Contract No.: ED-01-CO-0039/0010
MPR Reference No. 6136-600

MATHEMATICA
Policy Research, Inc.

Supporting Justification for OMB Clearance of Information Collection Forms for the Evaluation of Reading Comprehension Interventions

Revision of Currently Approved Collection (OMB #1850-0812)

Section B

June, 2007

Submitted to:

U.S. Department of Education
Institute of Education Sciences
555 New Jersey Ave., NW, Rm. 308
Washington, DC 20208
(202) 208-7078

Project Officer:
Audrey Pendleton

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave., SW, Suite 550
Washington, DC 20024-2512
(202) 484-9220

Project Director:
Jerry West, Ph.D.
Deputy Project Director:

Wendy Mansfield, Ph.D.

CONTENTS

Page

TABLES AND FIGURES

Appendix Tables

SUPPORTING STATEMENT
REQUEST FOR CLEARANCE OF INFORMATION COLLECTION FORMS
FOR AN EVALUATION OF READING COMPREHENSION INTERVENTIONS

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

The study does not aim to form a nationally representative sample of schools. Rather, the goal is to achieve a purposive sample that includes students and schools, in selected school districts, eligible for Title I funds and to achieve internally valid comparisons by randomly assigning schools to treatment conditions.

To address the Title I policy goal of determining the best way to help low-income children meet state academic-achievement standards, schools with a substantial portion of economically disadvantaged students were selected. These schools are also large enough to support the study design, which allows for subgroup analysis. The districts are also geographically dispersed, so that the results will be relevant for different regions of the country. While a sample of as few as four school districts may have been optimal (Glazerman and Myers 2004), we decided to spread the schools over a larger number of districts (10) so that no single district would have a large burden.

For the first year of the evaluation, we used the Common Core of Data (CCD) to identify districts meeting the above criteria. Specifically, we identified districts with 10 or more elementary schools with at least 40 percent of its students eligible for free or reduced-price lunch or schools operating school-wide Title I programs. A total of 157 districts met these criteria. We also contacted organizations such as the Council of Chief State School Officers (CCSSO) to obtain information on potential districts and asked developers of reading comprehension

programs to provide information about districts that expressed interest in using their intervention. In all, we contacted 71 school districts to assess their interest in participating in the study.

We began recruitment for Year 1 at the state level by sending a letter (Appendix H) to CCSSOs in the relevant states. The letter briefly described the study and noted that we would be calling the CCSSO to provide more information. With the state's support, we began district recruitment efforts with introductory letters (Appendix I) and then telephone calls to district superintendents.

We selected the 13 of the most eligible districts that showed an interest in participating in the study and then conducted site visits to those districts. Based on the profiles and additional information gathered during the site visits and subsequent discussions, we finalized the set of 10 districts and 89 schools that were best suited for the study.

Once the schools were selected, we enrolled all consenting students in all fifth-grade classes. The rationale for not sampling within schools was that the fixed costs per student or classroom are low, and we wanted to encourage support within the fifth-grade teaching team and promote its commitment to optimal implementation of the intervention across all classrooms.

Both components of Year 2 of the evaluation are based on the initial random assignment of schools conducted for the study. For the upcoming Component 1, we will test the original cohort of fifth graders at the end of their sixth-grade year. This component does not require the implementation of the reading comprehension interventions in sixth grade. It will, however, require the recruitment of middle schools into the study in 7 of the 10 districts where sixth grade is part of middle school. Based on preliminary data that participating districts provided regarding the feeder structure of schools, we expect that the number of middle schools will range from 4 to 12 per district. In three districts, this component will involve only the elementary schools that are already participating in the study, as these schools include grade six.

To assess the impact on fifth graders of being taught by *teachers* with a full year of experience with the interventions, in Component 2 we will repeat the original study using the same schools and teachers but a new cohort of fifth graders. The only difference between this approach and the Year 1 study is that some of the original teachers will likely have left the study schools, and some of the original schools may not agree to participate for a second year. If all of the original study teachers have left a school, that school will not be included in the analysis for this component (as there would be no teachers with experience using the interventions in those schools). If no original teachers remain in any of the control schools in a district, the analysis for this component will focus on comparing the impacts of the four intervention groups (as comparisons of the treatment and control groups would not be possible). To assess the impact on fifth graders of being taught in *schools* with a full year of experience with the interventions, we will include *all 5th-grade teachers* (both new and original) and a new cohort of fifth graders.

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

For Year 1 of the evaluation, we estimated that achieving the evaluation objectives would require a sample of 100 schools. This sample size requirement assumed that we would implement the study in 10 districts (10 schools per district) and find an average of three fifth-grade classrooms per school with 26 students per classroom. We estimated that this design would produce an overall sample of 300 classrooms and 7,800 students.[1] Based on these assumptions, the study could detect an effect size (impact) of 0.25 for comparisons of any of the interventions with the control group and 0.26 for comparisons of the impacts of the different interventions.[2]

[1] All sample sizes refer to the number of units that were used in the analysis. If some schools dropped out of the study or failed to comply with their treatment assignment, then a higher initial sample size would have been required.

[2] We have calculated this minimum detectable effect assuming a higher than normal threshold for statistical significance because we are making multiple comparisons. See Box, Hunter, and Hunter (1978) for a discussion of multiple comparison problems in experimental research and James-Burdumy et al. (2006) for an application to the

Now that we have completed Year 1 of the study, we were able to update some of the original assumptions used to determine the minimum detectable effects. Based on the 89 participating schools and on the updated ICC assumptions calculated from the baseline data, the study can detect an effect size of 0.23 for comparisons of any of the interventions with the control group and 0.24 for comparisons of the impacts of two interventions. With respect to Year 2 of the study, we calculate that—when comparing each intervention to the control group—we will be able to detect, with high probability, sustained impacts on student achievement that are at least 0.25 of a standard deviation.

This threshold for policy relevance of one-quarter of a standard deviation, although somewhat arbitrary, represents a reasonable floor for considering an intervention to be "effective." In 2000, the National Reading Panel (NPR) reviewed rigorous studies of comprehension interventions and found effect sizes for impacts on student achievement ranging from 0.24 to 1.70. For standardized test, the median effect size of the six studies of reciprocal teaching reviewed by the NRP was 0.34. For the seven studies of reading comprehension interventions, the median effect size was 0.35. Larger effect sizes were found with assessments that tested comprehension directly.

Once recruited for Year 1, the participating schools were randomly assigned to either an intervention or a control condition. Compared to randomly assigning classrooms or students, randomly assigning schools required a larger sample of schools to disentangle the treatment effects from school-level characteristics. Randomly assigning students would have been the most statistically efficient research design when compared to the random assignment of schools, for example, because the clustering of students within schools must be taken into account in the latter design. However, it was not feasible for this study because the interventions are not pull-

current problem.

out tutoring programs or individual instruction programs that could be administered to some students, but rather classroom-wide instruction programs for all students in the class.

Randomly assigning classrooms would have been the next most efficient design, allowing us to compare classrooms in the same school and thus eliminate school climate and other school-level factors as an influence on student outcomes. But there are contamination or spillover effects associated with this approach, as teachers would have been aware that one or more of their colleagues was delivering instruction differently or receiving some special intervention, which could have influenced their behavior. Additionally, intervention and nonintervention students interact, possibly closing the gap between their differences and thus their outcomes. In either case, the impact estimates would usually be biased toward zero.

Cost efficiency and fairness are two other factors that argued for implementing the same intervention in all classrooms at a school. Implementation costs are determined to a greater extent by the number of schools than the number of classrooms. Additionally, principals may be reluctant to have their school participate in a study where some teachers are provided with special training or materials and others are not. Even if principals allowed the differential treatment within a school, there might be pressure to permit some practices to spill over into nonintervention classrooms, thus biasing impact estimates. There might also be pressure to allow students perceived to "deserve" one treatment over another to transfer (cross over) to the "classroom of interest," also biasing the impact estimates.

Accordingly, we opted for a design that would randomly assign schools to intervention groups and a control group. Although school-level random assignment—compared to student-level or classroom-level assignment—required a larger sample of schools and thus increased costs for securing schools' cooperation, it eliminated many of the threats to the study's feasibility and internal validity. To reduce the number of schools needed for the evaluation and to increase

the precision of the impact estimates, we used a randomized block design, which is analogous to stratification techniques used to make statistical sampling more efficient. One blocking technique we implemented was to first identify schools that could be paired or grouped according to similarities in the characteristics that are considered crucial to outcomes and then conducted random assignment within pairs or groups. A key blocking factor was average student's pre-treatment reading scores across intervention groups, as pre-treatment reading ability is typically highly correlated with post-treatment reading ability.

Another critical blocking factor that we used was the district. That is, we conducted random assignment of pairs or groups of schools within districts to hold constant district policies such as teacher hiring, compensation, and professional development. Most important, conducting random assignment within a school district holds constant the curriculum and standard texts used in the classroom.

3. Methods to Maximize Response Rates and to Deal with Nonresponse

Within the participating schools, the teacher surveys and school records data collections are expected to yield about a 90 percent response rate. We estimate that we will assess 95 percent of the students in fall 2007 and spring 2008 and will complete 95 percent of the classroom observations during that school year. These expected response rates are based on our experience in conducting the first year of data collection in the study. Experienced staff administering the student tests will be trained and monitored by MPR supervisory staff. Sampled students absent on test day will be revisited at least twice for assessment purposes. Telephone follow-up for nonresponse will begin about two weeks after the second mailing of teacher surveys. We will also prompt schools by telephone to complete and return the school records forms.

MPR has developed and refined a wide range of methods to minimize attrition from survey samples and to maximize response. These methods focus on minimizing burden on respondents

and conducting intensive locating efforts, but also include techniques for avoiding refusals. MPR has found that the following techniques are major contributors to a high completion rate: establishing positive relationships with respondents and school and program staff; sending advance letters; establishing efficient and flexible scheduling; and making multiple attempts to schedule data collection from students who are absent from school when data is collected for most students.

4. Tests of Procedures and Methods to Be Undertaken

We will use the same instruments in Year 2 that were used in Year 1 of the study. In our design, we drew heavily on questions and instruments used successfully in previous studies. Consequently, most of the survey questions have been thoroughly tested on large samples with prior OMB approval. In addition, each instrument was pretested with up to nine respondents to determine what problems might arise in providing requested information and to make appropriate changes to the questionnaires, as needed. Responses and comments on the instruments were collected by mail and telephone from teachers and school personnel. The results of the pretest were used to make revisions to the instruments.

We also tested the classroom observation forms during the pilot year, and we revised them forms accordingly. In addition, we taped classroom instruction for each intervention during the pilot year, and used them in the training of observers for the full implementation. This allowed us to assess classroom observers for consistent and reliable coding of classroom instruction for each intervention.

5. Individuals Consulted on Statistical Aspects of the Design

The following individuals were consulted on the statistical aspects of the Evaluation of Reading Comprehension Programs:

Dr. Thomas Cook, Northwestern University

Dr. David Francis, University of Houston

Dr. Larry Hedges, University of Chicago

Dr. Mark Dynarski, MPR

Dr. Steve Glazerman, MPR