

Contract No.: ED-04-CO-0112/007
MPR Reference No.: 6398-113

MATHEMATICA
Policy Research, Inc.

**An Impact Evaluation of
Moving High-Performing
Teachers to Low-
Performing Schools**

***Part B: Supporting Statement
for Paperwork Reduction Act
Submission***

July 2009

Submitted to:

U.S. Department of Education
Institute of Education Sciences
555 New Jersey Avenue, NW
Washington, DC 20208
Telephone: (202) 208-7169

Project Officer: Elizabeth Warner

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave. SW, Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Steve Glazerman

CONTENTS

Chapter	Page
PART B: SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION.....	1
B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS.....	1
1. Respondent Universe and Sampling Methods.....	1
a. District Selection.....	1
b. School Selection.....	2
2. Statistical Methods for Sample Selection and Degree of Accuracy Needed.....	3
a. Evaluating the Impact on Student Achievement.....	4
b. Identifying Factors that Predict Teacher Transfers to Low-Performing Schools.....	5
3. Methods to Maximize Response Rates.....	6
4. Testing.....	6
5. Individuals Consulted on the Statistical Aspects of the Design.....	7

PART B: SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION

This submission is a request for approval of data collection activities that will be used to support An Impact Evaluation of Moving High-Performing Teachers to Low-Performing Schools. This evaluation is being funded by the Institute of Education Sciences (IES), U.S. Department of Education (ED); it is being implemented by Mathematica Policy Research, Inc. (MPR) and its subcontractors – The New Teacher Project (TNTP) and Optimal Solutions Group (OSG). The program being evaluated is called the Talent Transfer Initiative (TTI). This program uses value-added student learning gains to identify teachers with consistently high performance and offers them recruitment and retention bonuses to transfer to schools identified as low-performing based on average student test scores.

This is the second submission of a two-stage clearance request. The first submission (approved on November 5, 2008 under OMB number 1850-0861) requested approval to recruit school districts for the study, collect student records data from recruited districts, and administer a survey to 64 teachers participating in a pilot study. In this package, IES is requesting approval for all data collection activities that will support the full-scale study.

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

The study targets low-performing schools within school districts that face a problem the Talent Transfer Initiative (TTI) is designed to address: the best teachers may not be serving the neediest students. The study does not aim to make statements that generalize beyond the districts and schools, so these are not statistically sampled. Rather, the process of district and school selection, described below, identifies the respondent universe for the administrative data collection covered by this request. Two types of administrative data are used in this study: student records (linked to teachers) and teacher rosters (usually supplied by schools).

a. District Selection

It is not generally known which school districts have a teacher quality imbalance because teacher quality is not routinely measured and reported. (That is something this study intends to do once MPR identifies the districts and obtains the data). For this reason, the study team identified districts for possible inclusion as those meeting objective criteria that we believed would predict suitability for the program and its evaluation, and then narrowed the list based on subjective assessments of district officials' willingness and ability to implement the program.

The objective criteria were based on data from the National Center for Education Statistics (NCES) Common Core of Data. The initial list of school districts was made up of those with at least 30 elementary schools altogether, with at least 10 elementary schools having a high

percentage of low-income students (more than 70 percent eligible for free or reduced price lunch) and at least 10 schools having a low percentage of low-income students (less than 40 percent free or reduced price lunch). This rule was used to capture the size and diversity because the study required that each district have a large enough pool of low-performing schools (the source of potential treatment and control groups) to support the experimental design of the study, and be balanced by a sufficiently large number of relatively higher-performing schools to generate a pool of potential high-performing transfer program teachers (i.e., those teachers to invite into the program).

The study team then limited the sample of districts to those with administrative data sufficient to support the program. Specifically, the data had to support the estimation of teacher value-added indicators, drawing on at least two, and preferably three, prior years of information. These variables included teacher and student identification codes that are unique, non-repeated, and consistent over time, as well as linked across years. The specific data elements required of each district were described in the 2008 submission to OMB covering pilot data recruitment. We will emphasize in the final reports that the study was conducted in districts that met these requirements and caution should be exercised in generalizing to other districts.

Finally, to determine districts' need for the program, MPR recruiters contacted the identified districts and queried officials—such as heads of human resources, chief academic officers, chief accountability officers, and superintendents—to gauge local demand for an intervention such as the TTI. Because the final stage was voluntary, the sample of districts is not statistically representative of a well defined population.

b. School Selection

The universe of schools for the study consists of low-performing elementary and middle schools, where the definition of low-performing is based on recent student achievement levels. Once districts were selected, schools were purposively selected to participate in the study using the following sequence:

- Schools were ranked by their average achievement in the most recent year for which data were available.
- They were then selected in rank order from the bottom up; approximately 20 percent of the eligible schools on the list were invited to be in the study.
- The subset of the eligible schools was identified based on those with a teaching vacancy in one of the tested grades and subjects (grades 4-5 in elementary schools and grades 6-8 math or English/language arts in middle schools), and a consenting principal.
- Eligible schools were then randomly assigned to either a treatment group that would participate in the program or a control group that would fill its vacancy using normal procedures.

The proposed primary data collection has three components (described in more detail in Part A and in the appendices), each of which has a different universe. For all three, the research team proposes a complete census of eligible respondents.

- **Teacher Background Survey (teachers).** The respondent universe for the teacher background survey (whose title is “Teacher Career and Satisfaction Survey”) will consist of all teachers in the targeted grades in TTI program schools (schools that were eligible to hire a TTI teacher) and in control schools. All teachers in targeted grades in control schools will be included in test score analysis. Target grades will be identified prior to random assignment of schools by compiling a list of teaching vacancies in each of the tested grades and subjects in eligible schools. The intention is to target all tested grades and subjects (typically, elementary grades 4-5—all subjects, grades 6-8 math, and grades 6-8 reading/language arts) in which there is at least one expected teaching vacancy. The expectation is that most schools will have one or two vacancies in those grades and subjects. Section 2 provides a detailed explanation of the sample sizes needed for the study.
- **Principal Survey (principals).** The respondent universe for the Principal Survey will consist of the principal or designated assistant or vice principal at each school (program and control) selected for the study. The designee must be familiar with the hiring practices from the previous year and with teacher performance and collegiality in the school.
- **Candidate Survey (teachers).** The respondent universe for the Candidate Survey consists of all teachers in selected districts who have been determined as eligible for the transfer incentive under the TTI rules. Eligible candidates will be the high-performing teachers who are not already teaching in low-performing schools or in schools that were exempted from the program by the district. Candidates who leave the district before being notified of the program opportunity also will be excluded. The intention is to conduct a complete census of eligible candidates.

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

The sample size requirements for the study were developed by identifying the numbers of teachers and schools necessary to answer the study’s main research questions with a reasonable degree of precision. The precision standards used by the Department of Education require that hypothesis tests be conducted with a significance level of five percent.

Based on experiences in the pilot study, MPR estimates that the study will require 120 schools, split evenly between treatment and control, as well as 200 teaching vacancies, also split evenly between treatment and control. In addition, MPR estimates that about 10 school districts will be required to generate a sufficient sample and will proceed to select the sample purposively as discussed above.

To understand how the survey samples are formed, it is necessary to understand how the sample is spread across and within schools. MPR estimates that the 200 teaching vacancies will be spread over 150 grade levels within the 120 schools, so the typical school would have one or two vacancies in the same grade/subject and possibly one additional vacancy in a different grade or subjects. These numbers imply 1.67 teachers per school (for example, 40 schools with one teacher and 80 schools with two teachers) and 1.33 grade levels per schools (for example, 90 schools with one grade level and 30 schools with two grade levels). One way to achieve this configuration is to have 40 schools with one teacher in one grade, 50 schools with two teachers in one grade, and the remaining 30 schools with two teachers in different grades. MPR also assumes there will be four teachers per grade/subject combination, which would result in a population of 600 affected teachers from whom we would seek a background survey. The sample size justification is described in more detail below.

a. Evaluating the Impact on Student Achievement

Estimation of impact on student achievement will draw on data from the teacher background survey, Principal Survey, and student records.

Size of Sample Needed to Achieve Statistical Precision. The sample available for estimating impacts on student achievement will consist of the 200 classrooms initially identified as teaching vacancies, plus the approximately 400 classrooms in the same (150) grades in the same (120) schools, assuming 4 classrooms per grade. With a conservative assumption that data will be available for 16 students per classroom, the sample will include 9,600 students in each of the two study years. These data will be used to estimate the total impact, which is based on comparing entire grades within treatment schools to the corresponding grades/subjects within control schools.

In addition to total impacts, the study will also compute direct and indirect impacts of the program, as discussed in Part A, using subgroups of the overall sample. The analysis of direct impacts will be conducted using classrooms that were filled by TTI candidates in treatment schools and by new hires in control schools, while the analysis of indirect impacts will be based on classrooms taught by the remaining classrooms in the same grade/subject. Thus, the portion of the sample available to estimate direct impacts will be approximately 200 classrooms with 3,200 students, leaving the remainder of 400 classrooms with 6,400 students for estimating indirect impacts.

Samples of these sizes allow us to detect impacts that are approximately 13 to 15 percent of a standard deviation in test scores. For the direct effect, we calculate that the study will be able to detect 14.5 percent of a standard deviation.¹ These minimum detectable effect (MDE) calculations are estimated for an 80 percent power level and a 5 percent statistical significance level (two-tailed test). The calculations take into account the clustering of students in schools, assuming that variance between schools accounts for 10 percent of the total variance in student achievement. We assume that prior test scores explain 50 percent of the variance in post-test scores, and that principal and teacher survey data will allow us to reduce the variance in outcomes at the school level by 20 percent.

The corresponding calculations for the indirect effect suggest an MDE of 13.8 percent and, for total effects, 13.5 percent. (The MDEs are similar because they are all based on the same number of schools and school/grade combinations).

These precision levels are adequate to address the study’s main research questions related to test score impacts. If we assume that the direct effect (including the role of any distributional effects) is 25 percent of a standard deviation and the indirect effect is 15, then the total effect will be approximately 18. This assumption is based on historical performance of high-performing teachers. The high-performing teachers selected for the study were those who consistently performed in the top 20 percent of the distribution of teachers in the district teaching similar grades and subjects over the same time period of two or three years. The research team determined based on pilot data and early full-scale value added estimates that teachers in this top tier tended to produce average student achievement gains that were 20 to 35 percent of a standard deviation (in terms of student-level variance) above the average for all teachers, depending on the district, grade, and subject of the test. If new hires tend to be below average because many are new to the profession, then the expected direct effect should be even larger. However, we require the greater precision than 20 to 35 percent of a standard deviation in order to measure how this impact is spread across direct and indirect effects.

b. Identifying Factors that Predict Teacher Transfers to Low-Performing Schools

Another goal of the study is to identify the factors that determine whether a teacher who is offered an incentive will apply to transfer to a low performing school, will interview, and will be placed in such as school. For this analysis we rely heavily on the Candidate Survey. The immediate goal of the analysis is to identify the strength of the relationship between teacher-

¹ This minimum detectable effect was calculated under the assumptions noted in the text, using the following formula (Schochet 2005):

$$2.8\sqrt{\frac{2\rho(1 - R_{BS}^2)}{.5s} + \frac{2(1 - \rho)(1 - R_{WS}^2)}{.5sknr}}$$

where ρ is the proportion of total variance in student achievement that lies between schools (that is, the intraclass correlation at the school level); R_{BS}^2 and R_{WS}^2 are the proportions of the between-school variance and the within-school (between-student) variance, respectively, that are explained by the regression model; s is the number of schools; k is the average number of teachers/classrooms per school; n is the average number of students per classroom; and r is the proportion of students for whom the study will have achievement data, an assumed 80 percent.

specific factors and the variables that characterize whether the teacher applied to, interviewed for, and was placed in a low-performing school.

Sample Size and Justification. The study has been designed to detect an effect of each explanatory variable of approximately 5 percentage points on the probability of a candidate's transferring to a low-performing school. Typical explanatory variables include a teacher's satisfaction with his or her current position, or an increase in commuting distance to a prospective position. The anticipated underlying probability is 10 percent, based on experience from the pilot study, in which approximately 10 candidates were needed to fill each elementary school slot. Using the same assumptions listed above, we calculate that the sample size must be approximately 600 teachers to detect an effect of this size. We do not yet know the size of the universe of high-performing teachers eligible for transfer, but it may be such that a complete census will be necessary to generate adequate sample size. Assuming an 80 percent response rate, we would need about 750 candidates to be eligible initially. (Eighty percent of 750 eligible sample members would yield 600 complete surveys).

If the number of identified candidates is substantially higher than 750, we will draw a random sample; this sample would be stratified to ensure the proportional representation of successful transfers, as well as candidates who did not apply (nonapplicants) or applied but did not interview or successfully transfer (withdrawals and screenouts). In other words, if necessary, we will draw samples with the same selection probability in each stratum (no oversampling) until we reach the target sample size.

3. Methods to Maximize Response Rates

Response rates of about 80 percent are anticipated for the Candidate, teacher background, and Principal Surveys. MPR expects to collect student demographics and test scores in all participating districts and teacher rosters in all schools. To ensure such high response rates on the survey, follow-up methods, including second mailouts, e-mail prompts, telephone prompts, and telephone interviews for nonrespondents, will be used. We also are requesting approval in this clearance to provide a \$25 incentive for the Candidate Survey and teacher background survey respondents to maximize response rates.

4. Testing

The survey instruments were designed by drawing heavily on questions from instruments used successfully in previous studies. Consequently, most of the survey questions have been tested thoroughly on large samples, with prior OMB approval. Furthermore, the Candidate Survey was fielded in the pilot study, and the other instruments will be pretested in spring 2009 with up to nine respondents to determine what problems respondents might have in providing the requested information and make appropriate changes to the questionnaires, as needed. Responses and comments on the instruments will be collected by mail and telephone from teachers and principals. The results of the pretest will be used to make revisions to the instruments prior to the full-scale study.

5. Individuals Consulted on the Statistical Aspects of the Design

The following Technical Working Group Members were consulted on various aspects of the statistical design:

- Dale Ballou (Vanderbilt University)
- Tom Kane (Harvard Graduate School of Education)
- Rob Meyer (University of Wisconsin Center for Education Research)
- Tony Milanowski (University of Wisconsin Center for Education Research)
- Jeff Smith (University of Michigan)
- Jake Vigdor (Duke University)