A GUIDE FOR DEVELOPING A PROSPECTIVE EVALUATION OF TEACHERS INSTITUTE SEMINARS

Ellen Eliason Kisker
Twin Peaks Partners, LLC

July 2006

CONTENTS

INTRODUCTION

As new Teachers Institutes are implemented in cities across the country, plans for evaluating them are being formed. This document outlines potential evaluation designs, discusses the advantages and challenges of the evaluation designs that could be considered, and describes the sample design, data collection plans, and analysis strategies that each imply. The discussion is necessarily general, as many details of program implementation and data availability are not yet known.

The next section describes the Teachers Institutes and their expected impacts. The following sections describe evaluation research questions, discuss potential study designs for studying program impacts, and describe the sample designs, recruitment strategies, data collection plans, and analysis strategies for the main design options.

TEACHERS INSTITUTES, THEIR THEORY OF CHANGE, AND EXPECTED IMPACTS

Teachers Institute seminars, led by university faculty, offer public school teachers opportunities to increase their knowledge and skills in what they teach. Topics are suggested by teachers through teacher Representatives who are responsible for canvassing their colleagues and soliciting ideas for seminar topics. The seminars, which meet over a period of no less than three months, focus on increasing content knowledge, improving writing and oral presentation skills, facilitating interactions between teachers in different grades and schools, and supporting teachers in developing substantial curriculum units. Teachers receive a stipend in recognition of the time they are giving to improve teaching and to cover their expenses of participation.

Teachers Institute seminars aim to improve student performance by improving teacher quality. Quality teachers are defined as those who "really know their subjects, not just 'how' to teach; …have good basic writing, math, and oral presentation skills; …expect their students to achieve; … are enthusiastic about teaching; and…can motivate even highly disadvantaged students to learn." Thus, Teachers Institute seminars are designed to improve teacher quality in the following ways:

- By focusing on content areas that teachers themselves have identified as areas they want to know more about for their teaching, participation in Teachers Institute seminars is expected to help teachers deepen their knowledge of the subjects they teach.
- By guiding each teacher in preparing a substantial curriculum unit based on research on a topic chosen by the teacher and informed by the seminar, and by encouraging discussions of these units among teachers in the seminar, participation in Teachers Institute seminars is expected to improve teachers' writing and oral presentation skills.
- Because teachers will later use the curriculum units they developed themselves, participation in Teachers Institute seminars is expected to provide teachers with curriculum materials that they are more strongly motivated to teach and that more effectively motivate students to learn.
- Through their service as teacher Representatives or seminar Coordinators, teachers are expected to develop leadership capabilities that may lead to further career development.
- Because teachers attend seminars with teachers from other schools and disciplines, Teachers Institute seminar participation is expected to promote the development of teacher networks that may offer new opportunities for learning about what other teachers are doing in their classrooms and help establish supportive peer relationships among teachers.

Teachers Institutes are designed to influence teacher outcomes directly, and these effects are expected to lead to improvements in student achievement. Thus, it is expected that program impacts on teacher outcomes will be more discernable than those on student outcomes, which are also influenced by many other educational and noneducational factors.

POTENTIAL RESEARCH QUESTIONS

The research questions of primary interest to evaluators of Teachers Institutes focus on the impacts of Teachers Institute seminars on teachers and students:

- Do Teachers Institute seminars improve students' short-term (and/or long-term) educational outcomes?
- Do Teachers Institute seminars improve teachers' careers?

To understand and interpret the answers to these questions, it is important to document participation in and completion of seminars, understand the experiences of teachers who participate, and identify the ways that their experiences may translate into changes in their teaching and students' responses to it. It is also important to document teachers' participation in alternative professional development experiences. Questions about implementation of the Teachers Institute seminars and comparison conditions, which can aid in interpreting the findings on program impacts include:

- What percentage of teachers selected for the Teachers Institute seminars completed them? What were the focus and characteristics of the curricula they produced?[1] How satisfied were participants with their seminar experiences?
- What professional development programs did teachers in the comparison group participate in, and what was the focus and intensity of the programs?

Funders asked to support the Teachers Institute seminars may have questions about how program costs compare with other professional development approaches with similar goals and similar impacts:

- What did it cost (in money and in-kind resources) to provide Teachers Institute seminars to teachers who participated?

STUDY DESIGN FOR ADDRESSING QUESTIONS ABOUT PROGRAM IMPACTS

The most rigorous evaluation design for addressing questions about program impacts is a random assignment evaluation (sometimes called a randomized control trial). In a random assignment evaluation, a lottery-like process is used to assign people or groups (in this case, teachers or schools) to two or more research groups whose outcomes are subsequently compared to determine the program's net impact. When implemented well, a random assignment evaluation provides irrefutable evidence for the effectiveness of a program or intervention. Because it is such a strong evaluation design, it may be helpful in convincing funders to support the evaluation, and its results may be helpful in convincing funders to support the program.

Random assignment evaluations are not always feasible. It is generally considered unethical to conduct a random assignment evaluation if it requires denying people access to services to which they are entitled. Also, to be ethical, there should be more applicants (or potential applicants) than available program slots, so the study does not lead to any reduction in the number of people served, only a reallocation of services among eligible applicants. Sometimes it is not possible to identify procedures for random assignment that do not change the program itself in important ways.

If a random assignment evaluation is not feasible, evidence suggests that a comparison-group study in which statistical methods are used to create equivalent research groups is the best alternative design (Coalition for Evidence-Based Policy [CEBP], 2006). Under this design, comparison schools or teachers who have been matched to participating schools or Fellows based on their observed characteristics are assumed to be similar in terms of their unobserved characteristics as well, and the potential biases in estimated program effects may be minimized.

If neither of these evaluation designs is feasible, a value-added approach could be taken to obtain a measure of the contribution of Teachers Institute Fellows to student achievement in their district. In a value-added approach,[2] the unique contribution of having a teacher who participated in a Teachers Institute to student achievement during a school year would be estimated using models in which a student's end-of-year achievement is a function of the student's achievement at the beginning of the school year, the student's background and characteristics, the school attended by the student, and an indicator of whether the student's teacher was a Teachers Institute Fellow. Although these models would provide an estimate of the contribution of Teachers Institute Fellows to student achievement relative to other teachers in the district, this contribution could not be attributed specifically to participation in Teachers Institute seminars. Like other valued-added models, this one would rely on observational rather than experimental data, making it possible

that other factors associated with teachers' opportunities and decisions to participate in Teachers Institute seminars or factors associated with students' class assignments contributed to the estimated differences in achievement associated with having a teacher who was a Fellow.

IS A RANDOM ASSIGNMENT EVALUATION OF TEACHERS INSTITUTES FEASIBLE?

Three types of random assignment evaluation designs can be considered for the evaluation of Teachers Institutes. In the first, teachers would be randomly assigned to the intervention group or a comparison group. Teachers in the intervention group would be encouraged to participate in the Teachers Institute. Teachers assigned to the comparison group would not be permitted to participate in the Teachers Institute for the period of the evaluation (but could participate in any other professional development activities that are available to them). Teachers in the comparison group could be offered priority for enrollment in a Teachers Institute seminar at the conclusion of the study (i.e., after outcomes have been measured for the last time).

Because recruiting teachers and designing seminars are closely intertwined processes, any evaluation design that requires recruiting teachers, randomly assigning them to research groups, and then designing seminars could dramatically alter the Teachers Institute being evaluated. A key feature of the Teachers Institute approach is the presence of a teacher Representative in each school whose responsibility is to identify all eligible teachers, assess their interests, work with Teachers Institute staff and eligible teachers to develop seminars that build on the identified interests, and recruit teachers to participate in the seminars. By design, seminars are developed to meet the demand for them among teachers, and program staff believe that recruiting more teachers than can be accommodated in seminars is not likely to be possible without reducing the number of seminars offered. Thus, this design will not be considered further.

In the second type of random assignment design, schools would be randomly assigned to the intervention group or the comparison group. In intervention schools, a teacher Representative would be identified and the process of recruiting teachers and developing seminars would proceed according to usual practices. In comparison schools, no Teachers Institute activities would occur, and teachers in those schools would not be eligible to participate in a seminar until after the conclusion of the evaluation period.

Random assignment of schools may be feasible if a Teachers Institute does not have the resources to offer seminars to all teachers in a school district. When a Teachers Institute is able to expand (or when a new Teachers Institute begins operation), twice as many schools as can be served can be identified, and the schools can be randomly assigned to the intervention group or to the comparison group.

Randomly assigning schools rather than teachers also has the advantage that spillover of program effects to comparison group members is less likely to bias impact estimates downward when schools are the unit of random assignment. Sharing information and program-created curriculum units is likely to occur among teachers within a school, but such sharing may be less likely to occur between teachers in different schools.

The estimated impacts produced using this school-level random assignment design will provide an estimate of the combined direct effects (through participation in a seminar) and indirect effects (through use of Teachers Institute curricula and interactions with participating teachers) of the Teachers Institute seminars on the outcomes of all teachers and students in the intervention schools.

In principle, a third type of random assignment design could be implemented. Specifically, school districts could be randomly assigned to the intervention or the comparison group. In intervention districts, Teachers Institute seminars would be implemented as designed, and in comparison districts, no Teachers Institute activities would be implemented. The primary advantage of this design is that it would virtually eliminate biases due to spillover effects. Because Teachers Institutes have already been developed in particular districts, however, this design is not feasible at this time.

THE BEST ALTERNATIVE IF A RANDOM ASSIGNMENT EVALUATION IS NOT FEASIBLE

If, after careful consideration, random assignment is determined to be infeasible, the best alternative is likely to be a matched comparison group evaluation design. CEBP (2006) has laid out principles to guide the design of comparison-group studies. First, intervention and comparison group members should be very closely matched on characteristics that may predict their outcomes. The best predictors of outcomes (and the most important to match on) are pre-intervention measures of the outcomes. Of secondary importance is matching on demographic and other background characteristics. Propensity score matching methods may be used to accomplish the matching.

Second, the comparison group should not be comprised of individuals, schools, or districts that had the option to participate in the intervention but declined, because they may differ systematically from individuals, schools, or districts that chose to participate in terms of their motivation and other important characteristics that may not be measurable. This principle poses the greatest challenge for the evaluation of the Teachers Institute seminars. It may not be possible to adhere to this principle, but it is worth considering approaches to recruiting that may enable adherence. For example, if some schools, although eligible, do not have a teacher Representative, and teacher recruiting is minimal or nonexistent there, consideration could be given to whether the schools are sufficiently similar to those with a

teacher Representative and whether teachers with characteristics and prior outcomes similar to those of intervention group teachers could be identified at those schools for the comparison group.

Third, the study should choose the intervention and comparison groups "prospectively" to avoid the possibility that the two groups are formed in such a way that they generate desired findings, consciously or unconsciously. In the case of the Teachers Institutes, comparison teachers, schools, or districts should be identified when intervention teachers have been identified, before the seminars are conducted and outcomes data are collected.

Selecting comparison school districts would have the benefit of eliminating spillover effects that might occur between intervention and comparison teachers within a school and between schools within a district. Forming a comparison group of matching school districts, however, would require gaining the cooperation of and obtaining school records data from school districts receiving no direct benefits. The costs of implementing a district-level comparison group design, even if feasible, are likely to exceed the resources available for evaluating the Teachers Institutes, so this design is not considered further here.

Selecting intervention and comparison schools, or selecting comparison teachers matched to Fellows from schools without teacher Representatives or participating teachers, would eliminate or minimize biases due to spillover effects between teachers within a school. Spillover effects between schools in the district through use of curriculum units created by Teachers Institute participants may still bias impact estimates downward.

OTHER EVALUATION DESIGN CONSIDERATIONS

Regardless of whether an experimental design or a nonexperimental comparison group design is chosen, it is important to consider what other professional development opportunities teachers have and whether there will be a meaningful difference in the professional development activities of teachers in the intervention and comparison groups. If teachers in the comparison group are able to access other programs or opportunities like those available to the teachers in the Teachers Institute seminars, then the evaluation may find no benefit of offering Teachers Institute seminars, even though the seminars themselves are of great value.

Because the curriculum materials produced by Teachers Institute participants are available to other teachers in their school district, spillover effects on comparison group teachers are a very real possibility. If these spillover effects are pervasive or large, comparisons of intervention and comparison teachers will be biased toward showing no impacts and could substantially underestimate the true program impacts. The design involving random assignment of schools or selection of com-

parison schools addresses some of the potential spillover effects by examining the effects of offering Teachers Institutes in a school, which encompass the effects on teachers who participate as well as effects on their colleagues within their school. To avoid bias due to the remaining spillover effects on teachers in comparison schools, the design would have to make school districts the unit of random assignment (or the unit for selecting members of the comparison group). As discussed above, such designs are not practical for evaluating the Teachers Institute seminars. If this remaining bias due to spillover effects cannot be eliminated, it would be useful to document the extent to which this may be an important source of bias, for example by documenting the use of curriculum units by other teachers and learning about the extent to which Fellows share what they've learned and materials they've developed or received with their colleagues.

Teacher and student mobility during the evaluation period may also present challenges for the evaluation. For both teachers and students, mobility out of the school district will result in sample attrition, and the analysis results will necessarily pertain only to the "stayers" and will not capture any impacts on those who left. If mobility is different in the intervention and comparison groups, the impact estimates may also be biased by differential attrition. It will be important to assess the amount of attrition and any differences in attrition between the intervention and comparison groups.

Among sample members, assignment to the intervention or comparison groups will be established at the beginning of the evaluation (through random assignment or matching), and to preserve the integrity of the design, these assignments must be retained in the analysis. Thus, if comparison group teachers move to intervention schools and participate in a seminar or benefit from the participation of new colleagues, estimated impacts on teachers and their students may be biased downward. Similarly, if students of comparison teachers change schools and are placed in the classrooms of teachers in the intervention group, estimates of impacts on students may be biased downward. If the extent of these crossovers is not trivial, the analysis can make adjustments to the impact estimates to account for this type of crossover between research groups.

SAMPLE DESIGN

The evaluation should be designed to ensure that the sample is large enough and that the study will follow people long enough to yield a reliable conclusion on whether the Teachers Institute seminars were or were not effective. This might require studying multiple cohorts of teachers or pooling samples across Teachers Institute sites.

STUDY POPULATIONS

Detailed criteria for defining the study populations will need to be developed after a study design is selected. If an experimental design in which schools are randomly assigned to the treatment and comparison groups is selected, the study population will consist of all teachers in the schools who are eligible for Teachers Institute seminars based on the grades and subjects they teach, and their students. If student records cannot be linked to teachers, the student population will necessarily include the entire population of students in the intervention and comparison schools.

If a comparison group design is selected, the treatment group will consist of Teachers Institute Fellows, and the comparison group will be constructed using propensity score matching to identify a similar group of eligible teachers. If comparison teachers must be selected from schools that also employ Teachers Institute Fellows, the inability to link students with teachers may make it impossible to assess the impact of Teachers Institute seminars on students.

TARGET SAMPLE SIZES AND POWER ESTIMATES

What is the magnitude of impacts expected from the intervention? What is the magnitude of impacts that would be policy-relevant? It is important in finalizing the study design to determine what the sample size required to detect these impacts would be, and to plan the size of the program or the duration of the evaluation such that these sample sizes can be achieved.[3]

Power to Detect Impacts with a Random Assignment Design (Schools)

Table 1 illustrates minimum detectable differences in teacher outcomes that might be expected with different school sample sizes. For example, in one district with 20 participating schools (10 intervention and 10 comparison schools) with teacher response rates of 80 percent, an increase in teacher retention of 9 percentage points or more could be detected statistically. If data from three districts of this size were pooled, an increase in teacher retention of 5 percentage points or more could be detected statistically.

Table 2 illustrates the corresponding minimum detectable differences in student outcomes that might be expected with the different school sample sizes in a random assignment design. For example, in one district with 20 participating schools (10 intervention and 10 comparison schools) with 80 percent of students scoring above a test score threshold, an increase in average student test scores of 1.64 normal curve equivalent points or more and an increase of 2.0 percentage points or more in the percentage of students scoring above the threshold could be detected statistically. If data from three districts of this size were pooled, an increase in student test scores of 0.95 normal curve equivalent points or more and an increase of 1.1 percentage points or more in the percentage of students scoring above the threshold could be detected statistically.

Power to Detect Impacts with a Comparison Group Design (Teachers)

Table 3 illustrates minimum detectable differences in teacher outcomes that might be expected with different school sample sizes in a comparison group design with three intervention and three comparison teachers per school. For example, in one district with 20 participating schools (10 intervention and 10 comparison schools) with teacher response rates of 80 percent, an increase in teacher retention of 18.1 percentage points or more could be detected statistically. If data from three districts of this size were pooled, an increase in teacher retention of 10.5 percentage points or more in the percentage of students scoring above the threshold could be detected statistically.

Table 4 illustrates the corresponding minimum detectable differences in student outcomes that might be expected with the different school sample sizes in a comparison group design with three intervention and three comparison teachers per school. For example, in one district with 20 participating schools (10 intervention and 10 comparison schools) with 80 percent of students scoring above a test score threshold, an increase in average student test scores of 4.02 normal curve equivalent points or more and an increase of 4.3 percentage points in the percentage of students scoring above the threshold could be detected statistically. If data from three districts of this size were pooled, an increase in student test scores of 2.32 normal curve equivalent points or more and an increase of 2.5 percentage points in the percentage of students scoring above the threshold could be detected statistically.

These minimum detectable differences in teacher and student outcomes are larger than those under the random assignment design, but they pertain to impacts on participating teachers, rather than impacts on all eligible teachers in the school. Thus, under this design, measured impacts would be expected to be larger than under the random assignment design.

RECRUITING/IDENTIFYING STUDY PARTICIPANTS

If a random assignment design is selected, the first step in establishing the study sample is to identify the schools that will be included in the study and gain their informed consent for participating in the random assignment and intervention if they are selected. The process for doing this is likely to require discussion and negotiation between Teachers Institute staff and school district personnel and is likely to vary among school districts and possibly schools.

Once the study schools have been recruited, they can be randomly assigned to the intervention and comparison groups. To maximize the comparability of the treatment and control group schools on the basis of their observable characteristics, it is advisable to select the sample by pairing similar schools using data on school and community characteristics, and randomly assigning one of each pair to the intervention group and one to the comparison group. The Teachers Institute would then proceed to identify teacher Representatives in the intervention schools, recruit teachers, and develop seminars following usual practices.

If a comparison group design is selected, the approach to identifying the intervention and comparison groups will require more consideration. First, intervention teachers should be identified through the usual Teachers Institute recruiting procedures. When the intervention teachers have been identified, but before the intervention has been implemented, comparison group teachers need to be identified. The best approach to identifying the comparison group may differ among Teachers Institute locations.

If there are any schools in which a teacher Representative was not identified or no teachers were recruited, it is worth exploring whether these schools (or a subset of them) are similar in terms of their key characteristics (type of school, grades offered, size, aggregate student characteristics) to the schools in which teachers were successfully recruited. If so, comparison teachers could be identified by (1) developing a model for predicting seminar participation using data from teachers in the schools that had a teacher Representative, (2) using the model to estimate a propensity score (the probability of participation) for each teacher in the intervention group and in the comparison schools, and (3) using the propensity scores to identify matched teachers from the comparison schools for the comparison group. This approach has the advantage that the comparison group members did not really have the option to participate in a Teachers Institute seminar, and they are less likely to differ from intervention group teachers in unobserved characteristics such as motivation.

If a teacher Representative was identified and active recruiting took place in all or most schools, then comparison teachers will have to be selected from among eligible teachers who did not apply. A propensity score matching process like that described above can be used to select the comparison teachers. In this scenario, the model for predicting seminar participation would be estimated using data for all eligible teachers.

DATA COLLECTION PLAN

To address the research questions, data on both teacher and student outcomes must be obtained. Relevant teacher outcomes that are likely to be available from school records include:[4]

- **Retention** — whether or not a teacher teaching in district schools during the previous school year was still teaching in the school district two years later.[5] Additional measures of retention (such as retention in the current school) can also be explored.
- **Promotion** to a leadership position — whether or not a teacher teaching in the school district during the previous school year was promoted to a leadership position
- **Awards** — whether or not a teacher teaching in the school district during the previous school year received an award or recognition
- **Attendance** — number of days present

Additional relevant teacher outcomes could be assessed by conducting a survey of teachers in the treatment and comparison groups. These outcomes include:

- **Motivation** — degree of motivation to teach
- **Communication** — extent of communication and collaboration with other teachers in the district
- **Teaching Practices** — strategies used to motivate students to learn and curricula used
- **Career Plans** — degree of commitment to teaching in the school district

Relevant student outcomes that are likely to be available from school records include:[6]

- **Achievement Test Scores** — scores on standardized tests in key subject areas, as well as indicators of low scores (for example, scoring more than one standard deviation below the mean) and high scores (more than one standard deviation above the mean)
- **Grades** — for grades in which students receive letter grades, the student's grade point average
- **Awards and Recognition** — whether or not a student received an award or recognition during the last year
- **Attendance** — number of days present
- **Grade Retention** — whether or not a student was held back in his or her current grade (this may also be examined in terms of graduation to the next grade or school level)

In addition to the outcomes data, data on teacher and student background characteristics are needed. These variables are important for establishing the comparability of treatment and comparison teachers and students initially and will be used as covariates in the statistical models.

DATA SOURCES

The primary source for these data will be the school records for teachers and students that will be requested from the school districts in which Teachers Institutes are working. Appendix A describes the range of data to be considered in the data requests to school districts.

These data could be supplemented by survey data collected from treatment and comparison teachers (or a random sample of the teachers). The survey data would include outcomes and background characteristics not available in the school records data, such as teacher motivation, practices and strategies for motivating students, curricula used, and communication with and support from other teachers during a recent period. A baseline survey of teachers, conducted prior to random assignment, could assess teacher outcomes prior to the intervention and collect other background information not available in school records. One or more follow-up surveys could assess the additional teacher outcomes after the Teachers Institute seminars have been completed.

The survey mode and procedures would depend on the resources available. The least expensive approach to conducting the survey would be to distribute self-administered questionnaires to teachers in school or by mail, asking them to return the completed survey by a deadline. To achieve an acceptable return rate of the questionnaires, incentives may need to be offered to teachers for returning a completed questionnaire. If sufficient resources are available, calling teachers who do not return a completed questionnaire to complete the questionnaire with them by telephone could enhance the survey response rates and may be critical for obtaining a reasonable response rate (at least 70 percent).

In addition to or instead of a survey, qualitative research methods could be used to collect information related to outcomes not measured in school records. For example, focus group discussions with teachers in the intervention and comparison groups could provide useful insights into ways in which Teachers Institute participation may have affected teachers' motivation, teaching practices, and communication. This information would not be representative or provide quantitative estimates of program impacts on these outcomes, but it would be very useful in interpreting other impact estimates.

ANALYSIS PLAN

DATA QUALITY ASSESSMENT

Prior to developing analytic models, it is important to assess the quality of the data. These analyses should assess the completeness of the data and examine the extent of data problems. Analyses of the completeness of the data will assess whether all teachers/students are represented in the data and whether all data items are available for each teacher/student. If significant amounts of data are missing, then methods for imputing data and analytic methods that can handle missing data need to be considered.

It is also important to examine the data for inaccuracies that could affect the analytic results. For example, are there inconsistencies in the values of related items in the data, or are there values that are out of range and clearly erroneous? All data problems should be examined and reasonable solutions considered, including the possibility of consulting district staff to resolve problems with school records data.

DESCRIPTIVE ANALYSES

Before estimating multivariate models, it is useful to conduct descriptive analyses to provide context for and suggest possible refinements to the more complex analyses discussed in the next section. These analyses may also reveal further data issues that should be resolved before more complex analyses are conducted.

Descriptive analyses of aggregate levels and trends in program participation and outcomes provide an important starting point for more complex multivariate analyses designed to isolate, insofar as possible, the effects of Teachers Institute seminar participation on teacher and student outcomes. Useful school-level descriptive analyses include analyses of:

- Average student background characteristics, achievement outcomes, and attendance outcomes by grade for the current year and several prior years, if available
- Student and teacher mobility between schools and into and out of the school district
- Average teacher retention, promotion, and attendance rates for the current year and several prior years, if available
- Average proportion of teachers who participated in Teachers Institute seminars

An examination of levels and trends in these school averages and correlations between the concentration of Fellows and teacher and student outcomes will provide a broad-brush look at what has been happening during the study period. If the

correlations are not positive, this is not necessarily an indication that Teachers Institute seminar participation has not had a positive influence. It may indicate that other factors, such as changes in the composition or distribution of students or teachers, may have had a greater influence, and that these need to be considered in the multivariate analyses.

For teachers, useful descriptive analyses include:

- Analyses of the characteristics and previous outcomes of teachers in the intervention and comparison groups
- Analyses of seminar participation by intervention group teachers
- Analyses of trends in outcomes of intervention and comparison teachers over time (if available), including retention, promotions, awards, and attendance
- Analyses of patterns of participation and outcomes over time for key subgroups of teachers and schools

Comparing the characteristics and previous outcomes of intervention and comparison teachers is important for assessing the initial equivalency of the two groups. Any observed initial differences between the treatment and comparison groups should be included in the statistical models. If the random assignment evaluation design is selected, examining differences between treatment and comparison schools is important for assessing the integrity of random assignment. Controlling for these differences when estimating impacts will increase the precision of the estimates. If a comparison group design is selected, comparing the characteristics and previous outcomes of Fellows and comparison teachers is important for assessing the balance in characteristics achieved through matching. If initial differences between the groups remain, it is important to control for these differences when estimating impacts. The comparisons may also provide insights into potential biases in comparisons of outcomes of treatment and comparison teachers that may remain even after observed differences are controlled.

For students, useful descriptive analyses include:

- Analyses of demographic characteristics of students of intervention and comparison teachers/schools, if the available data permit student and teacher data to be linked (again, to assess the integrity of random assignment or to understand any observed differences between the treatment and comparison groups that may be useful for interpreting estimates of program effects on students and assessing potential biases in these estimates)
- Analyses of patterns of student exposure to intervention teachers
- Analyses of patterns of student outcomes in intervention and comparison classrooms/schools
- Analyses of patterns of student achievement and attendance outcomes for key subgroups of students, teachers, and schools

MULTIVARIATE ANALYSES TO ESTIMATE THE EFFECTS OF INSTITUTE PARTICIPATION ON TEACHER OUTCOMES

If a random assignment evaluation design is selected, simple descriptive comparisons of outcomes for intervention and comparison schools will provide unbiased estimates of the effects of offering Teachers Institute seminars. The precision of these estimates can be improved by estimating impacts using a multivariate model that controls for variations in school and teacher characteristics that arise by chance. Because the study results will not be generalized beyond the study schools, school effects can be considered fixed. Thus, the teacher outcome measured at an appropriate follow-up time can be viewed as a function of the pre-intervention level of the outcome, an indicator of whether the teacher taught in an intervention or comparison school, teacher characteristics, and school characteristics:

$$Y^1 = a + bT + cY^0 + E_m d_m X_m + E_n e_n S_n + E_p f_p B_p + r_i$$

where :

Y^1 = Teacher outcome (retention, promotion, attendance, award receipt)

Y^0 = Pre-intervention teacher outcome

T = Treatment indicator equal to 1 if the teacher is in a school assigned to the treatment group and 0 if the teacher is in a school assigned to the comparison group

X_m = m teacher characteristics measured prior to the intervention

S_n = n school dummy variables or characteristics measured prior to the intervention

B_p = indicator variables for pairs used in pairwise matching prior to random assignment

a, b, c, d, and e = coefficients to be estimated

r_i = error term

In this model, b represents the unbiased regression-adjusted impact on eligible teachers of offering Teachers Institute seminars in their school. Because the Teachers Institute seminars are conducted centrally with all accepted teachers (and the intervention is the same across schools and teachers), each teacher is weighted equally in the analysis.

If the study design is not an experimental design, simple comparisons of outcomes for intervention and comparison teachers cannot be relied on to provide unbiased estimates of the effects of participation in Teacher Institute seminars. To minimize the potential biases in these estimates, it is important to conduct multivariate analyses that control for as many other observed differences among teachers as possible. Thus, the teacher outcome measured at an appropriate follow-up time can be viewed as a function of the pre-intervention level of the outcome, an indicator of whether the teacher was in the intervention or comparison group, teacher characteristics, and school characteristics:

$$Y^1 = a + bT + cY^0 + E_m d_m X_m + E_n e_n S_n + r_i$$

where:
Y^1 = Teacher outcome (retention, promotion, attendance, award receipt)
Y^0 = Pre-intervention teacher outcome
T = Treatment indicator equal to 1 if the teacher is in the intervention group and 0 if the teacher is in the comparison group
X_m = m teacher characteristics measured prior to the intervention
S_n = n school dummy variables or school characteristics
a, b, c, d, and e = coefficients to be estimated
r_i = error term

In this model, which is similar to the one above except that T is an indicator of the teacher's group assignment, the estimated coefficient b provides an estimate of the effect on teachers of acceptance to participate in a Teachers Institute seminar.[7] Because teachers were not selected randomly for the intervention and comparison groups, this impact estimate may be biased by unobserved differences between the intervention and comparison groups.

If there are crossovers (comparison group members that moved to intervention schools or were permitted to participate in seminars), the impact estimates can be adjusted to account for this by dividing them by one minus the comparison group crossover rate. This procedure assumes that the outcomes of comparison group crossovers would have been the same if they had instead been in the intervention group to start with.

With the exception of attendance, the teacher outcomes are binary variables, and the models can be estimated using logistic regression methods. For attendance, a continuous variable, the models can be estimated using ordinary least squares methods.

Subgroup Analyses

If the number of schools and teachers included in the study is large enough, it may be possible to examine whether the effects of participation in Teachers Institute seminars differ among teachers with different characteristics and in schools with different characteristics. The descriptive analyses and models described above can be estimated using data just from teachers with the teacher characteristics of interest (such as new and experienced teachers) or schools with characteristics of interest (such as elementary, middle, K-8, and high schools).

MULTIVARIATE ANALYSES TO ESTIMATE THE EFFECTS OF INSTITUTE PARTICIPATION ON STUDENT OUTCOMES

If the available school records permit linking data for students with data for their teacher(s), three-level hierarchical models can be specified to estimate the effects of Teachers Institute seminars on students. The exact specification of the models

will need to be worked out when features of the data are known; for simplicity, reduced form random intercept models are presented here.

If the random assignment design is implemented or the comparison group is formed by identifying intervention schools and matched comparison schools, the effects on students in schools where teachers were offered Teachers Institute seminars can be estimated using a three-level hierarchical linear model. Because the study results will not be generalized beyond the study schools and all eligible teachers in the study schools will be included, school and teacher effects can be considered fixed.[8] The three levels can be aggregated into a unified model, which in its simplest form can be expressed as follows:

$$Y^1 = a + bY^0 + cT + E_n d_n X_n + E_m f_m Z_m + E_p g_p W_p + r_{ijk}$$

where:
Y^1 = the achievement test score of student i of teacher j in school k
Y^0 = the previous achievement test score of student i of teacher j in school k
T = Treatment indicator equal to 1 if school k is in the intervention group and 0 if school k is in the comparison group
X_n = n characteristics of student i of teacher j in school k
Z_m = m characteristics of teacher j in school k
W_p = p characteristics of school k
a, b, c, d_n, f_m, g_p = coefficients to be estimated
r_{ijk} = student error term

In this model, the coefficient *c* is an estimate of the effect on students of offering Teachers Institute seminars to teachers in their school.

If student data cannot be linked with teacher data, two-level hierarchical linear models (student and school levels) that include aggregate teacher characteristics as school characteristics in the model will have to be used to estimate program impacts. For example, instead of including teacher j's years of experience as a teacher-level variable, the average number of years of experience of all eligible teachers could be included as a school-level variable.

If the comparison group design in which Fellows form the intervention group and comparison teachers are identified through matching is implemented, the effects on students of having a teacher who was accepted to become a Teachers Institute Fellow also can be estimated using a three-level hierarchical linear model. Because the study results will not be generalized beyond the study schools, school effects can be considered fixed. Again, the three levels can be aggregated into a unified model, which in its simplest form can be expressed as follows:

$$Y^1 = a + bY^0 + cT + E_n d_n X_n + E_m f_m Z_m + E_p g_p W_p + r_{ijk} + u_{jk}$$

where:

Y^1 = the achievement test score of student i of teacher j in school k

Y^0 = the previous achievement test score of student i of teacher j in school k

T = Treatment indicator equal to 1 if teacher j is in the intervention group and 0 if teacher j is in the comparison group

X_n = n characteristics of student i of teacher j in school k

Z_m = m characteristics of teacher j in school k

W_p = p characteristics of school k

a, b, c, d_n, f_m, g_p = coefficients to be estimated

r_{ijk}, u_{jk}, e_k = student and teacher error terms

This model is identical to that for the random assignment design analyses above, with the exception that T denotes the treatment group status of teacher j rather than school k, the coefficient *c* is an estimate of the impact on students of their teachers' acceptance into a Teachers Institute seminar, and a teacher-level error term is included.[9]

Under this comparison group design, if student data cannot be linked with teacher data, it may not be possible to estimate the effects of Teachers Institute seminars on students. Two-level hierarchical linear models (student and school levels) that include aggregate teacher characteristics, including the percentage of teachers accepted into the Teachers Institute seminars, as school characteristics in the model may be investigated. In these models, the coefficient on the percentage of teachers accepted into the Teachers Institute seminars is an estimate of the impact on students of increasing the percentage of teachers who are Fellows in their school by one percentage point. This model would be similar to a value-added model, except that the schools included in the analysis would be limited to schools in which intervention and comparison group teachers taught.

The hierarchical linear models can be estimated using HLM6 software (Raudenbusch, Bryk, Cheong, and Congdon, 2004) or other major software packages, such as SAS or Stata.

Students who remain within the school district during the period between the previous and current testing occasions should have data for both the previous and current tests and can be included in the analyses of student impacts. Students who enter the school district after the previous testing occasion or leave the district before the current testing occasion will not have outcome data for both time points and cannot be included in the analyses. Thus, the results of the analyses will apply only to students who attend school in the district for the full period. If student mobility into and out of the district during the evaluation period is high, the results will pertain to a relatively limited portion of the student population.

Because teachers who are not Fellows are encouraged to use the curriculum resources developed by Fellows, it is possible and even likely that the teaching of comparison teachers and the achievement of their students may be improved by the Teachers Institute seminars. If curriculum units developed by Fellows are deposited in all schools, disseminated by school Representatives and Contacts, and made available on the internet, as they are in existing Teachers Institute sites, spillover effects within a school district may be significant. In the random assignment design and in a comparison group design that selects intervention schools and creates a comparison group by identifying matching schools, the estimated impacts of offering Teachers Institute seminars in a school include both the direct effects of teachers' participation and the spillover effects within the school, and only spillover between schools has the potential to bias the impact estimates. In the comparison group design that focuses on Fellows and selects a comparison group by identifying matching teachers, potential spillover effects may include spillover effects between teachers within a school as well as between schools. Including the percentage of teachers who are Fellows as a school-level characteristic in the models using comparison group data may help capture this within-school spillover.

If student data cannot be linked with data for teachers that the student had, a two-level hierarchical linear model may provide the best possible estimates of the effects of teacher participation in Teachers Institute seminars on student achievement. In this analysis, the sample of students necessarily includes all students in schools where eligible teachers taught.

Subgroup Analyses

To examine whether the effects of Teachers Institute seminars on student outcomes differ among students, teachers, and schools with different characteristics, the models described above can be estimated using data just from the students, teachers, and schools with the characteristics of interest.

INTERPRETING THE ANALYSIS RESULTS

The analyses described above will provide estimates of the effects of participation in Teachers Institute seminars on teachers and students. Descriptive analyses of data collected by the Teachers Institute from participating teachers about their perceptions of the usefulness and effectiveness of the seminars and their perceptions of the effects on students of using the new curriculum units provide additional information about the possible nature and magnitude of the effects of seminar participation. To the extent that these analyses paint consistent pictures of the effects of Teachers Institute seminar participation, the analyses together strengthen confidence in the findings of either analysis individually. To the extent that the analyses offer inconsistent information about program effects, they may raise questions that need to be addressed in future evaluations.

APPENDIX

TABLES

TABLE 1: ILLUSTRATIVE CALCULATIONS: IMPACTS OF OFFERING SEMINARS ON ALL TEACHERS IN SCHOOL MINIMUM DETECTABLE DIFFERENCES IN TEACHER OUTCOMES RANDOM ASSIGNMENT DESIGN (SCHOOLS)

Outcome	p	Number of Schools	Number of Eligible Teachers per School	Design Effect Due to Clustering	Variance sigma**2	Sample Size		Response Rates RR	Minimum Detectable Difference MDD	(a) MDD with Covariates
						Fellowsnt	Comparisonnc			
Proportion of teachers retained in district for at least two years: One Site										
.1 / .9	0.1	20	25	1.64	0.09	250	250	0.8	0.095	0.067
.2 / .8	0.2	20	25	1.64	0.16	250	250	0.8	0.127	0.089
.3 / .7	0.3	20	25	1.64	0.21	250	250	0.8	0.146	0.102
.4 / .6	0.4	20	25	1.64	0.24	250	250	0.8	0.156	0.109
.5 / .5	0.5	20	25	1.64	0.25	250	250	0.8	0.159	0.111
.1 / .9	0.1	10	25	1.64	0.09	125	125	0.8	0.135	0.094
.2 / .8	0.2	10	25	1.64	0.16	125	125	0.8	0.18	0.126
.3 / .7	0.3	10	25	1.64	0.21	125	125	0.8	0.206	0.144
.4 / .6	0.4	10	25	1.64	0.24	125	125	0.8	0.22	0.154
.5 / .5	0.5	10	25	1.64	0.25	125	125	0.8	0.225	0.157
.1 / .9	0.1	6	25	1.64	0.09	75	75	0.8	0.174	0.122
.2 / .8	0.2	6	25	1.64	0.16	75	75	0.8	0.232	0.162
.3 / .7	0.3	6	25	1.64	0.21	75	75	0.8	0.266	0.186
.4 / .6	0.4	6	25	1.64	0.24	75	75	0.8	0.284	0.199
.5 / .5	0.5	6	25	1.64	0.25	75	75	0.8	0.29	0.203
Proportion of teachers retained in district for at least two years: Two Sites										
.1 / .9	0.1	40	25	1.64	0.09	500	500	0.8	0.067	0.047
.2 / .8	0.2	40	25	1.64	0.16	500	500	0.8	0.09	0.063
.3 / .7	0.3	40	25	1.64	0.21	500	500	0.8	0.103	0.072
.4 / .6	0.4	40	25	1.64	0.24	500	500	0.8	0.11	0.077
.5 / .5	0.5	40	25	1.64	0.25	500	500	0.8	0.112	0.079
Proportion of teachers retained in district for at least two years: Three Sites										
.1 / .9	0.1	60	25	1.64	0.09	750	750	0.8	0.055	0.038
.2 / .8	0.2	60	25	1.64	0.16	750	750	0.8	0.073	0.051
.3 / .7	0.3	60	25	1.64	0.21	750	750	0.8	0.084	0.059
.4 / .6	0.4	60	25	1.64	0.24	750	750	0.8	0.09	0.063
.5 / .5	0.5	60	25	1.64	0.25	750	750	0.8	0.092	0.064

Note: The calculations assume an alpha=.05, power=.80, one-tailed test, regression R-square of .2, and intraclass correlation of 0.1.

TABLE 2: ILLUSTRATIVE CALCULATIONS: IMPACTS OF OFFERING SEMINARS ON ALL STUDENTS IN SCHOOL MINIMUM DETECTABLE DIFFERENCES IN STUDENT OUTCOMES RANDOM ASSIGNMENT DESIGN (SCHOOLS)

(Student data linked to teacher data)

Outcome	p	Number of Schools	Number of Students per School	Design Effect Due to Clustering	Variance sigma**2	Sample Size: Students in TI Schools nt	Sample Size: Students in C Schools nc	Response Rates RR	Minimum Detectable Difference MDD	(a) MDD with Covariates
Student test scores in normal curve equivalents										
		60	500	7.13	443.52	15000	15000	0.8	1.8	1.61
		40	500	7.13	443.52	10000	10000	0.8	2.21	1.98
		20	500	7.13	443.52	5000	5000	0.8	3.13	2.8
		10	500	7.13	443.52	2500	2500	0.8	4.42	3.95
		6	500	7.13	443.52	1500	1500	0.8	5.71	5.1
Proportion of students scoring above a threshold										
.1 / .9	0.1	60	500	7.13	0.09	15000	15000	0.8	0.026	0.009
.2 / .8	0.2	60	500	7.13	0.16	15000	15000	0.8	0.034	0.011
.3 / .7	0.3	60	500	7.13	0.21	15000	15000	0.8	0.039	0.013
.4 / .6	0.4	60	500	7.13	0.24	15000	15000	0.8	0.042	0.014
.5 / .5	0.5	60	500	7.13	0.25	15000	15000	0.8	0.043	0.014
.1 / .9	0.1	40	500	7.13	0.09	10000	10000	0.8	0.031	0.011
.2 / .8	0.2	40	500	7.13	0.16	10000	10000	0.8	0.042	0.014
.3 / .7	0.3	40	500	7.13	0.21	10000	10000	0.8	0.048	0.016
.4 / .6	0.4	40	500	7.13	0.24	10000	10000	0.8	0.051	0.017
.5 / .5	0.5	40	500	7.13	0.25	10000	10000	0.8	0.052	0.018
.1 / .9	0.1	20	500	7.13	0.09	5000	5000	0.8	0.045	0.015
.2 / .8	0.2	20	500	7.13	0.16	5000	5000	0.8	0.059	0.02
.3 / .7	0.3	20	500	7.13	0.21	5000	5000	0.8	0.068	0.023
.4 / .6	0.4	20	500	7.13	0.24	5000	5000	0.8	0.073	0.024
.5 / .5	0.5	20	500	7.13	0.25	5000	5000	0.8	0.074	0.025
.1 / .9	0.1	10	500	7.13	0.09	2500	2500	0.8	0.063	0.021
.2 / .8	0.2	10	500	7.13	0.16	2500	2500	0.8	0.084	0.028
.3 / .7	0.3	10	500	7.13	0.21	2500	2500	0.8	0.096	0.032
.4 / .6	0.4	10	500	7.13	0.24	2500	2500	0.8	0.103	0.034
.5 / .5	0.5	10	500	7.13	0.25	2500	2500	0.8	0.105	0.035
.1 / .9	0.1	6	500	7.13	0.09	1500	1500	0.8	0.081	0.027
.2 / .8	0.2	6	500	7.13	0.16	1500	1500	0.8	0.108	0.036
.3 / .7	0.3	6	500	7.13	0.21	1500	1500	0.8	0.124	0.042
.4 / .6	0.4	6	500	7.13	0.24	1500	1500	0.8	0.133	0.044
.5 / .5	0.5	6	500	7.13	0.25	1500	1500	0.8	0.135	0.045

Note: The calculations assume an alpha=.05, power=.80, one-tailed test, regression R-square of .2, and intraclass correlation of 0.1.

TABLE 3: ILLUSTRATIVE CALCULATIONS: IMPACTS OF SEMINARS ON PARTICIPATING TEACHERS MINIMUM DETECTABLE DIFFERENCES IN TEACHER OUTCOMES COMPARISON GROUP DESIGN (GROUPS FROM SAME SCHOOLS)

Outcome	p	Number of Schools	Design Effect Due to Clustering	Variance sigma**2	Sample Size: Fellows nt	Sample Size: Comparison nc	Response Rates RR	Minimum Detectable Difference MDD	(a) MDD with Covariates
Proportion of teachers retained in district for at least two years: One District									
.1 / .9	0.1	20	1.12	0.09	60	60	0.8	0.161	0.136
.2 / .8	0.2	20	1.12	0.16	60	60	0.8	0.215	0.181
.3 / .7	0.3	20	1.12	0.21	60	60	0.8	0.246	0.208
.4 / .6	0.4	20	1.12	0.24	60	60	0.8	0.263	0.222
.5 / .5	0.5	20	1.12	0.25	60	60	0.8	0.268	0.227
Proportion of teachers retained in district for at least two years: One District, Two Cohorts									
.1 / .9	0.1	20	1.24	0.09	120	120	0.8	0.12	0.096
.2 / .8	0.2	20	1.24	0.16	120	120	0.8	0.16	0.128
.3 / .7	0.3	20	1.24	0.21	120	120	0.8	0.183	0.147
.4 / .6	0.4	20	1.24	0.24	120	120	0.8	0.196	0.157
.5 / .5	0.5	20	1.24	0.25	120	120	0.8	0.2	0.16
Proportion of teachers retained in district for at least two years: Two Districts									
.1 / .9	0.1	40	1.12	0.09	120	120	0.8	0.114	0.096
.2 / .8	0.2	40	1.12	0.16	120	120	0.8	0.152	0.128
.3 / .7	0.3	40	1.12	0.21	120	120	0.8	0.174	0.147
.4 / .6	0.4	40	1.12	0.24	120	120	0.8	0.186	0.157
.5 / .5	0.5	40	1.12	0.25	120	120	0.8	0.19	0.16
Proportion of teachers retained in district for at least two years: Two Districts, Two Cohorts									
.1 / .9	0.1	40	1.24	0.09	240	240	0.8	0.085	0.068
.2 / .8	0.2	40	1.24	0.16	240	240	0.8	0.113	0.091
.3 / .7	0.3	40	1.24	0.21	240	240	0.8	0.13	0.104
.4 / .6	0.4	40	1.24	0.24	240	240	0.8	0.139	0.111
.5 / .5	0.5	40	1.24	0.25	240	240	0.8	0.141	0.113
Proportion of teachers retained in district for at least two years: Three Districts									
.1 / .9	0.1	60	1.12	0.09	180	180	0.8	0.093	0.079
.2 / .8	0.2	60	1.12	0.16	180	180	0.8	0.124	0.105
.3 / .7	0.3	60	1.12	0.21	180	180	0.8	0.142	0.12
.4 / .6	0.4	60	1.12	0.24	180	180	0.8	0.152	0.128
.5 / .5	0.5	60	1.12	0.25	180	180	0.8	0.155	0.131
Proportion of teachers retained in district for at least two years: Three Districts, Two Cohorts									
.1 / .9	0.1	60	1.24	0.09	360	360	0.8	0.069	0.056
.2 / .8	0.2	60	1.24	0.16	360	360	0.8	0.092	0.074
.3 / .7	0.3	60	1.24	0.21	360	360	0.8	0.106	0.085
.4 / .6	0.4	60	1.24	0.24	360	360	0.8	0.113	0.091
.5 / .5	0.5	60	1.24	0.25	360	360	0.8	0.116	0.093

Note: The calculations assume an alpha=.05, power=.80, one-tailed test, regression R-square of .2, and intraclass correlation of 0.1.

TABLE 4: ILLUSTRATIVE CALCULATIONS: IMPACTS OF SEMINARS ON STUDENTS OF PARTICIPATING TEACHERS MINIMUM DETECTABLE DIFFERENCES IN STUDENT OUTCOMES COMPARISON GROUP DESIGN (GROUPS FROM SAME SCHOOLS)

(Student data linked to teacher data)

Outcome	p	Number of Schools	Number of Teachers per School	Number of Students per Teacher	Design Effect Due to Clustering	Variance sigma**2	Sample Size: Students of Fellowsnt	Sample Size: Students of Comparisonnc	Response Rates RR	Minimum Detectable Difference MDD
Student test scores in normal curve equivalents										
		60	6	18	3.19	443.52	3240	3240	0.8	2.6
		40	6	18	3.19	443.52	2160	2160	0.8	3.18
		20	6	18	3.19	443.52	1080	1080	0.8	4.5
		10	6	18	3.19	443.52	540	540	0.8	6.36
		6	6	18	3.19	443.52	324	324	0.8	8.21
Proportion of students scoring above a threshold										
.1 / .9	0.1	60	6	18	3.19	0.09	3240	3240	0.8	0.037
.2 / .8	0.2	60	6	18	3.19	0.16	3240	3240	0.8	0.049
.3 / .7	0.3	60	6	18	3.19	0.21	3240	3240	0.8	0.057
.4 / .6	0.4	60	6	18	3.19	0.24	3240	3240	0.8	0.06
.5 / .5	0.5	60	6	18	3.19	0.25	3240	3240	0.8	0.062
.1 / .9	0.1	40	6	18	3.19	0.09	2160	2160	0.8	0.045
.2 / .8	0.2	40	6	18	3.19	0.16	2160	2160	0.8	0.06
.3 / .7	0.3	40	6	18	3.19	0.21	2160	2160	0.8	0.069
.4 / .6	0.4	40	6	18	3.19	0.24	2160	2160	0.8	0.074
.5 / .5	0.5	40	6	18	3.19	0.25	2160	2160	0.8	0.076
.1 / .9	0.1	20	6	18	3.19	0.09	1080	1080	0.8	0.064
.2 / .8	0.2	20	6	18	3.19	0.16	1080	1080	0.8	0.085
.3 / .7	0.3	20	6	18	3.19	0.21	1080	1080	0.8	0.098
.4 / .6	0.4	20	6	18	3.19	0.24	1080	1080	0.8	0.105
.5 / .5	0.5	20	6	18	3.19	0.25	1080	1080	0.8	0.107
.1 / .9	0.1	10	6	18	3.19	0.09	540	540	0.8	0.091
.2 / .8	0.2	10	6	18	3.19	0.16	540	540	0.8	0.121
.3 / .7	0.3	10	6	18	3.19	0.21	540	540	0.8	0.138
.4 / .6	0.4	10	6	18	3.19	0.24	540	540	0.8	0.148
.5 / .5	0.5	10	6	18	3.19	0.25	540	540	0.8	0.151
.1 / .9	0.1	6	6	18	3.19	0.09	324	324	0.8	0.117
.2 / .8	0.2	6	6	18	3.19	0.16	324	324	0.8	0.156
.3 / .7	0.3	6	6	18	3.19	0.21	324	324	0.8	0.179
.4 / .6	0.4	6	6	18	3.19	0.24	324	324	0.8	0.191
.5 / .5	0.5	6	6	18	3.19	0.25	324	324	0.8	0.195

Note: The calculations assume an alpha=.05, power=.80, one-tailed test, regression R-square of .2, an intraclass correlation of 0.1, and an p correlation intragroup correlation of .07.

DATA NEEDS FOR EVALUATION

Below is a list of data needed for evaluating the effectiveness of Teachers Institute Seminars. This list includes all data that would be useful for the evaluation. It is unlikely that all items will be available in any particular school district, but it is important in planning for evaluation to explore the availability of these data.

TEACHER DATA

Linking Information

Teacher data need to be linked to school data, so identifiers that enable teachers to be identified according to the school in which they taught are important.

Data for forming a comparison group[10] of teachers and their students

- Teacher background characteristics (age or birth date, gender, marital status, race/ethnicity)
- Teaching assignments (grade levels, subjects, schools)
- Years of teaching experience overall and in the district
- Highest degree completed
- Field of study in college
- Field of study in graduate school (if applicable)
- Certifications received
- Eligibility for Teachers Institute seminars

Data for documenting teachers' participation/receipt of services[11]

- Dates of participation
- Which seminar(s) attended
- Attendance in each seminar
- Whether completed requirements for each seminar
- Use of Teachers Institute curricula, and dates of use (for both participating and comparison teachers)
- Participation in other professional development activities (for both participating and comparison teachers)

Data on teacher outcomes

- Retention - whether left the district, and why
- Promotions to leadership positions
- Awards and recognition
- Attendance at school
- Satisfaction with the seminar

- Motivation to teach
- Extent of communication and collaboration with other teachers in the district
- Teaching practices
- Perceived effects of seminar participation on teaching
- Career plans

STUDENT DATA

Linking Information

Student data needs to be linked to teacher data, so identifiers that enable the students of each teacher to be identified are important.

Student background data

- Age
- Grade
- Gender
- Race/ethnicity
- English language proficiency
- Special education identification
- Talented and gifted identification
- Eligibility for free or reduced-price lunch
- Date of entry into current school
- Previous test scores (prior to study period)
- Previous grades (prior to study period)
- Previous grade retention (prior to study period)
- Previous level of attendance (prior to study period)

Data on student outcomes

- Achievement test scores
- Grades in key subject areas
- Behavior/motivation/effort grade
- Academic awards and recognition
- Attendance
- Grade retention
- Whether graduated
- Whether admitted to college
- Number of absences
- Number of times tardy
- Number of suspensions
- Number of expulsions
- Whether participated in extracurricular activities, by type of activity

SCHOOL DATA

School type

- Level/grades offered
- Whether school is a magnet school
- Whether school is a vocational school

Enrollment

- Number of students enrolled
- Percentage of students enrolled in bilingual and English as a Second Language services
- Percentage of students enrolled in special education services
- Percentage of students enrolled in compensatory education
- Percentage of students enrolled in a talented and gifted program
- Student mobility between schools and in/out of district
- Student participation in extracurricular activities

Aggregate Student Demographics

- Percentage of students by gender
- Percentage of students by race/ethnicity
- Percentage of students with a non-English home language
- Percentage of students who are eligible for free or reduced-price meals

Aggregate Student Test scores, by Subject

- **Aggregate Student Behavioral Indicators**
- Number of disciplinary referrals
- Number of suspensions
- Number of expulsions
- Attendance rates
- Percentage of students tardy
- Number or percent of students receiving awards or prizes, by type of award/prize
- Percentage of students retained or held back last year
- Graduation rates
- Percentage of graduates who enter college

Aggregate Teacher Characteristics

- Number of full-time teachers
- Number of part-time teachers
- Percentage of teachers from minority racial/ethnic backgrounds

- Average number of years of teaching experience
- Percentage with a masters degree or higher
- Average days absent due to illness or personal time
- Percentage of certified staff assigned to the same school in the previous year

School Characteristics

- Number of instructional days or hours per year
- Estimated hours of instruction per year by subject area
- Number of classrooms
- Student-teacher ratio
- Average class size
- Per pupil expenditures
- Title I eligibility and participation

NOTES AND REFERENCES

NOTES

1. If criteria and tools can be developed for assessing the quality of the curriculum units, understanding the quality of the curricula also may help explain the observed impacts on student performance.

2. Value-added modeling (VAM) is "a collection of complex statistical techniques that use multiple years of students' test score data to estimate the effects of individual schools or teachers" on student achievement (McCaffrey et al. p. xi). VAM tries to separate the effects of individual teachers and schools from the effects of noneducational factors such as family background for use in teacher or school accountability systems. The idea is that VAM can be used to identify especially effective teachers for special recognition and/or especially ineffective teachers for training or professional development programs, or dismissal. VAM can also provide estimates of how much difference teachers and schools make overall in student achievement, taking into account other factors that can influence achievement.

3. To be conservative, the illustrative calculations presented in this section incorporate design effects due to clustering. As discussed in the analysis plan below, if all teachers are included and results will not be generalized beyond the study schools, site effects may be considered fixed rather than random in the analyses.

4. The list of teacher outcomes will need to be refined based on the data actually available from the school district in which the evaluation is being implemented.

5. To be eligible, teachers must be assigned classes in the school district for the following year in which they can teach the curriculum units they develop during the Teachers Institute.

6. The list of student outcomes will need to be refined based on the data actually available from the school district in which the evaluation is being implemented.

7. If all or most teachers who are accepted follow through and participate in a Teachers Institute seminar, b provides an estimate of the effect on teachers of participating in a Teachers Institute seminar.

8. If teacher attrition is high, it may be appropriate to specify the teacher effects as random.

9. If all or most teachers who are accepted and assigned to the intervention group follow through to participate in a seminar, this is an estimate of the effect on students of their teacher's participation in a Teachers Institute seminar.

10. The comparison group could be formed by randomly assigning eligible applicants to participate in the Teachers Institute seminars or be part of the comparison group (the most rigorous option), or by identifying a matched group of teachers using school records information.

11. From a management information system or other program records

REFERENCES

Coalition for Evidence-Based Policy [CEBP]. "What Constitutes Strong Evidence of a Program's Effectiveness?" www.whitehouse.gov/OMB/part/2004_program_eval.pdf, accessed August 2006.

Darling-Hammond, Linda. *Teacher Quality and Student Achievement: A Review of State Policy Evidence*. Seattle, WA: University of Washington Center for the Study of Teaching and Policy, December 1999.

McCaffrey, Daniel F., J.R. Lockwood, Daniel M. Koretz, and Laura S. Hamilton. *Value-Added Models for Teacher Accountability*. Santa Monica, CA: RAND Corporation, 2003.

Raudenbush, Stephen W., and Anthony S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition.* Thousand Oaks, CA: Sage Publications, 2002.

Raudenbush, Stephen W., Anthony S. Bryk, Yuk Fai Cheong, and Richard T. Congdon, Jr. *HLM6: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International, Inc., 2004.

Snijders, Tom, and Roel Bosker. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage Publications, 1999.