

Contract Number:  
ED-IES-10-C (0001)

Mathematica Reference Number:  
06748-510

Submitted to:  
Institute of Education Sciences  
U.S. Department of Education  
555 New Jersey Avenue, NW  
Washington, DC 20208  
Project Officer: Melanie Ali

Submitted by:  
Mathematica Policy Research  
600 Maryland Avenue, SW  
Suite 550  
Washington, DC 20024-2512  
Telephone: (202) 484-9220  
Facsimile: (202) 863-1763  
Project Director: Philip Gleason

**Responses to OMB Questions  
on A Study of Teacher  
Residency Programs**

March 10, 2011



**MATHEMATICA**  
Policy Research, Inc.

The questions we received from OMB on 3/7/2011 on our information collection package for the A Study of Teacher Residency Program (TRP), along with our responses, are presented below.

**1. How likely is it that you will be able to recruit the target number of 8 districts for the outcomes study? Will the analysis be feasible or the sample size sufficiently large if you are unable to recruit the target number of districts for the outcomes study?**

We believe there is a good chance that we will be able to recruit enough programs (and thus districts) to conduct the planned outcomes analyses in a way that will provide meaningful information to policymakers and program administrators. The current environment of tight budgets and great attention to fiscal discipline at all levels of government will present some challenges in our recruiting of districts, but we have successfully recruited a large number of programs and districts in prior studies and we will make use of that experience in the recruiting effort for this study. The target of 8 programs is not a hard and fast goal or critical target. It is merely an estimate that has guided the study since its inception, an approximation of how many programs might be needed to obtain a *teacher sample* size large enough to conduct worthwhile analyses of student outcomes. The number of eligible teachers from TRPs is actually more important than the number of programs or districts. A final sample of 6 or 7 programs might be fine, if it included the largest programs, in terms of number of eligible teachers. Similarly, if some of the largest programs, in terms of number of eligible teachers, either decline to participate in the outcomes study, or lack the technical capacity to participate (e.g., cannot provide linked student-teacher test scores with consistent ID numbers over time), it may be useful to recruit more than 8 programs.

**2. Are there aspects of this study that will duplicate data collection by the program staff (i.e. retention rates)? Can this be streamlined?**

While some of the data needed for the study will overlap with data that grantees submit for GPRRA requirements, we believe it is essential to collect retention information and other potential overlapping items through the study. First, the level of detail included in the annual performance reports will vary significantly across grantees because they are not collected in a systematic way; the variation will make it challenging to present descriptive characteristics of the programs that use consistent definitions of the characteristics and rely on uniformly high quality data. Second, information collected through the annual reporting requirements may not have the timeliness or quality that this study needs. It is possible that different programs obtain information

on attributes such as retention in different ways, with varying degrees of care and attention. For example, some programs may stop following their teachers at the end of the commitment period. Third, the mobility survey will allow us to collect nuanced data on outcomes like retention, such as whether teachers remain the profession even if they leave the district, or whether teachers who remain in the district change schools, grades, or subjects. In addition, we may learn if a teacher made these changes because he or she felt inadequately prepared. Lastly, data that programs collect will generally cover only TRP teachers, while we are interested in benchmarking key outcomes among TRP teachers against these same outcomes among non-TRP novice teachers in the district.

**3. Is the only purpose of the request for teacher employment data to verify whether the novice teachers are still employed by a district? Is it not possible to verify this by using just the teacher mobility survey without having to request the employment data?**

The purpose of the teacher employment data is to verify whether novice teachers are still employed by the district. While the mobility survey is designed to provide similar information on teacher retention, a limitation is that we will not be able to complete the survey for all teachers in the sample. In fact, the teachers who we suspect will be most difficult to locate and complete the mobility survey will be those who will have changed jobs and potentially moved out of the area. If we rely fully on the mobility survey, we will not have information on retention among those teachers who do not complete the survey. In this way, the teacher employment verification data will complement data from the mobility survey.

To reduce the burden of the data collection effort, a possible change we could make would be to eliminate any overlap between the employment verification data and data from the mobility survey. We would initially collect the employment verification data for the full set of teachers. However, if that source of data indicated that a teacher is still teaching in the district, we would drop that teacher from the mobility survey sample. In the end, we expect that we would administer the mobility survey to about 10 - 20% of the original sample. Eliminating the overlap between these data sources in this way will have a cost in terms of information lost. We would lose two pieces of information on teachers who are still in the district: (1) for those who have changed schools, the reason for that move; (2) the chance to get any updated contact information (relevant for the first administration of the survey). In addition, for teachers reported to be still teaching at a given school in the district, we would lose the chance

to confirm the accuracy of that information (in particular, the teacher's current school).

**4. Does IES plan to collect any cost data as part of the evaluation? Given the high costs of a residency program relative to other teacher preparation programs, it would be very important to put any differences in value-added in the context of differences in costs. Would it be possible to conduct an exploratory analysis using value-added and some cost metric or another form of cost-benefit analysis?**

The current data collection plan for the Teacher Residency Program study includes a modest amount of cost information collection, but the extent of cost information that is collected as part of the study could be expanded. Below, we summarize the cost information being collected under the current plans for the study, and outline several options for expanding the collection of this type of information. We also describe some of the issues that would arise if this information were used as part of a cost-benefit analysis.

The main instrument we currently use to collect cost information (or information that could be used to generate cost estimates for specific program components) is the program survey, scheduled to be administered to all grantees and any non-grantee programs participating in the in-depth implementation study in spring 2011. Program survey questions that either directly solicit information on program costs or solicit information on program activities that might conceivably be translated to costs using publicly available information include:

- Amount of training provided to new mentors (C5 and C6)
- Compensation to classroom and other mentors (C9 through C12)
- Stipend/salary to residents during residency year (C15)
- Additional payments to resident after they become teachers of record (C16-C17)

The program director interview does not currently include any questions that directly seek information on program costs, though some of the interview questions may provide detailed information on program activities that might be translated to costs. The resident and mentor surveys do not include any questions that provide direct information on program costs.<sup>1</sup>

---

<sup>1</sup> The resident survey does include a question about whether the cost (to the resident) of other teacher preparation programs is higher than that of the TRP they are attending (item

We believe there are three options we could pursue to supplement the existing plans for cost data collections, which are listed below along with their pros and cons. Since these were not included in our data collection plans under our original design, they would add somewhat to the cost of the study as well as to the burden imposed on study respondents.

- ***Supplementing the Existing Cost Questions on the Program Survey***

Since the program survey currently includes some cost questions and is the only data collection instrument that covers the full population of grantees (along with selected non-grantees), the most straightforward way of collecting additional cost information would be to supplement those currently included on that survey. The resulting information would provide a somewhat more complete picture of program costs and would do so for a larger number of programs than either of the other options.

A limitation of asking cost questions on a closed-ended survey, however, is that it limits the extent to which the nuances of real-world program costs can be captured. For example, different programs may have different systems for tracking costs or may not track the specific costs in which we are interested. This is especially true given that TRPs are partnerships of various organizations, and costs of activities conducted primarily by one of the partner organizations may not be known to program staff. For example, the recruitment of students to the program may be conducted by university staff and while program staff may be aware of recruiting activities that are conducted, they may not know the costs of these activities. While these same issues are likely to arise if we attempt to collect cost data using semi-structured interviews, the nature of the interview could allow us somewhat more flexibility to obtain sufficient information to be able to generate reasonable cost estimates.

Given this limitation, adding questions to the program survey would allow us to obtain additional information about the costs of specific components, but we would not recommend attempting to use cost information collected solely on the program survey to develop an overall cost estimate of the program. There would be categories of costs for which we suspect we would not be able to reliably capture cost information through a survey and so the cost information would be incomplete. Further, if we only use the survey to collect cost data,

---

3a) and another question about how the out-of-pocket cost of the program affected their decision to enroll in the TRP (item 5). While these questions are relevant to program costs, we would not be able to use them directly to estimate the overall costs of operating the program.

we may not be fully aware of important differences between programs in the kinds of costs that are covered or are not covered by their responses.

- ***Including a Module of Cost Questions on the Program Director Interview***

The program director interview does not currently include specific questions on the costs of program activities. To address this limitation, we could include a module of questions that would ask directors about the specific activities included in each category and the cost of these activities. In asking these questions, we could incorporate any information directors or their staff will have already provided on the program survey, by confirming these answers or allowing the directors to refine or clarify their responses. We could also ask questions about program activities and their associated costs not currently included on the program survey. For example, we might ask some questions that would provide enough information to better understand program management or student recruitment and selection activities and costs. Since programs may differ markedly in these activities and how their costs are covered, asking questions like these in a semi-structured interview (rather than a closed-ended survey) would be more likely to provide more complete and comparable information across programs.

We do not recommend trying to add to the program director interview a sufficient number of cost questions to allow us to generate an overall cost estimate of the program. The problem with trying to do so would be that collecting enough information to fully understand the complex nature of program activities and costs would likely require an extended set of questions best answered by various staff at the program or at a program partner. It also seems likely that cost information is maintained in different forms at different programs, and so obtaining comparable information would require a flexible approach. For these reasons, we believe that the best way to obtain an overall program cost estimate would be to dedicate a separate interview (and perhaps a site visit to the program). Trying to tack the full set of necessary questions onto the existing program director interview would be likely to weaken the quality of data we obtained in the interview. In fact, we believe that the current version of the program director interview is already fairly long, and we would hesitate to add a substantial module for obtaining cost information without dropping some of the other questions currently in the interview protocol.

- ***Conducting a Separate Interview or Interviews to Obtain Cost Information***

Collecting data of sufficient breadth and nuance to produce overall program cost estimates that would be comparable across programs would require a separate interview (or set of interviews) dedicated entirely to this purpose, perhaps conducted during a site visit. This data collection effort would be structured around the components of program costs. In particular, we would develop an outline of program activities leading to costs—including such categories of program activities as student recruitment and selection, coursework, residency, and additional support and resources provided to graduates once they become teachers of record—and then ask questions and/or collect program data allowing us to develop cost estimates for each set of activities. We anticipate that it would be quite challenging to develop cost estimates for some categories of program activities. For example, coursework offered through the program may be offered as part of a university's regular course offerings, and it may be difficult to separate out costs for the TRP courses from other university costs. Thus, we expect that this cost data collection effort would require interviews and/or collection of administrative records from several different informants. Moreover, we would need to investigate the most appropriate design for collecting this information from programs.

Regardless of the specific approach used, obtaining a sufficient amount of cost information to generate an overall estimate of program costs would be a substantial undertaking. Thus, it is important to understand how this information would be used in the context of this study. One might imagine a TRP cost estimate to be used for one of three purposes: (1) a formal cost-benefit analysis; (2) a comparative study of costs of alternative forms of teacher preparation; and (3) simply as one part of a descriptive study of TRPs.

*Cost-Benefit Analysis.* We do not recommend attempting to use cost information collected as described above as a basis for conducting a cost-benefit analysis for two main reasons. First, we do not plan to collect information that would allow us to calculate program benefits as part of this study. Calculating true program benefits would require that the study's design allow us to calculate program impacts. Through the changes we have made to the study design, we have explicitly moved away from calculating program impacts. We decided it was not feasible to do an impact study, and so have altered the design to generate estimates of program outcomes—in particular, TRP teachers' value added scores. To use the value added scores in a cost-benefit analysis would suggest that we view those scores (or the contrast between the TRP teachers' scores and those of non-TRP novice teachers) as amounting to an impact estimate, which we fear would be misleading.

Second, even if we did generate impact estimates as part of this study (using, for example, the original experimental design), the nature of

those estimates would also make a cost-benefit analysis potentially misleading. In particular, that design would generate estimates of TRP teachers rather than the program alone. In other words, we would be measuring differences in the performance of students of TRP teachers with the performance of otherwise equivalent students who had non-TRP teachers. These differences could have arisen because of either aspects of the program and its effectiveness in preparing teachers for the classroom or aspects of the individual teachers who went through TRP programs. Thus, the cost side of such a cost benefit analysis would reflect costs of the program itself whereas the benefit side would reflect benefits of the program plus benefits (to students) of the individuals who went through the programs.

*Comparative Study.* Rather than comparing the costs of TRPs to the benefits of the program, an alternative approach would be to compare the costs of TRPs to the costs of other teacher preparation programs, traditional or alternative. We believe that such a comparison of costs could be useful, but we will not have access to the costs of other teacher preparation programs as part of this study. Thus, we could only conduct half of such a comparative study.

*Descriptive Study.* Rather than trying to compare TRP costs to either TRP benefits or costs of other teacher preparation programs, we could simply use any additional cost information collected as part of this study to enrich the study's descriptive analysis of TRPs. In other words, in addition to describing features of the programs such as the coursework and residency component, we would also describe the programs by providing a sense of their costs, either separately for individual parts of the program or—if the most intensive cost data collection strategy is pursued—the overall costs of the program. In conjunction with details on the characteristics of the programs and their participants as well as the resulting outcomes among TRP teachers, the cost information would provide a fuller picture of the programs.

**5. Please provide another paragraph or two for the SS giving IES's working definition of an outcome study, why it is the appropriate research design in this case, and what it can tell us and what it can't.**

IES considers an outcomes study to be a study in which data on key outcomes are analyzed and presented for some or all participants in a program. Similar outcomes may also be analyzed for nonparticipants in the program, but if so, this is done without using a design that would enable researchers to say with any real certainty whether any differences between outcomes associated with participants and those associated with nonparticipants are attributable to program



participation—that is, any comparisons do not involve an experimental design or a strong quasi-experimental design. In this context, the purpose of presenting outcomes for nonparticipants is to provide context, or a benchmark, to the level of outcomes realized by participants.

An outcomes study is appropriate for this evaluation in part because an experimental design is not feasible at this point in time, and a sufficiently rigorous quasi-experimental design is also not possible. That is, IES believes that without the possibility of a more rigorous design, an outcomes study is the next best option. The key outcomes for this study are student achievement and teacher retention.

The outcomes study in this case is also an appropriate design because the program being studied is relatively new and relatively small. TRPs have not been around for very long—the model was established in 2001 and until ED issued grants to 19 programs in fall 2009 and 9 more in spring 2010, only about 10 TRPs were in operation, and several of these had just started earlier in 2009. TRPs are training an average of just over 20 resident teachers this school year. Any new program may take a few years to smooth out key operations and hit their stride, so the results from a rigorous study of impacts at this point in time might not fairly reflect the results these programs could achieve a few years in the future. An outcomes study, though, is a useful preliminary step in collecting information about this program and their potential influences on student outcomes, especially given the federal investment in supporting TRPs and the great importance of continuing to broaden our knowledge of various approaches to training new teachers.

The planned outcomes study will tell us about (1) the average value-added associated with a set of teachers who chose to pursue their initial certification through a selected set of TRPs, and (2) the average retention rates of those same teachers. TRP teachers' value-added and retention rates will be benchmarked against the value-added and retention rates of non TRP teachers who (in the case of the value-added analysis) teach similar subjects, in the same districts, and who have similar levels of teaching experience. The outcomes study will not tell us about the “effectiveness” of TRP teachers, but it will provide a starting point for understanding these programs, which represent a new and growing approach to training teachers—an approach of potential interest to many policymakers and educators, and one on which almost no empirical information exists.

**6. Please provide more statistical justification about why 15 and 8 (6+2) are the right number of districts for the various stages of the study.**

To expand on the answer to question 1 above, these targets reflect subjective judgments, based in part on prior experience conducting similar studies; they are not based on considerations of statistical generalizability. We have tried to clearly and consistently describe this study as involving a purposefully selected set of programs, not a statistically representative set. From the beginning, when an impact study was proposed, it was clear that probably only a subset of grantees would meet the criteria necessary to participate in that component of the evaluation—e.g., programs that would be training teachers for whom it would be possible to collect student existing test scores. And having gathered preliminary information on the number of eligible teachers being trained by each program, we continue to believe that the right set of about 15 programs would yield useful information about program implementation, and the right set of about 8 programs would yield useful information about student outcomes associated with eligible teachers. For planning purposes, we have assumed that the 8 programs included in the outcomes analysis would allow us to conduct value added analysis for 150 TRP teachers.

The estimated targets of 15 and 8 programs also reflect the reality of resource constraints. While, theoretically, even a program training just one teacher in a commonly tested grade and subject could be included in the detailed implementation study and the outcomes study, we felt this would not be an efficient use of research funds, because recruitment costs, and even data collection costs to some extent, are similar no matter the program's size. Thus, we have prioritized targeting of programs based primarily on the number of eligible teachers they are training, with larger programs favored over smaller ones. Purposeful sampling such as this has been a common approach on major evaluations sponsored by ED over the past several years.

**7. Please justify why you propose a census of all teachers in the 8 districts. Also, please strike language in Part B that suggests that the justification is to make this convenience sample “as generalizable as possible,” since that isn’t a statically defensible statement.**

To be clear, we do not propose conducting a census of all teachers for the teacher of record survey. Instead, we propose to survey all *novice* teachers in districts participating in the outcomes study. We do not have definitive information on the number of novice teachers in participating districts (and, in fact, do not know which districts will ultimately participate in the outcomes study), but have assumed that across the 8 districts we will include 150 novice TRP teachers and 650 novice non-TRP teachers. The survey will have two main purposes. First, although we will obtain information from districts on teachers hired within the last two years (thus, potentially meeting our definition

of novice), we will not know whether these teachers have experience teaching prior to being hired by the district. The survey will be used to obtain that information and screen out teachers who are not in their first or second year of teaching overall (even if they are in their first or second year teaching in the district). Second, we will obtain information on the teachers' experiences within and outside the classroom so that we can contrast the experiences of novice TRP teachers with those of other novice teachers.

We will strike the language in Part B related to making the sample "as generalizable as possible."

**8. Related, please also justify why you require records of all students in the district (in tested grades).**

We will request records for all students in districts participating in the outcomes study (not just the students of novice teachers) due to the way that value added calculations are conducted. The value added calculation for a given teacher is a relative measure. It measures the gains of that teacher's students relative to what one would expect the gains to be given the characteristics and prior test scores of the students. And the expectations for the gains of a given set of students are determined based on other all students in the district with similar characteristics. Variants of this value added approach have been used by a number of prominent researchers (e.g., Meyer 1997; Sanders 2000; McCaffrey et al. 2004). So while the focus of the outcomes study will be on the value added scores of TRP teachers (and novice non-TRP teachers, as a benchmark), data on test scores and characteristics of all students in the district will be required to calculate the value added scores for these teachers.

Meyer, Robert H. "Value-Added Indicators of School Performance: A Primer." *Economics of Education Review*, vol. 16, no. 3, 1997, pp. 283-301.

Sanders, William L. "Value-Added Assessment from Student Achievement Data—Opportunities and Hurdles." *Journal of Personnel Evaluation in Education*, vol. 14, no. 4, 2000, pp. 329-339.

McCaffrey, Daniel F., J.R. Lockwood, Daniel Koretz, Thomas A. Louis, and Laura Hamilton. "Models for Value-Added Modeling of Teacher Effects." *Journal of Educational and Behavioral Statistics*, vol.29, no. 1, 2004, pp. 67-102.

**9. Also, according to the record specs, the student data will include a lot of PII, although some direct PII is not being requested. This seems at odds with the response to questions document.**

We do not need or have plans to collect direct student PII, such as student name and contact information. In addition, we have removed the date of birth item from the student data request documents. The student information that the study plans to collect through school records is the minimum necessary to address research question three. Of particular importance are items that we will use in the value-added model to get a more precise measure of growth in student achievement levels. In the value-added model:

$$Y_{ijt} = \alpha + Y'_{i(-t)}\delta + X'_{ij}\beta + \sum \gamma_j T_j + \mu_{jt} + \varepsilon_{ijt}$$

$Y_{ijk}$  is the test score of student  $i$  in a class taught by teacher  $j$  in year  $t$ ,  $Y_{i(-t)}$  is a vector of the previous two years of test scores for student  $i$ ,  $X_{ijk}$  is a vector of student baseline characteristics, the  $T_j$ 's are indicator variables for each teacher  $j$ ,  $\mu_j$  is a classroom-specific random error term,  $\varepsilon_{ij}$  is a student-level random error term, and  $\beta$ ,  $\delta$ , and  $\gamma$  represent parameters to be estimated. The variables included as part of the study's school records request include characteristics that will be included in  $X$ , the vector of student baseline characteristics. The purpose of this information is to measure the value-added scores of TRP teachers in a way that accounts for the characteristics of students in their classrooms.

**10. Please clarify and provide estimates of the number of teachers overlapping in each of the surveys. For example, you indicate that the teacher of record survey includes both teachers from the resident survey and those not. It also appears that some (all?) of the teachers of record are intended to be part of the teacher mobility surveys.**

For non-TRP teachers, the teacher of record survey will be the first survey in which they will participate. For TRP teachers, there will be an overlap of 75 - 100 who will do the resident survey and the teacher of record survey. This overlap is not complete because of exclusions from the teacher of record survey of TRP residents who complete the resident survey but who end up teaching grades or subjects with no test scores—and thus do not complete the teacher of record survey--and respondents to the resident survey from the 7 programs in the in-depth study sample of 15 programs but not among the 8 programs in the outcomes study. On the other hand, the teachers in the teacher of record survey sample will include those in their first and second years of teaching (as of the 2011-2012 school year), and only those in their first year of teaching will have been residents in the 2010-2011 school year when the resident survey was conducted. In other words, the second year teachers will complete the teacher of record survey but will not have completed the resident survey.

See the response to question 3. The extent of the overlap in the mobility survey will depend on whether we end up surveying only those who left the district. If we do, we expect that the sample for the mobility survey will be a subset of about 10 to 20 percent of the sample for the teacher of record survey. Otherwise, the sample for the mobility survey will include all teachers who completed the teacher of record survey other than those who screened out of that survey because they had prior teaching experience before being hired by the district.

**11. Please use the standard confidentiality pledge in A10 and all questionnaires. In particular, please do not use the term “confidentiality” rather the phrase “will not disclose to anyone outside the research team in identifiable form....”**

We will use the confidentiality statement in A10 (below) on all questionnaires and other data requests.

Per the policies and procedures required by the Education Sciences Reform Act of 2002, Title I, Part E, Section 183, responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law. Any willful disclosure of such information for nonstatistical purposes, without the informed consent of the respondent, is a class E felony.

---

**MATHEMATICA**  
Policy Research, Inc.

---

**www.mathematica-  
mpr.com**

---

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research