

Responses to OMB Questions on the Study of Emerging Teacher Evaluation Systems in the U.S. (#201111-1875-001)

1. *OMB's question: Can ED clarify further the process they will use to choose the case study sample? What are the selection criteria and how will they be factored in the decision to select a site for the case study? How will they decide one district is a better candidate for the study over another one when all things are equal? It would be helpful to learn more about the selection process because it is not entirely clear that it is systematic.*

The case study sample will include five sites that are fully operational in spring 2012 and four sites that are in the early phase of implementing a new teacher evaluation system. PPSS has specified that the following criteria will guide selection of the study sample:

- The use of measures of student achievement and measures of teacher effects based on value-added measures or other calculations of gains in the achievement of students
- The capacity to make distinctions among teachers at different levels of performance
- A formative component that provides timely feedback to teachers to help them improve
- Alignment with other parts of the district human capital system
- Requirement for annual evaluations of all teachers
- Data on individual student growth and student growth aggregated to the class and school levels available to teachers and principals
- Use of evaluation data to inform decisions in areas such as professional development, tenure, promotion, and compensation

Sites that are fully operational will include all of the characteristics listed above and completed the first, full evaluation cycle for all teachers in the 2010-2011 school year or earlier; sites that are in the early implementation phase will either have all of these features in place or have concrete plans to implement them in the 2011-12 school year. In addition, the study team will take into consideration diversity of geography, size, poverty level, urbanicity, student demographic characteristics, and state policy climate when selecting sites. Finally, sites must be willing to participate to be included in the study sample.

Based on an extensive and ongoing internet search and discussions with ED staff, members of the project's technical working group, and other experts, we have determined that there are only about 4-6 districts nationwide that have fully operational teacher evaluation systems that meet the selection criteria listed above. Therefore, the selection process for these is straightforward.

The number of sites that are in the early implementation phase is somewhat larger than the fully operational sites, although based on an ongoing review of state education agency and district websites, we have found that a number of these sites are reporting significant implementation challenges or have plans that call for initial implementation of core components of their teacher evaluation systems in 2012-13 or later. Therefore, selecting sites that are in an early implementation phase will require determining that (a) their plans for a teacher evaluation system include all of the components listed above, (b) implementation is likely to occur as scheduled and (c) their plans call for taking all system components to scale in 2012-2013.

Finally, although the implementation and operation of teacher evaluation systems can always face unexpected problems and challenges, a final consideration in selecting sites will be to identify sites that appear relatively stable and which have not faced strong and continuing opposition from teachers and other stakeholders. While ED recognizes that these challenges—and how they are addressed—can provide important insights about the ongoing reforms in teacher evaluation, the primary purpose of this study is to learn about how these systems operate and how they are implemented; focusing on protracted opposition and delays in implementation will detract from achieving the study's primary purpose.

- 2. OMB's question: We have been emphasizing for the past several months the importance for measuring the cost to implement such systems and were told by the Department that this study would be a good mechanism for collecting this information. We understand that it is somewhat difficult to operationalize, but is there any way to collect cost information and include it as a research question? We also would be interested in learning about the time it takes to implement these systems as well.*

First, it is important to note that IES is also currently conducting a study of teacher evaluation systems (a demonstration, impact evaluation) and is planning to collect data on the cost of implementing these systems. IES is currently preparing an OMB clearance addendum package for that impact evaluation that will describe their approach to collecting and analyzing the cost data.

With respect to the PPSS study, we believe that we could collect some information on the staff time and other costs involved in implementing these new types of evaluation systems in our case study sample. Accordingly, the questions below focus primarily on resources (costs and time) associated with annual administration that we believe respondents will be able to estimate. It is important to note, however, that these questions would only be administered to the staff in the fully operational districts.

Additional Research Questions:

- *How long did it take for the district to develop and implement the new teacher evaluation system?*
- *What resources are involved in the annual administration of the new teacher evaluation system (e.g., contracts, staff time) compared with the previous system?*

Below are the additional interview questions that we propose adding to the district-level protocol in order to address the new research questions above. Because some of these questions require respondents to estimate average time commitments for other staff in their district, we expect that respondents may not be able to answer all of them during the interview. The interviewer will review the questions with the respondent during the interview and will collect any data that can be collected at that time. However, the interviewer will also allow respondents to complete and submit the cost and staff time estimates after the interview if they prefer.

Time to Develop New Teacher Evaluation System

When did your state/district begin developing the [name of teacher evaluation system]? To the extent you know, how long did the development process take? What steps were involved in that process, and were there some steps that were the most time consuming to complete?

[If applicable] Once the teacher evaluation system was developed, how many weeks, months, or years were required to fully implement the system? Was the system pilot tested? How long did the pilot testing take to complete?

Ongoing Administrative Resources Spent on Teacher Evaluations

1. **Contracts.** Does the district have any contracts with outside organizations for analysis of teacher evaluation data or for assistance with the administration of the evaluation system? If so, what are the purposes of these contracts and what are the annual costs of these contracts? If the contracts serve purposes beyond those of the teacher evaluation system, what percentage of the contracts' total costs are estimated to be for teacher evaluation work?
2. **New Positions.** Did the district create any new positions that are devoted largely or entirely to the work of the new teacher evaluation system? If so, what are these positions? Were these positions added to the total number of positions in the district, or were existing positions converted into these positions? What is the nature of the work conducted by these staff?
3. **Tests Administered Solely for Teacher Evaluation Purposes.** Does the district administer additional standardized tests for the purposes of measuring teacher effects on student achievement for the teacher evaluation system, or is the district using existing tests (with no additional test administrations)? If additional tests are administered only for the purposes of the teacher evaluation, please explain which tests are administered, to whom and how frequently during the school year.

4. **Staff Involved in a Typical Annual Evaluation for a Teacher.** Please fill in the chart below by indicating the type of staff involved in a teacher’s annual evaluation in the current system and in the previous system. Please estimate the amount of time (in hours or minutes) that each of these staff spends on one typical teacher’s evaluation each year. If you estimate that the time that staff spends varies by school level, please let us know; we would like to get estimates for elementary and for secondary if you believe the estimates would be different. Please note that you or we (the study team) can also ask other respondents (both district-level and/or school-level respondents) for some of these time estimates if you believe that others are in a better position to respond.

Components of Teacher Evaluation Process	Type(s) of Staff Involved* (e.g., principal, assistant principal, instructional coach, external reviewer)		Approximate Amount of Time Per Annual Review for One Teacher (minutes or hours)	
	Current System	Previous System	Current System	Previous System
Developing individual teacher performance plans, goals, criteria, etc.				
Conducting classroom observations				
Reviewing teacher materials such as lesson plans, student work products, etc.				
Reviewing student achievement results (SPECIFY: _____)				
Reviewing other data (SPECIFY: _____)				
Written evaluation summary				
Meeting(s) with teacher to discuss feedback on performance				
School leadership review of draft performance ratings				
Other (specify): _____				

*Note: Please do not include the work/time of the teacher who is being evaluated in this table.

5. **Other District Staff Who Work on the Teacher Evaluation System.** Are there any other district-level staff persons who are not directly involved with an individual teacher’s evaluation and whose time is therefore not captured in the table above? If so, what is the nature of their work on the system and approximately how many hours do they work on the evaluation system each school year? (Please use chart below.)

Activities Involved in Implementing Teacher Evaluation System	Type(s) of Staff Involved (e.g., Statisticians, professional development staff, communications staff, etc.)		Approximate Amount of Time Per Year (hours)	
	Current System	Previous System	Current System	Previous System
Developing and conducting training for staff involved in implementing the teacher evaluation system (e.g., principals, supervisors, performance evaluators)				
Statistically analyzing student achievement data in order to produce teacher effect for evaluation				
Monitoring district-level contracts for work on teacher evaluation system				
Other (specify): _____				

3. *OMB’s question: It seems to us that some of the questions asked in the State protocol, while very good, should also be asked in the district protocol; we could see that some of the decision-making asked about could very well occur at the district-level, especially Questions 5-7 on the State protocol. Was there a specific rationale for not asking about evaluation system design in detail at the district-level?*

We agree that questions 5-7 from the State protocol should be added to the District protocol. Specifically, we propose to add the following, slightly modified versions of Questions 5 and 6 from the State protocol to the District protocol:

[If applicable] Who was involved in determining the relative weight assigned to various factors to be included in the overall rating of teacher effectiveness (e.g., changes in student learning, classroom instruction, professionalism, and community engagement)?

What other issues did the planning group address in developing the teacher evaluation system? [*Follow-up questions for respondents who identify other issues*]: How were these issues resolved? What challenges did the group face in resolving them?

In addition, we agree that Question 7 from the State protocol is a better question and should take the place of the current Question 5 in the District protocol (i.e., “What framework, standards, or

domains of good teaching inform the new teacher evaluation system?”). Specifically, we will add the following, slightly modified version of the Question 7 to the District protocol:

Please describe how and to what extent the teacher evaluation model uses standards and/or a specific model of teaching practice (e.g., Charlotte Danielson’s framework, National Board for Professional Teaching Standards, Classroom Assessment Scoring System, state-developed standards/model) to assess teacher performance, including:

- Whether and how an existing framework/standards has been tailored/modified to address district needs and priorities
- Strategies for communicating about the framework/standards with teachers, school leaders, and other stakeholders
- Required and/or recommended strategies for collecting data on teacher performance in the areas included in the framework/standards (e.g., classroom observations, review of professional portfolios, peer review, student and/or parent surveys)
- The relative weight of measures of professional practice in the overall assessment of individual teachers

4. *OMB’s question: We feel that an important policy question to ask is what approaches districts/states used to ensure that their evaluation systems are valid and reliable and should be included in the protocol(s).*

We expect to capture a significant amount of information regarding the validity and reliability of the components of teacher evaluation systems through comprehensive document reviews. Nevertheless, we will add questions to the protocol that ask about district- and state-level efforts to ensure that the individual components of their teacher evaluation systems provide valid and reliable measures of teacher effectiveness. Specifically, we propose to add the following questions to the state and district protocols:

How, if at all, has the state/district tested the validity of the teacher evaluation system? That is, to what extent has the system’s capacity to validly measure teacher performance been tested?

Follow-up questions for respondents who say they know something about state/district efforts to test the validity of the teacher evaluation system: Has the teacher evaluation system been tested to determine whether teacher ratings are sensitive to external/contextual factors such as class size, curriculum, student demographics, etc.? Are teacher ratings relatively stable over time? Can the scores be manipulated? Who conducted the validity tests of using gains in student achievement and classroom observation data to rate teacher performance and what were the results? Were any adjustments made to the teacher evaluation system based on these results? To what extent are plans in place to regularly test the validity of the teacher evaluation system?

How, if at all, has the state/district assessed the validity and reliability of using classroom observations to rate teacher performance? That is, to what extent has the state/district

determined whether classroom observation ratings are a valid predictor of value-added measures of student achievement? In addition, to what extent has the state/district tested the consistency of the ratings that principals and other observers (internal and/or external) assign teachers based on their observations of teachers' classroom practices?

Follow-up questions for respondents who say they know something about state/district tests of the reliability of using classroom observations to rate teacher performance: Who conducted the tests of reliability of classroom observation data and what were the results? Were any adjustments made to the system (i.e., to the classroom observation instruments/protocols, the types of individuals conducting the observations, the number of observations conducted, or the inter-rater reliability training provided) based on the results of the reliability tests? To what extent are plans in place to regularly test inter-rater reliability ?

5. *OMB's question: Do any of the protocols ask if districts/states are developing other tools or approaches to assess student learning of teachers of subjects and grades that are not tested? If not, can you provide the rationale for not including it? We think it would be an interesting policy to learn more about.*

Yes, both the state and district protocols ask about the tools and approaches districts/states have developed to assess the effectiveness of teachers in untested subjects and grades. Specifically, Question 8 in the State protocol and Question 11 in the District protocol both ask about whether the design of the teacher evaluation system is appropriate for all teachers; for which teachers it is less appropriate and why; and whether any efforts are underway to address the issue. In addition, the study is designed to capture the perspective of teachers in untested subjects and grades by conducting focus groups with these teachers in each of the nine districts. Specifically, focus groups will be conducted with each of the following groups of teachers in every district:

- Special education teachers who may not have standardized student achievement tests as part of their evaluation measures
- Teachers who teach English language learners in language instructional education programs
- Teachers in grades 4 and 5
- Middle school teachers who teach in at least one tested subject and grade
- Middle school teachers who teach non-tested core academic subjects and grades
- High school teachers who teach in the tested grades and subjects
- High school teachers who teach in non-tested academic subjects and grades

6. *OMB's question: For the questions about the purpose of the teacher evaluation system, can a respondent choose more than one response?*

Yes. Indeed, the interview protocols ask open-ended questions that are designed to illicit more than one response.