B. Collections of Information Employing Statistical Methods

1. Respondent Universe and Sample Selection

Case study sites will be selected from among the population of 2006 grantees, which number about 120 in total. We will use two data sources for site selection: 1) student assessment data from states that have agreed to provide us data for the state data analysis task; and 2) the 2008 Annual Performance Reports, which are the most recent Annual Performance Reports available for the 2006 grantees.

Site Selection to Document Changes in Student Achievement

We propose to utilize the following method below to choose case-study grantees for site visits to document practices in increasing student achievement. We anticipate having access to data from approximately 60 grantees.

We will calculate regression-adjusted differences in average pre- and post-TAH assessment scores for all of the TAH-grantee districts. The advantage of using regression-adjusted differences is that we can control for differences in student demographic characteristics between TAH grantees. These differences, which are likely to be correlated with student performance on American History assessments, would bias pre-post differences and attribute to the TAH program achievement gains (or losses) that were actually unrelated to TAH.

For each grantee, the average pre-post difference in assessment scores will be modeled as:

$(t_{2j} - t_{1j}) = \alpha + \beta X + \varepsilon$
where:

t_{1j} = American History average assessment score for all schools in grantee district j in pre-TAH year y_1;

t_{2j} = American History average assessment score for all schools in grantee district j in post-TAH year y_2;

α = intercept;

X = vector of baseline student characteristic variables;

β = vector of baseline student characteristic coefficients;

ε = random error term.

After running the linear regression in equation (1), a regression-adjusted pre-post difference in mean assessment scores for each grantee will be calculated by computing a predicted pre-post difference in average American History scores using the estimated coefficients and the mean baseline student characteristics:

$$\hat{\Delta t}_j = \hat{\alpha} + \hat{\beta} \overline{X}_j$$

where:

$\Delta \hat{t}_j$ =   predicted pre-post difference in average American History scores

$\hat{\alpha}$ =   estimated value of the intercept;

$\hat{\beta}$ =   estimated vector of baseline student characteristic coefficients;

$\overline{X}_j$ =   vector of mean baseline student characteristics for students in grantee district j.

We will then rank the grantees from highest to lowest according to their regression-adjusted pre-post average difference in assessment scores. Depending on the distribution of the rankings, we will group the grantees into different categories such as:

- Previously high-achieving districts that experienced a large change in assessment scores;
- Previously low-achieving districts that experienced a large change in assessment scores;
- Previously high-achieving districts that experienced no change in assessment scores;
- Previously low-achieving districts that experience no change in assessment scores.

Depending on the number of grantees in each category, we propose to select at least one grantee in each category for the case studies. There are a number of reasons why we feel it would be important to select case study grantees from all categories of assessment score achievement:

- By studying districts that experienced a large change in assessment scores, one can identify possible TAH-related practices that influenced the change in American History achievement. However, it is only by comparing program implementation in low and high-achieving districts that one can identify which practices may have affected assessment score performance in the high-achieving districts.
- There are any number of other factors besides TAH program implementation or student baseline characteristics that could affect changes in assessment scores across the pre- and post-TAH years: by studying low and high-achieving districts, one can examine how much of the pre-post average difference might be attributed to TAH and how much could possibly be attributed to other factors.

It is important to note that the proposed method does not represent a completely unbiased method for choosing case study grantees; for example, it does not control for regression to the mean, which is the natural tendency for extreme values to move closer to the average upon subsequent measurement. Therefore, for statistical reasons, grantee

districts that have lower than average scores in the pre-TAH year will tend to score higher in subsequent years. In addition, we stated previously that by performing case studies across a range of pre-post average assessment score differences, we could identify possible previously unobserved factors that could cause those differences, but the proposed selection method in itself does not control for these unobserved factors at all.

For these reasons, we do not view the proposed method for choosing case study grantees as entirely objective and rigorous, but we do view the proposed method as an informative way to categorize and choose grantees across a wide distribution of achievement levels.

Site Selection to Document Increases in Teacher Content Knowledge

To select case study sites for the site visits to document practices in increasing teacher content knowledge, we will use data from the 2008 APRs. Our approach for selecting the sample is designed to identify grantees that have evaluation designs and measurement tools capable of determining teacher learning gains (or lack of gains). A template to be used to review and compare grantees using evaluation data in the APRs is attached.

Using the template, we will first sort grantees based on the strength of their evaluation designs. We will use the following categories:

- Experimental
- Quasi-experimental
- Case study
- Pre-post, no control
- Other
- Not enough information.

For grantees that employ either experimental or quasi-experiment designs, we will then identify the assessment instrument in use. We will sort grantees by:

- National teacher test (e.g., Praxis)
- State teacher test
- Structured classroom observations
- Project developed teacher test
- Teacher self-assessment
- Satisfaction and effectiveness survey
- Other.

We will then identify grantees who used a national teacher test, a state teacher test, or structured classroom observations, AND who used an experimental or quasi-experimental design and report significance levels of results. Next, we will sort grantees into two groups by reported results: one with statistically significant positive (+) results, and a second with negative (-) results or no significant difference (0). Those grantees in the first group will be ranked based on effect sizes. The four grantees with the largest

statistically significant effect sizes will be chosen for the "high performing" case studies. Four grantees with no significant effects or negative effects will be chosen from the second group, either randomly or—if possible—matched to the first group based on geographic region.

Our preliminary review of a sample of 2007 APRs suggests that the majority of grantees employ pre- post designs with no controls. If too few grantees meet our design criteria for selection into the sample, we will add a sample of grantees with these weaker designs who still report significance level of results and use strong assessment instruments.

Like our method of site selection based on increases in student achievement, this approach will not draw on the full universe of 2006 grantees and will therefore not be fully representative. Since site selection necessarily will be restricted to a subsample of grantees who have the more rigorous evaluation designs, and who report outcomes by the end of the second grant year, it is possible that all of these grantees are higher performing than average, or alternatively, that the highest performing grantees are being excluded from consideration. Nevertheless, we believe the selection approach described above allows us to consider as many grantees as possible while also comparing grantee outcomes based on reasonably sound evaluation methods.

2. Data Collection

The sampling issues related to data collection activities are covered in the previous section and described in the data collection tasks and deliverables described in Exhibit 2. Research staff will ask grant directors, in advance of the site visits, to suggest a list of participating teachers who are diverse with respect to teaching experience, teaching levels and types of grant activities in which they have participated. Site visitors will work with the grant directors to determine the best method of contacting the teachers and inviting them to participate in interviews. The grant directors and their staff may choose to contact the teachers, or the site visitors may contact the teachers to schedule the interviews, as preferred.

3. Methods to Maximize Response Rates

Our strategy for maximizing response rates is to capitalize on the popularity of the TAH Grants Program among members of the history education community. For TAH project directors, for example, the novelty of the federal government's substantial investment in these grants is well known; the program's name recognition is high; and the program's approval is widespread. We will work with project directors to identify teachers and training providers who would be willing and available to participate in the case studies. Finally, we have taken (or will take) the following steps to maximize the response rates for the data collection activities:

- We have constructed all data collection instruments as concisely and tightly as possible. To the extent possible, we will coordinate data collection activities with each other to ensure that they impose a manageable burden on

respondents, while yielding data that collectively answer the evaluation questions of most interest to the government and the field.

- We will send letters of introduction to project directors in summer 2009, informing them of the study and describing all data collection activities. The letters will include contact information for BPA and SRI staff members who can answer questions about the study, will provide information about OMB clearance, and will include contact information for the study's project officer at ED.

As a result of all of the above efforts, we anticipate the response rate to be between 95% and 100%.

4. Pilot Testing

The teacher protocols will be pilot tested with two American history teachers participating in TAH grants that are not part of the current study. If possible, face to face interviews will be conducted with participants of grantees in the Washington DC or San Francisco area. If it is not possible to locate such teachers, phone interviews will be conducted to pilot the protocol.

5. Contact Information

The contact person at the Department of Education is Ms. Reeba Daniel. The primary contractor of this study is Berkeley Policy Associates, based in Oakland, California. SRI International, based in Menlo Park, CA, is the subcontractor. The principal investigator of the study is Dr. Daniel Humphrey and the project director is Dr. Phyllis Weinstock. Data collection will be conducted by researchers at both Berkeley Policy Associates and SRI International under the direction of Dr. Weinstock. The contact information for these individuals is as follows:

Reeba Daniel

U.S.Department of Education
Policy and Program Studies Services
Phone: (202) 401-3416
E-Mail: Reeba.Daniel@ed.gov

Daniel Humphrey, Ed.D.

SRI International
Phone: (650) 859-4014
E-Mail: Daniel.Humphrey@sri.com

Phyllis Weinstock, Ph.D.
Berkeley Policy Associates
510-465-7884 x221
E-Mail: Phyllis@bpacal.com