being targeted for this study) at baseline.

Given the above assumptions, Optimal Design software (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006) calculated that 48 schools (approximately 24 treatment and 24 control clusters) were necessary to achieve the desired power of $> .80$ for the student achievement outcomes related to Sample A: Longitudinal Cohort of “stayers” and 47 schools for Sample B:

Typical School Cohort comprised of “stayers and in-movers”.

Proximal Outcomes: School Improvement Practices — The four proximal outcomes of school improvement practices include data-based decision-making, effective practices, purposeful community, and shared leadership. Little empirical evidence could be found regarding estimates of effect size, intraclass correlation, and proportion of posttest variance explained by baseline measures of these school improvement practices. Consequently, the parameter estimates for these analyses, shown in Table 8, were chosen for the following reasons.

With regard to the minimal detectible effect size, we can assume that there is some effect, but the magnitude of that effect is unknown. Rigorous studies of CSR to date have not been designed to examine changes in school practices over time. However, in one study of CSR models and distributed (shared) leadership (Camburn, Rowan, and Taylor, 2003), the authors “tentatively” claim that the CSR programs they studied (Accelerated Schools, America’s Choice, and Success for All) configure leadership in their schools differently than non-CSR schools, but do not give direct estimates of those effects for individual programs. Given that *Success in Sight* is a comprehensive school improvement approach that deals directly with changing school practices, a conservative estimated effect size of .30 was deemed appropriate.

No evidence of ICC estimates for school improvement practices was found. Thus, a value of 0.10 was used as the estimate of the ICC based on the assumption that the amount of variance between clusters (schools) would be lower, given the restricted sample of schools in the study. In other words, the schools in this study are similar in student performance (i.e., low- to moderate-performing) which reflects the extent to which they engage in practices that support improved student achievement (i.e., not as much as high-performing schools)

Finally, estimates for postintervention variance explained by preintervention measures of school improvement practices were conservatively set at .30 for several reasons. First, it is assumed that school practices would tend to be relatively unstable due to variations in implementation; specifically, the manner and timelines by which the leadership teams at each school would “scale up” to involve the whole school in the change process. Second, the intervention itself anticipates variations in the choice of school improvement goals to be selected by participating schools as their focus for improvement. Third, the baseline scores on school practices will be used as the covariate for school practices measures at the end of Years 1 and 2, and the correlation between these measures are not directly known, but only hypothesized on the basis of intercorrelations. A study of the Effective Schools CSR model, which included 38 high schools, 32 middle schools, and 134 elementary schools across 22 school districts, reported high intercorrelations among school environment (i.e., school practices) variables (Witte & Walsh, 1990). This study examined four scales related to effective schools, teacher control or influence, and parental involvement. Intercorrelations for teacher control and teacher ratings of school effectiveness ranged from .52 at the elementary level to 0.82 at the middle school level.

An assumption of 20 teachers per cluster was used to estimate final sample size for the power

analysis for reports of school practices with an effect size of .30, the proportion of postintervention variance explained by preintervention test scores of .30, and an intraclass correlation of .10. Using the above parameter estimates, Optimal Design calculated that 42 clusters were necessary to achieve a power $> .80$ using the above assumptions.

To accommodate both the main and secondary analyses and accommodate school attrition, a total of 52 schools will be included in this study.

Unusual Problems Requiring Specialized Sampling Plans. We do not anticipate any unusual problems requiring specialized sampling procedures.

Use of Periodic Data Collection Cycles to Reduce Burden. To reduce the burden on respondents, all data collection is scheduled at naturally occurring data collection points during the school year; that is, when schools traditionally participate in student testing and other data collection required by the state (i.e., either spring or fall of the school year).

3. METHODS TO MAXIMIZE RESPONSE RATES AND DEAL WITH NON-RESPONSE

General Factors Influencing Feasibility of the Study. Several factors may influence the feasibility of this study. First, an adequate sample of eligible schools must be recruited to participate. Specifically targeting schools that need to improve student achievement, as based on adequate yearly progress status, and highlighting the benefits of involvement, will help elicit school participation. A large enough sample of schools will be recruited to provide reliable data even in the case of school attrition. The desired sample of 52 includes four more schools than needed for the statistical analyses.

Second, the schools must be willing to participate in random assignment. To increase the willingness to participate in the study and abide by the random assignment, a comprehensive set of tangible and intangible benefits are planned to encourage and support teacher participation and retention for the duration of the project.

Tangible benefits of participation in the project include:

- Teachers in treatment schools will receive all of the training, mentoring, resources and materials offered through *Success in Sight* at no charge.
- Participants in both treatment and control sites will receive partial compensation for participating in data collection that is over and above their regularly assigned duties at the school and occurs outside of the school day. Per OMB, the customary rate for

such participant compensation is \$25/hour for teachers and \$35/hour for administrators, payable annually.²

Intangible benefits of participation include:

- Satisfaction of participation in a project with high visibility.
- Potential benefit of intervention to participant's students (intervention groups initially, all groups eventually).
- Potential for satisfaction from professional growth.

Response Rates and Non-Response. Obtaining high response rates also will be critical to the success of this study. It will be particularly important to obtain response rates that are not only high overall, but that are approximately equal in the treatment and control groups. It is anticipated that the response rate for the online teacher survey will be at least 80%.³ (This expected response rate is reflected in the sample size for the analyses involving data gathered from teachers.) It is also expected that student participation in the standardized achievement test, administered immediately following the intervention, will reflect traditional levels of participation on the state assessment. This assumption is based on the fact that the target school districts have a history of administering standardized assessments at the district level as a regular part of their assessment portfolio. As such, they require high levels of participation in these assessments. Currently, one of the target districts is administering the standardized assessment being utilized in this study; the other district recently ended the administration of a similar assessment. Finally, participation in the site visit interviews and focus groups will, by design, be representative of the target respondent groups. That is, site visits will not be considered complete until all school principals have been interviewed, leadership teams have been interviewed with the majority of members present, and focus groups conducted with an appropriate cross-section of teachers. To ensure such participation in the site visits, the research team will work with the schools from the beginning to not only set the expectation for involvement, but to plan for it through ongoing communication. In addition, participation in data collection is expected to be high among both treatment and control groups because of the tangible benefits noted above and efforts to reduce the burden of data collection on participants. Should non-response bias occur despite these efforts, the data will be appropriately weighted to adjust for the specific nature of this bias (i.e., particular subgroups).

² A review of other OMB packages for similar whole school studies revealed customary rate of participant compensation for data collection activities to be \$26-\$30/hour for teachers and \$36-\$48/hour for administrators. Studies reviewed included Reading First (Abt Associates, Inc., 2004); Longitudinal Analysis of Comprehensive School Reform Implementation and Outcomes (LACIO) (WestEd & Cosmos Corporation, 2006); and Trends in International Mathematics and Science (TIMMS) (Windwalker Corporation & Westat, Inc., 2005).

³ The research team has previously conducted similar studies with the schools being targeted for this study. In these studies, the response rate for the online teacher survey exceeded 80%.

Ongoing Communication. All study participants (treatment and control schools) will be briefed early on and throughout the project about the expectations for data collection, including the nature of the data collection, schedule, and anticipated time needed to complete various tasks. All such communications have been (or will be, for documents developed during the project) reviewed to ensure they are at the appropriate reading level for the recipient. The research firm conducting this study has worked with Minnesota schools for many years and is aware that these schools have a history of supporting data-based decision-making and participating in educational evaluations. As such, with the proper notification, assurances of confidentiality, and support of the district, most parents have typically agreed to have their child participate in additional testing for the purposes of program evaluation.

Minimal Data Collection. All instrumentation is designed to gather only the most necessary data and in a reasonable amount of time. Existing student achievement data from the state assessment will be used for the baseline and one of the two outcome measures. The teacher survey and site visit protocols have been carefully evaluated and revised where necessary to ensure minimal time burden and clarity of directions.

Minimal Burden. Online data collection will be used whenever possible to reduce respondent burden (i.e., baseline student assessment and teacher survey). Protocols for the online teacher survey reflect best practices in survey research (e.g., three-wave mailings including initial and two follow-ups) while addressing potential disadvantages of Web-based administration (see Part A-3 — *Use of Improved Information Technology to Reduce Burden*). Online administration allows for easy notification of the survey posting, reminders, and direct follow-up with non-responders via email. The online student assessment to be administered at the end of the intervention is one already in use in many Minnesota school districts as it is aligned with state standards and correlates with the state assessment results. Standard administration of this assessment includes opportunities for students who are absent to participate in make-up testing. Consequently, many parents are already familiar with the assessment and its administration.

4. TEST OF PROCEDURES OR METHODS

As much as possible, existing instruments with documented reliability and validity were identified for use in this study; this includes the teacher survey and all student achievement tests. Existing measures were selected with construct validity in mind as well as their degree of sensitivity to the intervention. Administration of the online teacher survey will follow protocols used successfully by the research team in previous studies. All student assessments will follow the standard protocols followed by the state and/or testing company. When existing measures were not available, as in the case of the site visit protocols for interviews and focus groups, original instruments were developed to provide the necessary alignment with and sensitivity to the intervention. The site visit interview/focus group questions were pilot tested with a small sample (less than 10) of individuals from the target population to ensure clarity of directions and questioning and efficient use of respondents' time. The feedback from this pilot test resulted in only minor changes to the wording of the interview/focus group questions to increase clarity.

5. INDIVIDUALS CONSULTED ON STATISTICAL ASPECTS OF THE DESIGN

The statistical aspects of the design have been reviewed thoroughly by staff at the Institute of Education Sciences, as well as by members of the study's expert panel listed in *Part A-8 — Federal Register Comments and Persons Consulted Outside the Agency*.

The following individuals have worked closely in developing the statistical procedures and will be responsible for overseeing data collection and data analysis:

Dr. Elisabeth Palmer, Principal Investigator, ASPEN Associates
(952) 837-6251

Dr. Frances Lawrenz, Associate Dean for Research,
University of Minnesota

Dr. Michael R. Harwell, Professor Psychological Foundations,
University of Minnesota

Dr. Stephanie B. Wilkerson, President, Magnolia Consulting

Jessaca Spybrook, Consultant, Optimal Design, University of Michigan

REFERENCES

- Bloom, H. S., Bos, J. M., & Lee, S.-W. (1999). Using cluster random assignment to measure program impacts: Statistical implications for the evaluation of education programs. *Evaluation Review*, 23(4), 445-489.
- Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2005). *Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions* (working paper). New York: MDRC. (ERIC Document Reproduction Service Number: ED486654).
- Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125-230.
- Camburn, E., Rowan, B., & Taylor, J. (2003). Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. *Educational Evaluation and Policy Analysis*, 25(4), 347-373.
- Liu, X., Spybrook, J., Congdon, R., Martinez, A., & Raudenbush, S. (2006). Optimal Design for

Multi-Level and Longitudinal Research (Version 1.77) [Computer software]. University of Michigan, Ann Arbor: Survey Research Center.

Raudenbush, S., Spybrook, J., Liu, S., Congdon, R. & Martinez, A. (2006). Optimal Design for Longitudinal and Multi-level Research (Version 1.77) [Computer software]. University of Michigan, Ann Arbor: Survey Research Center. Retrieved June 1, 2006, from <http://sitemaker.umich.edu/group-based/files/odmanual-20060517-v156.pdf>

Schochet, P. Z. (2005). *Statistical power for random assignment evaluations of education programs* (No. 6046-310). Princeton, NJ: Mathematica Policy Research Inc. (ERIC Document Reproduction Service No. ED489855)

SPSS, Inc. (2003). SPSS for Windows (Version 12.0.1) [Computer software]. Chicago: Author.

Witte, J., & Walsh, D. (1990). A systematic test of the effective schools model. *Educational Evaluation and Policy Analysis*, (12)2, 188-212.