Part B: Supporting Statement for Paperwork Reduction Act Submission

Study of Teacher Preparation Experiences and Early Teacher Effectiveness Phase II–Data Collection

July 18, 2014

<div style="background-color:red; color:white">

Part B: Supporting Statement for Paperwork Reduction Act Submission

</div>

Table of Contents

Table of Exhibits

Appendices

Appendix 1:   District Administrative Records Collection Protocol
Appendix 2:   Teacher Survey

B. Collection of Information Employing Statistical Methods

Introduction

The U.S. Department of Education (ED) is conducting a study examining the relationship between teacher preparation experiences and early teacher effectiveness (The Study of Teacher Preparation Experiences and Early Teacher Effectiveness, formerly known as The Study of Promising Features of Teacher Preparation Programs). This Information Collection Request (ICR) is the second of two ICRs for the study. The first ICR (*Phase I—Recruitment*; OMB-PRA: 1850-0891) requested clearance for recruitment activities. This second ICR, *Phase II—Data Collection*, requests clearance for data collection activities (obtaining teacher contact information from districts, collecting data from teachers on preparation experiences via an online teacher survey, and obtaining student data from districts). The packages have been submitted separately because the process of recruiting districts had to begin before the teacher survey was developed and piloted.[1]

This study is sponsored by ED's Institute of Education Sciences (IES) and is being implemented by Abt Associates Inc. and its partners: Belmont Research Associates, The Bench Group, Dillon-Goodson Research Associates, Education Analytics (EA), and Pemberton Research (together, the "study team").

Overview of the Study

Title II, Part A of the Elementary and Secondary Education Act—the Improving Teacher Quality State Grants program—focuses on improving teacher quality and increasing the number of highly qualified teachers. This study, which is the first of its kind, is about how to improve teacher quality through more effective preparation of new teachers. It will look at how the nature and intensity of preparation experiences are related to teacher effectiveness in promoting student achievement. For the purposes of this study, "teacher effectiveness" is defined as teacher value-added (TVA). The study will measure TVA using analytic techniques that isolate teacher contributions to student test score gains. The study has three primary questions:

1) What are the relationships between teacher preparation experiences and teacher effectiveness in the first year of teaching, measured by teacher value-added? (RQ1)

2) What are the relationships between teacher preparation experiences and teacher effectiveness with English learners in the first year of teaching, measured by teacher value-added for English learners? (RQ2)

3) Do relationships between teachers' preparation experiences and teacher effectiveness in the first year of teaching differ depending on teachers' assessments of the usefulness of the preparation experiences? (RQ3)

Measuring Preparation Program Experiences

To answer the research questions, the study needs to identify preparation experiences that could reasonably be hypothesized to affect student test scores. The challenge is that little or no research has examined relationships between preparation experiences and test scores. (The 2010 National Research

---

[1]   At the time of the first ICR, the study design called for an experiment; however, efforts to recruit teacher pairs which met the experiment's requirements were unsuccessful, and therefore a value-added approach is now the focus of the study.

Council report on teacher preparation highlighted this lack of research.) However, the recent Measures of Effective Teaching (MET) Study (Kane & Staiger, 2012) did identify a set of teacher practices related to test score gains. The study team used the MET study as a framework for identifying preparation experiences that would foster these practices. The survey asks teachers to rate the frequency of these experiences in their teacher preparation program.

Specifically, in the MET study, teachers were rated on their instructional practices using four different observation systems. This process resulted in each teacher being rated on practices in 32 domains of instruction. The study team distilled these 32 domains into 12 broad topic areas, which form the basis for the teacher survey questions about preparation experiences. This reduction was based primarily on data from the MET study on (a) the distribution of teacher ratings on each practice, (b) the relationship of teacher ratings on each practice to teacher value-added scores, and (c) correlations of the practices across the four measures. For each of the 12 topic areas, the study team developed a small set of specific instructional strategies from the coding manuals for the four MET measures. The strategies align with teaching behaviors that coders are trained to look for as evidence of a teacher's skill with the measure-specific practices. Finally, the study team conducted cognitive interviews with teachers to ensure that the instructional strategies were being interpreted as intended.

The same kind of evidence provided by the MET study to inform the identification of instructional practices for the general classroom does not exist for English learner-specific practices. Some evidence does exist, however, at the individual strategy level. The study used four research syntheses and the 2007 and 2014 IES Practice Guides on instruction for English learners to identify English learner-specific instructional strategies that are supported by evidence of improving English learner achievement (August & Shanahan, 2006; Baker et al., 2014; Calderón, Slavin & Sanchez, 2011; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Genesee, Lindholm-Leary, Saunders & Christian, 2006; Gersten, Baker, Shanahan, Linan-Thompson, Collins & Scarcella, 2007). A set of English learner-specific strategies were selected by 1) eliminating redundancies across strategies that were identified in multiple reviews, 2) consulting content experts to guide and prioritize the selection of strategies, and 3) conducting cognitive interviews with teachers to ensure that the strategies were being interpreted as intended. The final set of English learner-specific instructional strategies were grouped together into one English learner-specific topic area on the survey.

Sample

An analysis of statistical power indicated that a sample of 6,450 teachers is needed to answer the primary research questions with reasonable precision. To meet this sample size, the study aims to include as many teachers as possible in recruited districts who teach reading/English language arts (ELA) and/or mathematics, the two subject areas consistently tested across states. Because analyses will be conducted separately by subject, focusing on elementary-grade teachers increases the likelihood that each teacher contributes to both the reading/ELA and mathematics sample, which reduces the cost of recruiting districts and teachers. Fourth grade is the lower bound for the teacher sample due to data availability— proficiency testing typically begins in third grade, which means fourth grade is the first grade for which two years of test scores needed to estimate TVA are available (i.e., test scores from grades 3 and 4). The study selected sixth grade as the upper bound because it is assumed to be the highest elementary grade.[2]

---

[2]   In some school districts, some or all of the sixth grade classrooms will be in middle schools. However, data from the Common Core indicates that a substantial proportion of sixth grade classrooms will be in elementary schools.

Limiting the study to first-year teachers would minimize issues with recalling preparation experiences. However, data on teachers entering the teaching force indicated that meeting the sample size requirement would require two years of data collection (i.e., recruiting teachers in 2014-15 and again in 2015-16). To limit data collection to one year, the study decided to include teachers in their second or third years of teaching for whom TVA in their first year of teaching could be estimated. The three-year upper bound is driven by concerns about recalling preparation program experiences for teachers who have been teaching for more than three years.

Teachers in their first through third years will be asked the same questions about their teacher preparation experiences. The primary analyses will examine the relationship of teachers' preparation experiences to effectiveness in their *first year*. For example, a teacher in her second year in 2014-2015 will be asked about her preparation experiences and those experiences will be related to her value-added in the *previous* year, which corresponds to her first year of teaching.[3]

The teacher sample will be recruited from a purposive sample of approximately 50 large and moderate-sized school districts across the country that can meet the student-teacher data linkage requirements for estimating TVA. Targeted districts include the 10 largest districts in the United States (based on the expected number of eligible teachers) and other moderate-to-large districts (or local education agencies). For the majority of these districts, the study team is currently conducting research projects, has existing partnerships, and/or has access to key decision makers. The targeted districts are in 25 states and 49 percent of them include 10 percent or more English learners. Districts will be asked to provide the study team with teacher contact information for administering the (online) teacher survey on preparation experiences in spring 2015. The study team will also request student data from districts to estimate teacher value-added. The data will include student scores on state assessments (reading/ELA and mathematics) and student demographic data.

Analysis and reporting

Hierarchical analysis (students nested within classes nested within teachers) will be used to estimate the relationship between preparation program experiences and student test scores, controlling for student, class, teacher, and school baseline covariates. Reading/ELA and Mathematics scores will be analyzed separately. For the subset of teachers with five or more English learners, the study will examine the relationship between general preparation experiences and the test scores of English learners as well as English learner-specific preparation experiences and the test scores of English learners. The study team will produce a study report that is expected to be available by fall 2017.

Phase II—Data collection request

This ICR seeks clearance to obtain the following data:

- From school districts:
    - Teacher contact information (emails, phone numbers) for first-, second-, and third-year teachers in the district in the 2014–15 school year;

---

[3] If second- or third-year teachers have not taught in an eligible grade or subject in their first year but have in their second and/or third year of teaching, they will still be surveyed and included in secondary analyses examining the relationship of teacher preparation experiences to teacher effectiveness in the second and/or third years of teaching. Only second- or third-year teachers who meet the eligibility criteria in their first year of teaching will be included in the sample used to answer the primary research questions.

- Student data:
  - Reading/ELA and mathematics state assessment data for students in grades three to six for four years (2011–12 through 2014–15); [4]
  - Student demographic data (English learner status, Special Education status, free/reduced price lunch status, gender, race/ethnicity) for students in each year of test data.
- From teachers:
  - Data to verify sample eligibility, to determine teacher preparation pathway, to assess the frequency and usefulness of preparation experiences, and to measure background characteristics.

This ICR provides a detailed discussion of the study data collections and the analysis and reporting of the data, as well as an overview of the study, including its design and data collection procedures. Copies of the data collection instruments (the teacher survey and the District Administrative Records Collection Protocol) are included in the appendices.

B.1   Respondent Universe and Sampling Methods

B.1.1   Respondent Universe

The study team will construct a sample of teachers in recruited school districts that meet the following eligibility criteria for the primary analyses—eligible teachers must be 1) in their first, second, or third year of teaching; 2) responsible for teaching reading/ELA or mathematics during their first year of teaching to at least one general education classroom of fourth- through sixth-grade students.[5] All teachers, regardless of pathway (e.g., traditional, alternative, Teach for America), will be eligible.

B.1.2   Sampling Methods

The study will recruit approximately 50 districts. Targeted districts include the 10 largest districts in the United States (based on the expected number of eligible teachers) and other moderate-to-large districts. For the majority of these districts, the study team is currently conducting research projects, has existing partnerships, and/or has access to key decision makers. The targeted districts are in 25 states and 49 percent of them include 10 percent or more English learners.

Given the large sample size needed to answer the primary research questions (n = 6,450), the study team intends to construct a purposive sample of teachers that includes all eligible teachers in districts that agree to participate in the study. Districts will be asked to provide contact information for all first-, second-, and third-year teachers in the district in the 2014-15 school year.[6] The study team will use the student-teacher linkages that are part of the student assessment data to further narrow the sample to first-, second-, and

---

[4]   For districts that administer state tests in the fall of each year, data will be obtained for students in grades four to seven for SY2012–13 through SY2015–16.

[5]   Second- and third-year teachers who have not taught in an eligible grade or subject in their first year but have in their second and/or third year of teaching will be surveyed and included in secondary analyses examining the relationship of preparation experiences to teacher effectiveness in the second or third year of teaching.

[6]   Since districts may only count years of teaching in their district when reporting teachers' years of experience, the survey will ask teachers to confirm that they are in the first, second or third year of their teaching career (not limited to their tenure in their current district).

third-year teachers who meet the grade and subject criteria listed above. This set of teachers will be sent the teacher survey and their eligibility will be confirmed via questions at the beginning of the teacher survey. While the anticipated study sample will not have formal representativeness in the strict statistical sense, it will span a large number of districts and states across the nation.

B.1.3   Expected Response Rates

The study plans for a 100 percent response rate for teacher contact information and student administrative data since participating districts will have agreed to provide these data. The study plans for an 80 percent response rate for the teacher survey. The high rate is based on 1) districts being asked to encourage teachers to participate in the survey; 2) the use of a multi-mode follow-up strategy that involves social media, email, phone calls, and opportunities for teachers to complete the survey on the phone or in hard copy; and 3) offering a modest incentive to complete the survey.

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

B.2.1   Sample Selection

Please see section B.1.2 above for a detailed description of the study's sampling strategy.

B.2.2   Estimation Procedures

This section presents the study's estimation approach for addressing the three primary research questions.

1. **What are the relationships between teacher preparation experiences and teacher effectiveness in the first year of teaching, measured by teacher value-added?**

To address the first research question, the study team will analyze relationships between teacher preparation program experiences (TPPEs) and teacher effectiveness, measured by value-added. TPPEs are defined as the frequency of specific types of preparation experiences related to the representative instructional strategies within each key instructional topic areas. For each topic area, four analytic variables representing the four types of preparation experiences will be created: 1) opportunities to "read about, hear about, or see a role play of the strategies, such as during coursework;" 2) opportunities to "observe a teacher using the strategies in a K-12 classroom (in videos or during fieldwork or student teaching);" 3) opportunities to "practice the strategies in a K-12 classroom *prior* to becoming a full-time teacher;" and 4) opportunities to "receive feedback from program staff or a cooperating teacher on your use of this strategy that included what you did well or how you could improve." These TPPE measures will be used as explanatory variables in the analytic models.

Overview of Approach

The relationship of TPPEs will be examined separately for reading/ELA and mathematics but the analysis approach will be identical. For both outcomes, the study team will combine data from different school districts and grade levels to estimate relationships between TPPEs and value-added. May et al. (2009) argue that the decision of whether to combine data across grades or states should be driven primarily by a study's research questions. Combining scores across grades or states is appropriate when the study's questions are about student performance on state tests. Combining scores is inappropriate when the study's questions are about attainment of specific skills. Because the current study's primary research questions fall into the first category—relating TPPEs to value-added, as measured by student performance on state tests—combining data across states and grades is appropriate.

The study will use a two-stage approach to analyze the relationship between preparation experiences and value-added. The first stage will produce TVA scores for each teacher. Each TVA score is a measure of the extent to which a teacher's students experienced achievement growth over a school year, adjusting for student characteristics. In this stage, achievement scores will first be standardized using state and year means and standard deviations. The study then will fit models using data from each unique combination of state, grade, and year, which will yield a set of state/grade/year-specific TVA scores. In participating districts, test scores of all students in relevant grades will be used in the stage 1 analysis. The use of all student data in the relevant grades will increase the precision of the estimated coefficients and resulting TVA score estimates.

In stage 2, the set of state/grade/year-specific TVA scores will be combined into a single data set. Using the combined data set, regression models will be estimated with the TVA score as the dependent variable and one or more TPPE measures as independent variables, as well as grade and state dummies, and other statistical controls. Only the TVA scores of teachers who have completed the survey on preparation experiences will be used in the second stage analysis. Each of the two stages is described in more detail below.

Analytic Model for Stage 1

The stage 1 analysis will produce a TVA estimate for each teacher. The approach will utilize a hierarchical linear modeling (HLM) framework, which is common in education research. The analytic model is a three-level hierarchical linear model with students (level-1) nested in classes (level-2), and classes nested in teachers (level-3).[7] The level-1 model has student prior year scores and other student characteristics as covariates. The level-2 model includes average student measures 'centered' at the teacher-level mean (the overall mean at the teacher level will be subtracted from the classroom mean). The level-3 model includes these covariates aggregated to the teacher-level. In the level-3 model, the dependent variable is the conditional TVA and the level-3 residual is the part of the TVA that has not been explained by covariates. The level-3 residuals from the stage 1 analyses are used as the dependent variable in the stage 2 analysis. Note that the model does not include dummies ("fixed effects") for schools.[8]

Specifically, the level-1 model, or student-level model is:

$$Y_{ijk} = \alpha_{0jk} + \sum_{m=1}^{M} \alpha_{mjk}(X_{mijk} - \overline{X}_{m.jk}) + e_{ijk}$$

The level-2 model or class-level model is:

$$\alpha_{0jk} = \beta_{00k} + \sum_{m=1}^{M} \beta_{0mk}(\overline{X}_{m.jk} - \overline{X}_{m..k}) + u_{jk}$$

---

[7]   The study assumes that more than half of fifth- and sixth-grade teachers will teach more than one class in a single school year.

[8]   Inclusion of fixed effects for schools limits the analysis sample to teachers from schools where there is more than one eligible teacher in the school and where two or more teachers have different values on the TPPE measures of interest. This approach could result in a very small analysis sample that would be a poor representation of the full sample of teachers that will be surveyed.

The level-3 model or teacher-level model is:

$$\beta_{00k} = \gamma_{000} + \sum_{m=1}^{M} \gamma_{00m}(\overline{X}_{m..k}) + \sum_{t=1}^{T} \psi_{00t}(Z_{tk}) + r_k$$

where, in the level-1 model,

Y_{ijk} is the spring reading/ELA or mathematics achievement test score from i^{th} student (i in 1,2,...,n) in the jth class, (j in 1,2,..., J), nested in the kth teacher (k in 1,2,...,K teachers).

$(X_{mijk} - \overline{X}_{m.jk})$ is the mth of M student characteristics (e.g., prior year test score, gender, race/ethnicity age, English learner status, free/reduced price lunch status), centered at the class-level mean.

e_{ijk} is the student-level error, assumed distributed normal with mean zero and variance σ_e^2

α_{0jk} is the covariate adjusted mean of achievement scores for the jth class of the kth teacher.

In the level-2 model,

$(\overline{X}_{m.jk} - \overline{X}_{m..k})$ are the student characteristics aggregated up to class-level means, and centered around the teacher's mean.

u_{jk} is the class-level error, assumed distributed normal with mean zero and variance σ_u^2

In the level-3 model,

β_{00k} is the conditional *mean* of achievement scores for the kth teacher (i.e., the conditional TVA score)

γ_{000} is the conditional grand mean (predicted mean for teachers when all teacher-level covariates are zero).

$(\overline{X}_{m..k})$ are the means of student characteristics for the kth teacher, averaged over all of the classes taught by the kth teacher.

(Z_{tk}) are measures of school characteristics (e.g., percent LEP, percent FRPL), and dummies for district.

r_k is the teacher-level error, assumed distributed normal with mean zero and variance σ_k^2

As mentioned above, this model will be fit separately for each unique combination of state, grade, and school year. The fitted models will produce an estimated TVA "score" (the level-3 residual r_k) and its standard error.

Analytic Model for Stage 2

In stage 2, a data set will be created that merges teachers' TPPE measures and their TVA scores from stage 1. Regression models will be used to estimate the relationships between TPPEs and TVA scores. These models will use the TVA as the dependent variable and TPPE measures; measures of perseverance; leadership; and prior achievement (ACT/SAT scores), which are collected via the survey to account for selection; and indicator variables for state, grade, and teachers' years of experience at the time of the survey as explanatory variables.[9] The outcome variable in the model (TVA score) will be weighted inversely proportional to the square of the standard error of the TVA score, which means teachers with more precisely estimated TVA scores will receive more weight in the estimation.

2. **What are the relationships between teacher preparation experiences and effectiveness with English learners in the first year of teaching, measured by teacher value-added for English learners?**

To address the second research question about relationships of TPPEs to value-added for English learners, the models described in the previous section will be fit to subsets of data consisting of student outcomes for English learners, and teachers of classes that have a minimum of five English learners. Specifically, in the stage 1 models that yield TVA estimates, all of the student-level (level-1) outcomes and covariate data will be specific to only English learners. For the class-level (level-2) and teacher-level (level-3) covariates that represent aggregates of student characteristics, the data from all students (English learner and non-English learner) will be aggregated such that these measures will represent the characteristics of all of the students in a teacher's class or classes. These models will yield state, grade, and year-specific TVA estimates that pertain to only English learners. Only teachers with at least five English learners and their English learner-specific TVA estimates will be used in the stage 2 models. Measures of TPPEs will include general measures (i.e., the same measures as will be assessed in the primary research question) and TPPEs that are ELL-specific.

3. **Do relationships between teachers' preparation experiences and effectiveness in the first year of teaching differ depending on teacher assessments of the usefulness of the experiences?**

In addition to rating the frequency of specific types of preparation experiences related to key instructional topic areas, teachers are also asked on the survey to rate the usefulness of their TPPEs.

To answer this research question, descriptive data will be tabulated from the usefulness items to describe teachers' perceptions of the usefulness of TPPEs for first, second and third year teachers separately, and for all teachers combined. The stage 2 analytic models described above will also be estimated, but in each model, the TPPE measure of interest will be replaced with a dichotomous variable for usefulness: 1 = useful (if TPPE is rated as useful or very useful), 0 = not useful (if TPPE rated as not useful or a little

---

9   Some teachers may have multiple TVA records. If there are more than a negligible number of teachers with multiple records, the model will be fit as a two-level hierarchical model with repeated observations at level-1 and teachers at level-2.

useful). The coefficient for the usefulness variable will represent the relationship between the perceived usefulness of a TPPE and TVA. That is, it will address the question of whether TPPEs that are rated as useful are more strongly associated with teacher effectiveness than TPPEs that are rated as not useful.

A secondary research question about the usefulness of TPPEs is whether teachers' assessment of the usefulness of their experiences changes over time. A descriptive analysis will compare the proportions of teachers who describe a TPPE as useful among the first-, second-, and third-year teacher samples.

B.2.3 Degree of Accuracy Needed

For power calculations, the study team used a simplified version of the two-stage analytic approach described above. This simplified version includes key concepts of the two-stage approach but omits complications from combining data across states and grade levels.[10]

To conceptualize the power and sample size calculations, the study team focused on the estimate of the effect of a *single* TPPE, controlling for other TPPEs. For clarity and notational convenience, this variable is referred to as the *TPPE measure*.

A review of the research on the relationship between TPPEs and teacher effectiveness (Price et al., 2013) indicated that effects of TPPEs on TVA are likely to be small. For power calculations, the study team operationalized a "small" effect to be 0.05 standard deviation units for the effects of a binary measure of a TPPE (the measure takes the value "1" if the teacher had the experience and "0" if not). To calculate the sample size needed to detect an effect of 0.05, the study has made assumptions about 1) the numbers of mathematics and reading/ELA classes taught by teachers in a school year; 2) the distribution of the *TPPE measure*; and 3) hypothesis tests, variance components, explained variance, numbers of classes per teacher, and number of students per class.

For the first research question, the results of these calculations indicate that to detect small effects (a minimal detectable effect (MDE) of 0.05 standard deviation units) of TPPEs on TVA for mathematics outcomes, the study needs data for 4,500 teachers. Similarly, to detect small effects of TPPEs on TVA for reading/ELA, the study needs data for 4,500 reading/ELA teachers. Because some teachers will teach both mathematics and reading/ELA while some will not, a total sample of 6,450 teachers is needed.

For the second research question concerning the relationships of TPPEs to TVA for teachers of English learners, the study is estimated to have data for 3,000 teachers of mathematics and 3,000 teacher of reading/ELA—two-thirds of the full sample of teachers who will have at least five English learners in their classes. Because some teachers will teach both mathematics and reading/ELA while some will not, the total sample will consist of 4,300 teachers with at least five English learners in their classrooms. This sample will yield an MDE of 0.062 standard deviation units.

> *Assumptions about numbers of reading and mathematics classes taught by each teacher*

The study will focus on fourth-, fifth- and sixth-grade teachers. Separate models will be estimated for TVA in reading/ELA and mathematics. Sample size calculations assume that one-third of completed surveys will come from teachers in each of the three grades. Because some fifth- and sixth-grade classes are departmentalized, with a teacher teaching only reading/ELA or mathematics, the sample size calculations account for the fact that not all teachers will contribute scores to both reading/ELA and

---

[10] This simplification entails combining the stage 1 and stage 2 models described above into a simplified three-level model that essentially adds the TPPE measures and selection controls to the stage 1 model. The resulting combined model is assumed to be estimated with data from a single state and grade level.

mathematics. Exhibit B-2 presents the assumptions used in sample size calculations concerning the proportions of fourth-, fifth-, and sixth-grade teachers who teach 1) a single class with both subjects (mathematics and reading/ELA); 2) two mathematics classes but zero reading/ELA classes; or 3) teach two reading/ELA classes but zero mathematics classes. Based on the assumptions presented in the exhibit, on average, mathematics teachers will be responsible for 1.4 mathematics classes and reading/ELA teachers will be responsible for 1.4 reading/ELA classes.

Exhibit B-2. Assumptions About Numbers of Mathematics and Reading/ELA Classes Taught

	Teach one class of reading/ELA and one class of mathematics	Teach two classes of reading/ELA and no classes of mathematics	Teach two classes of mathematics and no classes of reading/ELA	Total
Fourth	100%	0%	0%	-
Fifth	50%	25%	25%	-
Sixth	20%	40%	40%	-
Total # of teachers	2,550	1,950	1,950	6,450
Reading/ELA sample size	2,550	1,950	-	4,500
Mathematics sample size	2,550	-	1,950	4,500

Assumptions about the distribution of the TPPE measure

The power analysis for a *TPPE measure* depends on the distribution of that variable. For example, for a dichotomously coded (yes or no) *TPPE measure,* power is greatest (required sample size is smallest) if the distribution of the *TPPE measure* is such that 50 percent of the sample has a TPPE value of 1, and 50 percent of the sample has a value of 0. The study team developed the sample size requirements based on a more plausible scenario in which the *TPPE measure* had values of 20 percent "yes" and 80 percent "no."

Assumptions about hypothesis tests, variance components, r-squares, classes per teacher, and students per class

The study uses standard statistical assumptions regarding conducting two-tailed hypothesis tests with alpha-level criterion $p < 0.05$ with 80 percent power. The study also makes assumptions about 1) values of the variance components; 2) the proportion of variance explained by covariates; 3) the average number of students per class; 4) the average number of classes per teacher; and 5) the correlation of covariates with the *TPPE measure,* which is termed the *Rsq.TPPE Dependent Measure*.[11]

For assumptions regarding the first four parameters listed above, the study team used ranges of plausible values from the literature and previous educational evaluations (Exhibit B-3). For the fifth parameter, previous literature did not provide guidance so the study team selected a range that was reasonable. Within the ranges of plausible values for these five parameters, there are an infinite number of scenarios, based on varying sets of assumptions. A typical power analysis would choose several exemplary scenarios and show sample sizes or MDEs for each of those scenarios. The study team adopted a more flexible approach. A plausible range of assumptions for each of the parameters was specified and a Monte Carlo methodology was used to make a random draw from a uniform distribution within the specified range for each parameter. For each of 5,000 Monte Carlo draws, the MDE that resulted for the draw was calculated.[12] The Monte Carlo procedure yields MDE estimates for a wide range of possible scenarios. All

---

[11] This label can be read as "R-square when the TPPE measure is regressed on the other covariates."

[12] In each of the 5,000 Monte Carlo draws, an MDE was calculated using a modified form of Equation 13 from Schochet (2008). The modification was the addition of a term in the denominator that inflates the variance of the

scenarios assumed a two-sided hypothesis test with alpha = 0.05 criterion, 80 percent power, and a three-level HLM model of the form described in the previous section. Exhibit B-3 summarizes the assumed plausible ranges of parameters used for the Monte Carlo simulations.

Exhibit B-3. Assumed Plausible Ranges for Variance Components, R-Squares, Classes per Teacher, and Students per Class

Parameter	Min	Max
Level 1 Variance	0.50	0.80
Level 2 Variance	0.00	0.30
Level 3 Variance	0.08	0.50
Total Variance	1	1
Level 1 R-square	0.50	0.75
Level 2 R-square	0.10	0.40
Level 3 R-square	0.40	0.70
Rsq.TPPE Dependent Measure	0.10	0.50
Classes Per Teacher	1.43	1.43
Students Per Class (RQ1)	18	25
Students Per Class (RQ2-ELLs)	5	15

Estimated MDEs

For RQ1, MDE estimates from the Monte Carlo simulations are summarized in Exhibit B-4. Across the 5,000 simulations, the mean of the MDE was 0.05 and ranged from around 0.031 to 0.078. These results indicate that with data from 4,500 mathematics teachers, the expected MDE is between 0.031 and 0.078. Averaged over all assumption sets, the MDE is 0.050. The same estimates apply to data from 4,500 reading/ELA teachers.

For RQ2, additional assumptions must be made about the proportion of teachers from the main sample that will have classes with five or more English learners. For these calculations, the study assumed that two-thirds of the teachers in the main sample will have classes with five or more English learners (the study's districts are large and likely to be urban, which supports this assumption). Across the 5,000 simulations, the mean of the MDE for English learners was 0.062 and ranged from 0.038 to 0.096.

---

treatment effect as a function of the correlation of covariates with the *TPPE measure* (i.e., as a function of the term *Rsq.TPPE Dependent Measure*).

Exhibit B-4. MDEs for Various Scenarios—RQ1

# Teachers w/ complete data	# Teachers used in analysis	Distribution of key variable	Other assumptions	MDE Mean (min, max)
2,550 ELA & mathematics 1,950 mathematics only 1,950 ELA only 6,450 total	2,550 mathematics and ELA 1,950 mathematics or ELA 4,500 total	Binary TPPE measure w/ 80%-20% split: 900 with *TPPE measure* = 1 (20%) 3,600 with *TPPE measure* = 0 (80%)	See Exhibit B-3	.050, (.031, .078)

Exhibit B-5. MDEs for Various Scenarios—RQ2

# Teachers w/ 5+ English learners and complete data	# Teachers used in analysis	Distribution of key variable	Other assumptions	MDE Mean (min, max)
1,700 ELA & mathematics 1,300 mathematics only 1,300 ELA only 4,300 total (2/3 of teachers w/ complete data)	1,700 mathematics and ELA 1,300 mathematics or ELA 3,000 total	Binary *TPPE measure* w/ 80%-20% split: 600 with *TPPE measure* = 1 (20%) 2,400 with *TPPE measure* = 0 (80%)	See Exhibit B-3	.062 (.038, .096)

B.2.4   Unusual Problems Requiring Specialized Sampling Procedures

No unusual problems that require specialized sampling procedures are anticipated.

B.2.5   Use of Periodic Data Collection Cycles to Reduce Burden

This is a one-time data collection effort.

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

This section describes the strategies and methods that will be used to maximize response rates for the teacher survey.

To obtain responses from the study teacher sample, the study team has developed strategies to facilitate communication with respondents during data collection activities and to maximize response rates. These strategies have proven successful in the study team's extensive experience conducting large-scale evaluation studies (e.g., The Reading First Impact Study, Evaluation of the U.S. Department of Education's Student Mentoring Program, Evaluation of the Massachusetts Expanded Learning Time Initiative, The Enhanced Reading Opportunities Study, The Career Academies Evaluation, and The Teacher Incentive Fund Evaluation).

B.3.1   Teacher Survey

The study team will administer online surveys to all eligible teachers in recruited districts. Teachers will be provided with an individualized link to their survey by email. The online survey technology is user-friendly, especially because new teachers are likely to have used the web extensively, and also minimizes the burden that is associated with returning paper surveys by mail.

Teachers may not want to take the time necessary to complete the survey. Alternatively, they may be willing to participate, but only if the study team makes it easy for them to do so, or offers them something in return. The study team will use four strategies to increase teacher response rates:

- *Encouragement from districts*. During recruitment, the study team will explore ways that the district might be willing to encourage teachers to complete the survey. This might include expressing their support for the study by sending out a letter of support or publicizing the study in a district newsletter or on district websites.

- *Personalized access*. Invitations to participate in the survey will be personalized and include a password embedded link to the survey.

- *Provide for multiple modes for completion*. Respondents will be able to complete the survey online, but if requested, they can receive a hard copy of the survey or complete the survey by phone.

- *Make use of reminders*. The study team will use reminder emails. Individual responses to the surveys will be tracked in the survey software, and targeted reminders will be sent every four to five days to those who have not completed the survey. The study team will also make phone calls to non-respondents after three email reminders to encourage them to complete their surveys. The call will be an opportunity to complete the survey by phone or the study team will ask teachers for additional contact information and the best time and method to allow them to complete the survey.

There will be an incentive payment to teachers. Incentives are appropriately used in federal statistical surveys with respondents whose failure to participate would jeopardize the quality of the survey data (Graham, 2006). Given the importance of obtaining a high response rate for this project, the study team intends to offer a $30 incentive to teachers who complete the survey. Based on previous studies, an 80 percent response rate is expected.

B.3.2   Extant Student Administrative Data

During recruitment, study team members will execute data release agreements with districts. Each district will be asked for four years of student data to be able to estimate value-added from 1) the first year of teaching for all eligible teachers (first-, second- and third-year teachers); 2) the second year of teaching for all eligible second- and third-year teachers; and 3) the third year of teaching for all eligible third-year teachers (see Exhibit B-6).

Exhibit B-6. Years of Student Outcome Data Needed to Calculate Teacher Value-Added

Teacher Sample	1st Year of Teaching	2nd Year of Teaching	3rd Year of Teaching
1st year teachers	Spring 2014 & 2015	NA	NA
2nd year teachers	Spring 2013 & 2014	Spring 2014 & 2015	NA
3rd year teachers	Spring 2012 & 2013	Spring 2013 & 2014	Spring 2014 & 2015

Student administrative data will be requested twice during the study period. The first data capture will be in late 2014. This data capture will request student data from spring 2012, spring 2013, and spring 2014. The second data capture will be in fall 2015, which will be a request for the final student data from spring 2015. Each data request will be for data from all students in the district in grades three through six, to allow for the estimation of teacher value-added for fourth-, fifth-, and sixth-grade teachers.[13]

B.4   Test of Procedures and Methods To Be Undertaken

B.4.1   Teacher Survey

During the 60-day comment period, the teacher survey was pilot tested. The pilot test results indicated that the survey instructions are understandable, the content is comprehensible, and the length is reasonable. The pilot test was conducted with a purposive sample of teachers who represented traditional and alternative certification preparation programs and were in their first, second or third year of teaching. These teachers were recruited from districts not being asked to participate in the study.

The teacher survey includes questions in four areas: eligibility, preparation pathway, preparation program experiences, and background characteristics (Appendix 2). In the first area, a series of questions are asked to confirm eligibility for the study. These questions ask teachers about their year of teaching (in any district) and whether they are or have been responsible for teaching reading/ELA and/or mathematics to at least one classroom of general education students in grades 4, 5 or 6 in any of the following school years: 2012-13, 2013-14 or 2014-15.

The second area includes questions on teachers' preparation pathway that come from surveys used in two previous studies:

---

[13] For districts that administer state tests in the fall of each year, data will be obtained for students in grades four through seven for SY2012–13 through SY2015–16.

- The Teacher Pathway Project (Boyd et al., 2009)—Survey of Program Graduates, Survey of First Year Teachers, Survey of Second Year Teachers;[14] and

- The Impacts of Comprehensive Teacher Induction Study (Glazerman et al., 2008).

The third area includes 13 items that ask for teachers' reports on the frequency of specific preparation experiences. The survey is a new measure developed for the current study; there were no prior surveys asking these same questions. The study team used the Measures of Effective Teaching (MET) study (Kane & Staiger, 2012) as a framework for developing 12 of the survey items. The MET study reports evidence on instructional practices observed in a large sample of teachers that are related to TVA. The study team developed one English learner-specific item based on the existing literature. The survey items ask about preparation experiences related to the development of selected general and English learner-specific instructional topic areas in new teachers.

The 13 survey items on teacher preparation experiences follow the same item format. Each item is organized around a single topic area. These topic areas were distilled from the larger set of instructional practices measured in the MET study. For each topic area, the respondent is asked about experiences related to a small set of (four to five) specific instructional strategies that are examples of the overarching topic area. The instructional strategies also were derived from the MET observation measures; they were drawn from the coding manuals for the MET observation measures, in which definitions were provided about specific instructional behaviors that observers were looking for in order to rate a teacher on a broader instructional practice. For each of the instructional strategies, respondents are asked to rate the frequency of four types of preparation experiences that teacher candidates might experience as part of their teacher preparation program: 1) opportunities to "read about, hear about, or see a role play of this strategy, such as during coursework;" 2) opportunities to "observe a teacher using this strategy in a K-12 classroom (in videos or during fieldwork or student teaching);" 3) opportunities to "practice this strategy in a K-12 classroom *prior* to becoming a full-time teacher;" or 4) opportunities to "receive feedback from program staff or a cooperating teacher on your use of this strategy that included what you did well or how you could improve." Teachers are instructed to include in their responses all preparation experiences prior to becoming a teacher of record for the first time as well as preparation experiences that might be received during their first, second or even third year in the classroom.[15]

The survey items on experiences include an additional question about teachers' perceptions of the usefulness of their preparation experiences in their classroom teaching. This question is asked for each instructional strategy.

The fourth area includes questions on demographics including age, gender, race/ethnicity and two background characteristics that will be used to control for selection bias. These two background

---

[14] Surveys retrieved from: http://cepa.stanford.edu/tpr/teacher-pathway-project#quicktabs-pathway_project=1

[15] Second- and third-year teachers who have not completed their programs at the time of the survey but who will be included in the primary analyses because they have first-year TVA will be reporting on preparation experiences that are happening after the outcome. The study team expects this to be a small subset of second- and third-year teachers and will conduct sensitivity analyses to determine if excluding these teachers from the analyses affects the overall findings.

characteristics were identified in research on Teach for America as being related to teacher effectiveness (Dobbie, 2011). The first is a measure of perseverance. The survey includes the eight-item version of the Grit Scale (Duckworth & Quinn, 2009), a validated measure of perseverance that has been used in a number of previous studies and has been shown to be related to measures of academic persistence and performance. The second is a measure of leadership skills. The survey includes two items about teachers' involvement in initiating or holding a leadership position in a club or organization during their undergraduate career. Finally, the survey asks teachers for the information that is necessary to provide to ACT Inc., or the College Board to locate teachers' ACT or SAT scores and release them to Abt Associates. For the ACT, this includes full name at time of testing, current name (if different), state in which the test was taken, and date of birth. For the SAT, this information includes full name at time of testing, current name (if different), and social security number. Teachers are assured that their scores will be used for analysis purposes only and will never be identified or shared outside of the study researchers. These scores are a proxy for prior academic skill and will be used as another covariate in the models to control for selection bias.

B.5   Individuals Consulted on Statistical Aspects of the Design

The following individuals were consulted on the statistical aspects of the study:

Name	Title/Affiliation	Telephone
Dr. Fatih Unlu	Senior Scientist, Abt Associates	617-520-2528
Dr. Mark Dynarski	Director, Pemberton Research	609-443-1981
Mr. Cristofer Price	Principal Scientist, Abt Associates	301-634-1852

The following individuals will be responsible for the data collection and analysis:

Name	Title/Affiliation	Telephone
Dr. Tamara Linkow	Associate, Abt Associates	617-520-2978
Dr. Fatih Unlu	Senior Scientist, Abt Associates	617-520-2528
Dr. Mark Dynarski	Director, Pemberton Research	609-443-1981
Mr. Cristofer Price	Principal Scientist, Abt Associates	301-634-1852
Dr. Robert Meyer	Director, Education Analytics	608-265-5663
Dr. Andrew Rice	Associate Director, Education Analytics	608-890-3789

References

August, D., & Shanahan, T. (2006). *Developing literacy in a second-language: Report of the National Literacy Panel on Language Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., Gersten, R., Haymond, K., Kieffer, M. J., Linan-Thompson, S., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://ies.ed.gov/ncee/wwc/publications_reviews.aspx.

Boyd, D.J., P.L. Grossman, H. Lankford, S. Loeb, and J. Wyckoff. (2009). Teacher preparation and teacher effectiveness. *Educational Evaluation and Policy Analysis*, 31, 416. DOI:10.3102/0162173709353129.

Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *The Future of Children, 21*(1), 103-127.

Dobbie, W. (2011). Teacher characteristics and student achievement: Evidence from Teach for America. *Mimeo, Harvard University*.

Duckworth, A. L. & Quinn, P. D. (2009). Development and Validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment, 91*(2), 166-174.

Francis, D. J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Genesee, F., Lindholm-Leary, K., Saunders, W. M., & Christian, D. (Eds.) (2006). *Educating English language learners: A synthesis of research evidence.* Cambridge University Press.

Gersten, R., Baker, S.K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Glazerman, S., S. Dolfin, M. Bleeker, A. Johnson, E. Isenberg, J. Lugo-Gil, M. Grider, and E. Britton. (2008). *Impacts of comprehensive teacher induction: Results from the first year of a randomized controlled study* (NCEE 2009-4034). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Graham, J.D. (2006). *Questions and answers when designing surveys for information collections*. Washington, D.C.: Office of Management and Budget.

Kane, T.J. & D.O. Staiger. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. *MET Project Policy and Practice Brief prepared for Bill and Melinda Gates Foundation*.

May, H., Perez-Johnson, I., Haimson, J., Sattar, S., & Gleason, P.. (2009). *Using state tests in education experiments: A discussion of the issues* (NCEE 2009-013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Price, C., Goodson, B., Dynarski, M., Unlu, F., Caswell, L., & Shivji, A.. (2013). A review of the research on the relationship between teacher preparation programs and teacher effectiveness. Unpublished manuscript. Bethesda. MD: Abt Associates.

Schochet, P.Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics, 33*(1), 62–87.