

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

MATHEMATICA
Policy Research, Inc.

**Evaluation Of Moving
High-Performing
Teachers To Low-
Performing Schools**

*Part B Supporting Statement
for Paperwork Reduction Act
Submission*

August 22, 2008

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. S.W., Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Steve Glazerman

CONTENTS

	Page
PART A: SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION	
A. JUSTIFICATION.....	A-1
1. Circumstances Necessitating the Collection of Information.....	A-1
2. Purposes and Uses of the Data.....	A-5
3. Use of Technology to Reduce Burden.....	A-6
4. Efforts to Avoid Duplication.....	A-6
5. Methods to Minimize Burden on Small Entities.....	A-6
6. Consequences of Not Collecting Data.....	A-7
7. Special Circumstances.....	A-7
8. Federal Register Announcement and Consultation.....	A-7
a. Federal Register Announcement.....	A-7
b. Consultations Outside the Agency.....	A-7
c. Unresolved Issues.....	A-7
9. Payments or Gifts.....	A-8
10. Assurances of Confidentiality.....	A-8
11. Additional Justification for Sensitive Questions.....	A-9
12. Estimates of Hours Burden.....	A10
13. Estimates of Cost Burden to Respondents.....	A-10
14. Estimates of Annual Costs to the Federal Government.....	A-10
15. Reasons for Program Changes or Adjustments.....	A-10
16. Plans for Tabulation and Publication of Results.....	A-11
a. Tabulating Descriptive Information.....	A-11
b. Estimating Teacher Value Added.....	A-11
c. Estimating Impacts of the Master Teacher Residency Program.....	A-13
17. Approval to Not Display the OMB Expiration Date.....	A-14
18. Explanation of Exceptions.....	A-14

CONTENTS *(continued)*

Page

**PART B: SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT
SUBMISSION**

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS.....	B-2
1. Respondent Universe and Sampling Methods.....	B-2
2. Statistical Methods for Sample Selection and Degree of Accuracy Needed.....	B-4
a. Evaluation of Impact on Student Achievement.....	B-4
b. Candidate Survey.....	B-5
3. Methods to Maximize Response Rates.....	B-5
4. Pilot Testing.....	B-6
5. Individuals Consulted on the Statistical Aspects of the Design.....	B-6

APPENDICES

**APPENDIX A: COVER LETTER FOR MASTER TEACHER RESIDENCY PROGRAM
CANDIDATES TO ACCOMPANY CANDIDATE SURVEY**

APPENDIX B: PRETEST MATERIALS

Pretest Draft for the Master Teacher Residency Program Candidates Survey

APPENDIX C: DRAFT RECRUITMENT MATERIALS

- C1. District Recruitment Protocol
- C2. Principal Recruitment Letter
- C3. Information Sheet For Principals
- C4. Letter of Introduction and Invitation to Apply to Program for Eligible
Teacher Candidates

APPENDIX D: CONFIDENTIALITY PLEDGE

APPENDIX E: REFERENCES

PART B: SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT SUBMISSION

This OMB package requests clearance to recruit school districts for an upcoming evaluation to test the effect of teacher incentives designed to move high-performing teachers to targeted low-performing schools (hence, the evaluation is titled, “Moving High-Performing Teachers to Low-Performing Schools.”) The U.S. Department of Education’s National Center for Education Evaluation is conducting the study, with its contractor Mathematica Policy Research, Inc. (MPR) and two subcontractors, The New Teacher Project (TNTP) and Optimal Solutions Group, Inc.

The evaluation aims to estimate the impact of the high-performing teachers on the low-performing schools that they transfer to. The evaluation design is a randomized experiment in which the researchers will randomly assign schools that have a teaching vacancy in targeted grades and subjects to an intervention group or a control group. High-performing teachers will be offered bonuses for transferring to and remaining in the intervention schools for two years. Control schools will fill their teaching vacancies the way they normally would if they were not part of a study. The intervention is called the Master Teacher Residency Program (MTRP) and the high-performing teachers are referred to as Master Teachers. We will compare student achievement and other outcomes between the intervention and control schools to estimate the impact of the intervention.

In addition to the clearance request for recruiting, we are requesting clearance to collect student records data from those recruited districts and administer a data collection form to a group of 61 teachers participating in a pilot study that will be conducted for the 2008-09 school year. We refer to this teacher data collection form as the “Candidate Survey.”

This request is the first of two. A future request will seek clearance to collect additional teacher and principal survey data associated with the evaluation. With the exception of the Candidate Survey covered in this request, we are not requesting any clearance for data collection forms during the pilot study that is currently taking place. The pilot study involves one district and will allow us to pretest three additional surveys—one of teachers, one of principals, and one of district human resources staff—on a sample of 9 or fewer individuals each.

The study is submitting the package in two stages because site identification and recruitment must begin before all the data collection instruments are developed and pretested and because implementing the Candidate Survey with a full sample of 61 teachers will allow us to learn the appropriate lessons from the pilot study before moving on to the planned full study. The draft letter requesting teacher participation in the Candidate Survey is contained in Appendix A, the draft Candidate Survey to be used in the pretest in Appendix B, recruitment materials in Appendices C1-C4, and MPR’s internal confidentiality pledge in Appendix D. References appear in Appendix E.

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

The study will not statistically sample districts and schools. Instead, the study will intentionally target districts with problems that the MTRP is designed to solve and in which the intervention would be feasible to implement. The evaluation does not aim to make statements that generalize beyond the districts and schools under study. Specific district recruitment criteria are discussed below and listed in more detail in a district recruitment protocol in Appendix C-1.

The MTRP study will recruit districts to participate in an evaluation designed to measure the impact of high-performing teachers on student achievement in low-performing schools. The study will recruit districts, schools, and high-performing teachers. Below we describe the process we will use to identify suitable districts and schools and to recruit them to participate in the study.

Identification of Districts and Schools. As a starting point for identifying potential districts and schools for inclusion in the study, the study will consider the suitability of the nearly 40 school districts of various sizes proposed by the study's subcontractors (The New Teacher Program) and the study's Technical Working Group. Approximately 30 additional large urban districts will be identified through the Common Core of Data Universe File (CCD) and other sources including information MPR has on hundreds of districts with whom we have worked on previous studies. The study will determine whether each district meets three criteria:

1. Encompasses a number of low-performing schools (the source of a potential treatment group) that is large enough to support the experimental design of the study and is relatively balanced by the number of high-performing schools (the source of potential participating teachers).
2. Maintains and will make available the linked student-teacher data (including test scores) that we will need to identify high-performing teachers and estimate impacts (see Section A.16); and, will provide the data in an appropriate format within the study's time requirement.
3. Demonstrates a willingness to adopt the MTRP. Participation in the evaluation effort is voluntary.

To address the first criterion, information from the CCD will be combined with other sources of information, including district and school report cards, to assess districts' suitability for the study. Schools will be defined as "low-performing" primarily based on failure to make adequate yearly progress (AYP) during the school year(s) preceding recruitment. Information on school AYP status will be gathered on district and state Web sites.

To address the second criterion, past experiences using district achievement data and contacts to district research and assessment staff or data warehouse staff as necessary will be made to ensure that recruitment efforts focus on districts with the ability to meet the data requirements of the study (including their ability to provide student-teacher linked achievement test scores).

To address the third criterion, we will judge willingness to adopt the MTRP based on the contractors' prior knowledge and understanding of teacher unions, district leaders, and the reform climate in districts around the country. For example, certain districts work MPR and TNTP and have stated that they are unwilling to adopt any policy changes that include salary augmentation tied to student achievement. Such districts will be dropped from the list.

District and School Recruitment. First, the study team will contact the estimated 70 districts that will be identified through the process described above. Second, we will visit an estimated 25 districts, those that signal an interest and willingness to execute the strategy we outline, to discuss in detail the purposes and requirements of the study—including how random assignment will be implemented. These visits will include high-ranking district officials to generate and gauge interest in the study and explore information on district suitability—for example, verifying the existence and identities of low- and high-performing schools and the suitability of student achievement data for value-added estimation and impact estimation. Third, if the study team is able to verify that the district meets the selection criteria, we will arrange a follow-up meeting with personnel such as the chief human resources officer, the director of research and evaluation, the legal counsel, and if necessary, school board and teachers' union representatives.

During the calls and visits, we will explain the purpose and importance of the study; describe the responsibilities associated with participation; outline the necessary conditions for district, school, and teacher participation; describe the timelines for various pilot study activities; convey appropriate guarantees regarding data confidentiality; highlight potential benefits to students, schools, districts, and education policymakers; and describe the reports that we will produce.

Once a district agrees to participate, the study team will work with the district to identify schools and obtain principal support. We will invite principals to information sessions. Many districts will want to recruit their own low-performing schools into the study and to centralize all communications with schools. For others, the study will contact principals by mail, by phone and/or in-person (see draft materials in Appendices C1 – C4). Key to sustaining principal support is to ensure that principals understand the purpose of the study, especially the random assignment and any data requirements, and agree to participate whether their schools are assigned to treatment or control conditions.

Estimating Teacher Value Added

In both the pilot and the full study, the credibility of the MTRP depends on our ability to identify high-performing teachers. Our approach relies on a critical first stage of selecting only those teachers with a proven track record of raising student achievement. We identify teachers with such a track record by using several years of achievement data to estimate teachers' value-added—the unique contribution that a teacher makes to student achievement growth in a typical year. Using the estimates of each teacher's value-added, we will identify a list of high-performing teachers from the upper tail of the performance distribution. We will then target those teachers for the next stage of teacher recruitment.

The value-added model of teacher performance is based on a student achievement growth model which controls for students' prior achievement and student background. Specifically, the

model will be a regression of a student’s test score in a given subject and year on the student’s previous-year score, background characteristics, and a teacher effect:

$$(1) \quad Y_{ijt} = \mu + \theta Y_{ijt-1} + \beta X_{it} + \gamma_j T_{ijt} + \tau_t + \varepsilon_{ijt}$$

where Y_{ijt} is the year t test score for student i in the subject taught by teacher j ; Y_{ijt-1} is the previous-year test score; X_{it} is a set of student characteristics included as control variables; T_{ijt} is a teacher dosage variable indicating the fraction of the year the student was taught by teacher j ; τ is a fixed year effect;¹ ε is a random error term; and μ , θ , β , γ , and τ are parameters to be estimated. Our goal is to estimate the coefficients γ_j on the teacher dosage variables to measure each teacher’s value added.

Volatility in Test Scores. Random, one-time events that are not related to achievement but that affect scores can be a problem in that, by necessity, teachers work with a small sample of students each year. Such volatility can undermine attempts to estimate meaningful rankings of teacher performance (Kane and Staiger 2002). To account for this imprecision, we will follow Kane and Staiger’s recommendation to aggregate test score information over several years to estimate teacher effectiveness. Specifically, we will pool at least three years of student learning gains in estimating the value-added model.

Errors in Variables and Attenuation Bias. Another concern with value-added models is the well-documented attenuation bias that results from including an explanatory variable (the pretest) that is measured with error (Meyer 1992). While the bias in the coefficient on the pretest is known to be attenuation, that is, biased towards zero, the bias in the teacher effects, which are the main parameters of interest, is of unknown direction.

We will deal with measurement error by estimating the reliability of the pretest and subtracting the estimated reliability from the diagonal elements of the cross-product matrix formed to compute the regression coefficients.² As a test of the robustness of this approach and of the estimates to assumptions about errors in variables in general, we will recompute the teacher value-added estimates using gain scores (post-test minus pretest) and errors-in-variables regressions with different values of the reliability estimate and each time examine how our list of top teachers changes. If a teacher initially identified as high performing fails to remain in that group under alternative models, then we will examine the specific patterns of achievement gains and student characteristics to determine if the problem lies with the base model or the alternative models and make necessary corrections.

Sampling Error. There will always be some margin of error in the estimation of teachers’ value-added, even with several years of data. We aim to determine the cutoffs for high performance in such a way that minimizes the probability of including a teacher who might have a high value-added score by pure chance. We will use the estimated standard error for each teacher effect to characterize the uncertainty with which the teacher falls in the high performing category. One way to use this information is to apply an empirical Bayes or “shrinkage”

¹ For simplicity we present the model here for a single grade level, but we will estimate a pooled model that puts all grade levels on a common scale and includes year-by-grade fixed effects instead of just year effects.

² This can be accomplished in Stata by using the `eivreg` command.

estimator that replaces each teacher effect with a weighted average of the mean teacher effect and the observed one, with the weights being a function of the standard error of measurement. Less precisely estimated teacher effects (those based on less data) will “shrink” closer to the mean teacher effect.

Identification of High-Performing Teachers and Control Teachers. Identification of high-performing teachers for potential inclusion in the study will occur after districts have been successfully recruited.

To be eligible for the MTRP, teachers must instruct students in a targeted elementary or middle school grade or subject (a grade or subject for which the district administers annual achievement tests). High-performing teachers in participating districts will be identified based upon their estimated value-added, a measure of the unique contribution that a teacher makes to student achievement growth in a typical year and measured by student achievement, controlling for student’s past-year achievement. Candidates who meet the criteria will be invited to participate in the MTRP.

In addition to eligible high-performing teachers (candidates for the MTRP), non-candidate teachers hired for vacancies in the targeted grades in the control schools will be included in the study, as will principals of intervention and control schools. The involvement in data collection of the newly hired non-candidates and the principals will be described in a future submission for OMB clearance, which will include the New Hire Survey and Principal Survey mentioned in section A.1.

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

Below we describe the sample sizes and degree of accuracy needed for the data collections for which we are currently requesting clearance.

a. Evaluation of Impact on Student Achievement

Sample Size. Following a pilot study, the full study will include 160 schools, each of which has a teaching vacancy in one of the targeted grades/subjects. We will use a random number generator to assign 80 schools to a treatment group in which the targeted slot may be filled with a Master Teacher and 80 schools to a control group that fills the targeted vacant teaching slots as they normally would. We anticipate that we will be able to successfully recruit 10 purposively selected school districts that meet the criteria described above. We will identify an average of 16 purposively but systematically selected schools (8 treatment and 8 control) per district. The average sample size does not preclude variation in sample size by district to achieve our overall sample size target.

All schools must have at least one teaching vacancy in a tested grade and subject in order to be included in the study. In total, we expect 200 vacant slots to be filled by study teachers (100 treatment and 100 control), or an average of 1.25 vacant slots per school. We will therefore include approximately 10 treatment teachers and 10 control teachers per district.

We assume an average of 23 students per classroom. We further assume that approximately 20 percent of students will be lost to the study through either attrition or missing data. Thus, we expect a sample of 3,680 students in the 200 classrooms of study teachers spread across the 160 schools.

Degree of Accuracy Needed. We have designed the study to be able to detect impacts of a magnitude that the Government has deemed policy relevant, between 15 and 20 percent of a standard deviation. (For a student starting at the 25th percentile of the achievement distribution, this would correspond to an increase to the 29th or 32nd percentile, respectively.) The corresponding sample size requirement of 160 schools (80 treatment and 80 control) is based on calculations that take into account the presence of covariates, the group randomized design, and the natural clustering of students in classrooms and classrooms in schools.

Our calculations suggest that the study design can detect an impact of 19 percent of a standard deviation.³ This minimum detectable effect calculation is estimated for an 80 percent power level and a 5 percent statistical significance level (two-tailed test). The calculations conservatively assume that variance between schools accounts for 20 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.

Subgroup sample sizes and the precision of the subgroup impact estimates will be lower than those for the full sample. A 50 percent subgroup of students from the 160 schools would provide us with a minimum detectable effect of 21 percent of a standard deviation.

b. Candidate Survey

Sample Size and Justification. Approximately 61 high-performing teachers (Master Teacher candidates) will be asked to complete the Candidate Survey during the pilot. This represents the universe of candidates and is not a statistical sample (i.e. it represents a census). Besides informing how we conduct the Candidate Survey for the full-scale study, the pilot Candidate Survey is a critical means for improving the intervention before it is rolled out on a larger scale for the full study. The information we gather on the reasons that candidates offer for their decision to apply or not apply for a transfer bonus, their reaction to the offer of recruitment and transfer bonuses, and the experiences that applicants have during the school-matching process represent novel data that cannot be found in any previous study. Information from all 61 candidates, which was the maximum number feasible to include in a pilot study, will be the minimum required to make sensible decisions about how to structure or restructure the intervention when it is scaled up for the full study.

3. Methods to Maximize Response Rates

We anticipate response rates of about 90 percent for the Candidate Survey and expect to collect student test scores in all participating districts. To ensure high response rates such as these on the survey, we will employ follow-up methods including second mailouts, telephone prompts, and telephone interviews for nonrespondents.

4. Pilot Testing

The Candidate Survey itself is a pilot and will be used to inform the full study request for clearance to be submitted in December 2008. However, given the novelty of the instrument and

³ This minimum detectable effect was calculated under the assumptions noted in the text using the following formula:

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

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.

our plans to administer the survey to 61 teachers, we intend to pre-test the instrument on 9 eligible respondents, de-briefing them on the duration of the questionnaire and the clarity of the items.

5. Individuals Consulted on the Statistical Aspects of the Design

John Hall, Senior Statistician, Mathematica Policy Research

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

Dale Ballou (Vanderbilt University)

Brad Jupp (Denver Public Schools)

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)

Louise Sundin (Minneapolis Federation of Teachers, formerly)

Jake Vigdor (Duke University)