

## **RESPONSE TO OMB COMMENTS FOR THE STUDY OF THE DISTRIBUTION OF TEACHER EFFECTIVENESS**

### **Comments**

1. Value-added measure - The value-added measure does not control for things like class size, teacher aids, parent aids, etc., so all of those measures would be included in teacher effects. Therefore, some of what we are calling "teacher value-added" might be these other things. And if there are differences in this "teacher value-added" between disadvantaged and non-disadvantaged students, it will be hard to determine if it is due to the teacher distribution or because of these other things.

- Controlling for class size should be relatively easy, and it might be possible to collect data on and control for some of these other things. One added benefit would be that we might be able to get good estimates of the value of these other things. And it might make sense to do some runs that include school fixed effects. (It would be wonderful if there was any way to control for things like the number of classmates with serious disciplinary issues or maybe lots of absences.)

We recognize that value-added estimates measure not only the effectiveness of the teacher but also the combined effect of all factors that affect student achievement in the classroom. The study reports will clearly define the value-added measure used for the study and will describe unobserved factors that might affect the measure. We address the recommendations to control for class size, include school fixed effects, and control for attendance and disciplinary problems below.

**Controlling for classroom-level characteristics.** Including additional classroom-level control variables, like class size, is substantively different than adding additional student-level background characteristics as control variables (note that we discussed this point at a meeting of the technical working group for the study). We plan to estimate the contribution of student-level control variables using within-teacher variation in types of student. This generally produces relatively precise estimates of student characteristics, and does not bias the teacher fixed effect estimates.

We did not propose to estimate classroom-level characteristics like class size for our main regression estimates because classroom-level characteristics can be systematically correlated with teacher effectiveness, so their inclusion can lead to biased estimates of teacher value-added effects if there is a one-to-one correspondence between classroom-level characteristic and teachers. In other words, it is not generally possible to identify the classroom-level characteristic separately from the teacher effect. For example, districts often assign smaller classes to teachers of lower-

performing students. But if lower-performing teachers are also assigned to lower-performing students (which is closely related to the main research question of the study), then including a control variable for class size will in part “control away” for these teacher effectiveness differences, resulting in biased teacher effects.<sup>1</sup>

Traditional fixed-effect value-added models like the one we are proposing to estimate can identify the effect of classroom-level characteristics only by using data on multiple classrooms for individual teachers. This can be accomplished by making use of within-teacher variation in classroom characteristics for teachers with multiple classes. One can exploit variation across multiple sections of a teacher’s courses within one year (primarily for middle school teachers), or by using multiple years of data.

The difficulty with this approach is that there is often little variation in student characteristics like class size across a teacher’s sections within a year or across a teacher’s classes in consecutive years. In such circumstances, estimates of classroom characteristics will typically be imprecise and unstable. However, we plan to investigate these classroom-level characteristics. We plan to check the sensitivity of our results by estimating value-added models that include classroom-level variables representing student peer effects and could also incorporate class size into these analyses.

**Including school fixed effects.** The process of estimating a teacher fixed effects model that also includes school fixed effects would essentially be identical to the teacher fixed effects model we will estimate, with an alternate interpretation. To estimate a model with school fixed effects, we would replace one teacher at each school with a school fixed effect (the omitted teacher for that school) and assign the same school fixed effect to other teachers at the same school. Then the interpretation of the model would change, as we would be measuring each teacher’s effectiveness relative to the omitted teacher in that grade at the same school rather than to all other teachers in the district for that grade. But apart from re-norming each teacher’s value-added to conform to this new interpretation, the estimates obtained from a school-and-teacher fixed effects model and from a teacher fixed effects model would be otherwise identical, so we would not gain new information from this approach.

A model that included both teacher and school fixed effects is not aligned with the study research questions that focus on the distribution of teachers across each district, as we would be measuring teacher effectiveness relative to the average teacher in the school, rather than the average teacher in the

---

<sup>1</sup> The same problem arose in the early class size literature: non-random assignment of small classes to lower-performing students made class size reductions appear to be an ineffective strategy to raise student achievement based on estimates using observational data. The Tennessee STAR experiment changed many researchers’ perception of the efficacy of class size reductions because it was based on random assignment of students to smaller classes.

district. As a result, this model would capture within-school variation in teacher effectiveness, but not between-school variation. Since between-school differences in teacher effectiveness are potentially an important source of inequity for the average teacher effectiveness gap, we think it is important to use a value-added model that captures this aspect of teacher effectiveness. We have proposed to distinguish between within- and between-school sources of teacher effectiveness gaps by comparing the ASEG to the ATEG.

**Controlling for attendance and disciplinary problems.** On the last point about including additional student-level control variables for attendance or disciplinary problems, we plan to estimate models that add additional student control variables in a sensitivity analysis. For the main analysis, we will rely on student background variables that will be available in all districts. Although we are not requiring that districts provide attendance or behavior data because of concerns about placing additional burden on study districts, we could test the sensitivity of the results to including student attendance as a control variable if it is easy for districts to provide these data or if they offer to provide them.

2. Identification issues - It seems prudent to include controls for disadvantaged students, but it makes identification strange. For example, if disadvantaged students were never mixed with non-disadvantaged students, it would be impossible to identify AEG. I think identification is alright if there is sufficient mixing of students, but IES and Mathematica should think carefully about whether this issue places any limitations on their analysis. It might be useful to do some simulation just to make sure that our intuition about identification is right.

If there were complete segregation of students, it would be impossible to identify the coefficient on FRL status because there would be no within-teacher (or within-school) variation in FRL status. Although we could estimate the ASEG and ATEG, we would not be able to distinguish between differences in teacher quality for disadvantaged and non-disadvantaged students that is due to teacher quality compared to differences that are due to other factors correlated with FRL status. However, we think it is very unlikely that districts will have complete segregation of FRL students. We will present information about the distribution of FRL and non-FRL students across schools in each district with the ASEG and ATEG results.

3. One-year value-added measure - I think that the rationale for using one-year rather than multi-year measures of value-added makes sense, but we probably shouldn't completely pre-commit here. It could be that the reduction in variance might be worth the loss in information when moving from one-year to multi-year value-added measures.

We anticipate the additional precision gains for the ATEG and ASEG will not be worth the bias introduced by including multiple years of data. For the AEG, we aggregate value-added measures for teachers into two groups: teachers of FRL students and teachers of non-FRL students. If a school district in our sample has 400 teachers per subject, on average the effective sample size for thinking about the precision gains is thus 200 times the sample size for an individual teacher. The precision gains from using multiple years of data are much smaller when calculating the average value added across 200 teachers than when examining individual value added estimates for teachers. Regarding bias, if we used multi-year value added estimates, one source of bias would be on the estimates of early-career teachers, who tend to improve over the first several years in the classroom, but whose value-added scores would reflect in part their effectiveness from prior years. If we were to include multiple years of data for early-career teachers, we could be systematically underestimating their effectiveness. To the extent that these teachers are systematically matched to disadvantaged students, this bias would be transferred to the AEG.

## Questions

1. Why isn't IES conducting a separate value-added analysis for elementary reading teachers?

We will conduct the value added analysis separately by district, grade, and subject. This means that a separate value added analysis will be conducted for elementary reading teachers **Please see revised Supporting Statement Part A, Section 16 aii, on page 12.**

2. What is the rationale for choosing the top and bottom 20% of the teacher distribution as cut points for identifying effective and ineffective teachers, respectively?

In previous versions of the distribution study OMB package, we described that our plan was to define effective and ineffective teachers as those whose VAM point estimates place them in the top or bottom 20 percent of the distribution. We had planned to use the generalized Index of Dissimilarity to quantify the extent to which the three groups of teachers (effective, average, and ineffective) are unevenly distributed within each district. However, based on advice received at a meeting with our technical working group (TWG) in late May, we revised our plan to examine teacher distribution and now plan to construct an Average Effectiveness Gap (AEG). This is the plan described in our current OMB package. **Please see revised Supporting Statement Part A, Section 16 aiii, on pages 14 and 15.**

We will use the Average Effectiveness Gap (AEG) to measure the distribution. The AEG is a summary measure of the distribution of teacher effectiveness between disadvantaged and non-disadvantaged students, as

defined by eligibility for free or reduced-price lunch. We selected this measure in part because it is not subject to an arbitrary cutoff that defines effective or ineffective teaching, thereby taking advantage of as much variation as possible in the value-added measures of teacher effectiveness. The AEG represents the amount by which the teacher or school quality experienced by non-disadvantaged students differs from the teacher or school quality experienced by disadvantaged students.

For the mobility analysis to be presented in the final report, we plan to group teachers by the top quartile, middle two quartiles, and bottom quartile of value-added measures when presenting descriptive statistics on which teachers have remained in their schools, left to teach in other district schools, or left the district. We are not aware of any empirically-based or policy-supported cut-offs. However, we will conduct sensitivity analyses to make sure that these results are not dependent on this choice of cut-offs. We will also conduct multivariate analyses that do not depend on these or any other cutpoints.

3. In terms of district policies that you intend to include as part of your analysis, will you also try to look at supported policies such as induction and mentoring?

We will gather information on comprehensive induction programs that provide mentoring and support for new teachers. We will also examine district policies that provide targeted professional development to teachers in high-need schools. **Please see the policy types included in Table 3 in revised Supporting Statement Part A (page 12). Please also see the revised District Interview Protocol, Section J Teacher Development, pages 13 and 14.**

Can you clarify how you will quantify district policies for the analysis? Will you try to measure variation in individual policies?

We will describe the strength or intensity of some policies based on their potential to have an effect on the teacher effectiveness gap. We will identify key attributes of each policy and develop a three-level rubric that defines the intensity of policies based on these key attributes. The levels will be defined based on the existing research when possible, and on approaches emphasized by policymakers. As an example, we can rate the intensity level of a bonus for teaching in high-need schools based on (1) the amount of the bonus, (2) the type of bonus (i.e. non-monetary, one-time stipend, or permanent salary increase), and (3) the required teaching commitment to earn the bonus. **Please see revised Supporting Statement Part A, Section 16 ai, on pages 11 through 13. This section discusses our approach to constructing variables for the analysis that will examine the relationship between district policies and the distribution.**