WORKING PAPER

Methods for Accounting for Co-Teaching in Value-Added Models

June 2012

Heinrich Hock
Eric Isenberg

MATHEMATICA
Policy Research

ABSTRACT

Isolating the effect of a given teacher on student achievement (value-added modeling) is complicated when the student is taught the same subject by more than one teacher. We consider three methods, which we call the Partial Credit Method, Teacher Team Method, and Full Roster Method, for estimating teacher effects in the presence of co-teaching. The Partial Credit Method apportions responsibility between teachers according to the fraction of the year a student spent with each. This method, however, has practical problems limiting its usefulness. As alternatives, we propose two methods that can be more stably estimated based on the premise that co-teachers share joint responsibility for the achievement gains of their shared students. The Teacher Team Method uses a single record for each student and a set of variables for each teacher or group of teachers with shared students, whereas the Full Roster Method contains a single variable for each teacher, but multiple records for shared students. We explore the properties of these two alternative methods and then compare the estimates generated using student achievement and teacher roster data from a large urban school district. We find that both methods produce very similar point estimates of teacher value added. However, the Full Roster Method better maintains the links between teachers and students and can be more robustly implemented in practice.

CONTENTS

I. INTRODUCTION

We consider methods to account for student sharing when estimating value-added models of teacher performance. Encouraged in part by the Obama administration's Race to the Top competition for federal funding, value-added models have begun to play a prominent role in teacher assessment systems used by school districts to make high-stakes decisions on performance pay and in some cases even tenure and retention (Steele et al. 2010). A vexing question for policymakers and value-added experts is how to adapt a value-added framework to evaluate teachers who share responsibility for student learning. To help answer this question, we lay out the relative theoretical merits of the statistical methods that can be used to estimate value-added models when there is shared responsibility for students. We then compare the results of two methods using data from a large urban district.

An important prerequisite for implementing a value-added model is accurate data linking students to teachers. As such, districts have begun to use a procedure known as roster validation, in which teachers are asked to confirm whether or not they taught the students assigned to them according to the administrative database of teacher-student links. Conducting roster validation boosts the face validity of value-added measures and increases the accuracy of the estimates (Hock and Isenberg 2010).[1]

Validated roster data provided to us by a large urban school district demonstrated extensive student sharing among teachers. Of the teachers in the tested grades who had at least 15 math students, 21 percent taught students who were also educated in math by another teacher that year. Almost seven percent of teachers shared all their students; six percent shared between 20 and 99 percent of their students, and nine percent shared more than zero and less than 20 percent of their students.[2] In some cases, two or more teachers were jointly responsible for a classroom of students at the same time. In other cases, groups of students were taught by one teacher for part of the year and another teacher for the rest of the year. We refer to both types of cases as "co-teaching."

The prevalence of co-teaching in the data suggests that one might consider adapting the basic fixed-effects approach that has been commonly used in the literature to model teacher-student links.[3] A conservative approach would be to discard students who spent less than 90 percent (or a similar threshold percentage) of the school year with a particular teacher (Steele et al. 2010) and continue to estimate a fixed-effects model. This method, however, would create an incentive for teachers to shift attention away from students who are shared with another teacher at the school or who transfer from another school. Instead, we consider how the basic value-added model may be adapted to account for teacher "dosage"—the fraction of a student's instructional time for which the teacher is responsible. We discuss three specific approaches that make use of dosage when

[1] We abstract in this paper from a broader consideration of the validity and utility of value-added models for policy decisions, both of which have been extensively discussed in the literature (Kane and Staiger 2002; McCaffrey et al. 2004; Rothstein 2010; Baker et al. 2010; Glazerman et al. 2010; Koedel and Betts 2011; Chetty et al. 2011; and Goldhaber and Chaplin 2012).

[2] Percentages cited are for math but were similar for reading.

[3] McCaffrey et al. (2004) describes some possible theoretical approaches for modeling student sharing.

modeling co-teaching: (1) the Partial Credit Method, (2) the Teacher Team Method, and (3) the Full Roster Method.

The Partial Credit Method modifies the fixed-effects approach to value added by allowing teacher variables to take continuous dosage values. This method presents problems when teachers claim many students in common but teach a few students individually—a common occurrence in the district we studied. In these cases, near-collinearities among the teacher measures result in statistically unstable coefficient estimates. For this reason, we do not consider this approach to be a viable one for estimating teacher value added in our study district.

Under the Teacher Team Method, we add team variables to the model to capture the joint effect of two or more teachers on student achievement. This method reduces the collinearities inherent in the Partial Credit Method. In its purest form, the Teacher Team Method does not make it possible to estimate the distinct contribution that each teacher makes to the team because teams are jointly assigned a dosage of one. However, within the Teacher Team approach, it is possible to use a Partial Credit technique in some circumstances. For example, one could use Partial Credit to capture the individual contributions of teachers at different schools on the achievement of a student who switches schools mid-year, while forming team variables for teachers who share students within the same school.

With the Full Roster Method, instead of adding variables for teacher teams, student records are replicated so that each teacher-student combination contributes one dosage-weighted observation to the regression. As with the Teacher Team Method, it is not possible to disentangle the individual effects of co-teachers on their shared students. Although the Full Roster Method is not as adaptable as the Teacher Team Method—there is no capacity to accommodate Partial Credit techniques—it is more robust to a wide variety of teaming arrangements. It also enables more links between teachers and students to be maintained than does the Teacher Team Method.

The rest of this paper proceeds as follows. In Section II, we define the three methods in general terms, discuss why the Partial Credit Method is not feasible for estimating teacher value added, and demonstrate the conditions in which the Teacher Team and Full Roster Methods yield identical point estimates. We relax those assumptions to show how and why these methods may differ. A more detailed derivation is given in the appendix. Section III describes the choices we made when implementing these two models with data provided by the school district. Section IV demonstrates that the point estimates of the value-added measures for teachers were nearly identical across these two methods. Section V concludes by discussing the practical advantages of the two methods in the context of producing value-added measures for high-stakes teacher assessments.

II. METHODS CONSIDERED

A. Partial Credit Method

Of the three methods we considered, the Partial Credit Method was the least effective for accounting for co-teaching in the school district. This method modifies a basic fixed-effects value-added model by allowing teacher variables to take continuous dosage values ranging from zero to one, rather than discrete values of zero or one. The estimating equation in this approach is

$$(1) \quad y_i = \boldsymbol{\pi}'\mathbf{z}_i + \boldsymbol{\psi}'\mathbf{w}_i + \varepsilon_i ,$$

where y_i is the gain score for student i and the vector $\mathbf{z}_i$ represents student-level covariates.[4] The $(K \times 1)$ vector $\mathbf{w}_i$ measures the student's dosage weight attributed to each of K teachers in the school district, with the kth element, w_{ik}, denoting the proportion of the school year that a student i spent with teacher k. Students who were co-taught or moved between teachers have multiple non-zero entries in $\mathbf{w}_i$. Each teacher weight is then the percentage of the year that the student spent in the particular teacher's classroom. For example, if a student split time equally between two teachers, each teacher would receive a 0.5 dosage for that student. Because teacher effects are additive and linear, the Partial Credit Method implicitly assumes that one unit of a teacher's time has a constant effect on each student, regardless of the fraction of the year that the student was in the classroom, and that there are no interactions between teachers. Thus, the model embodies the notion that each teacher is *individually responsible* for a distinct component of the achievement gains of shared students.

Although the Partial Credit Method theoretically allows one to estimate distinct effects for all K teachers in the system, this method did not yield reliable estimates for teachers in the school district we studied. We have successfully used the Partial Credit Method when calculating *school* value added in many districts—the method works well in this application because relatively few students switch between any pair of schools. The method does not work as well when calculating teacher value added because teachers tended to share blocks of students in the district we studied. This form of co-teaching resulted in a fairly strong correlation between the dosage variables for teachers who shared students. As a result, an outlier estimate for one teacher (which could arise solely by chance) would skew estimates for others who co-taught students with that teacher. More problematic is that some teachers in the study district shared almost all their students, teaching only a few individually. In such cases, teacher estimates were essentially identified from the changes in achievement of a small number of solo-taught students, resulting in unstable coefficients. We therefore do not consider this approach a viable option.

B. Teacher Team Method

The Teacher Team Method is an alternative that involves modifying the column space to add team variables when teachers' students overlap. If two teachers share all their students, the teacher

[4] For ease of exposition and notation, we ignore differentiation by grade and treat the dependent variable as a simple gain score. In practice, our approach was to adjust the gain score by including a grade-specific decay parameter estimated in a preliminary errors-in-variables regression. We also assume throughout that each z_i has been mean-centered.

team variable replaces the variables corresponding to the individual teachers. The estimating equation for the Teacher Team Method is:

(2) $y_i = \boldsymbol{\pi}'\mathbf{z}_i + \boldsymbol{\gamma}'\mathbf{c}_i + \xi_i$.

Each element of the ($M \times 1$) vector $\mathbf{c}_i$ corresponds to a teacher team or to an individual teacher. Each team variable represents the joint effect of two or more teachers on student growth, and can be thought of as an interaction term in the regression (which implies that $M > K$). Once teams are specified, teacher and team coefficients may be estimated by applying ordinary least squares (OLS) to equation (2).

In the purest form of the Teacher Team Method, either a team or an individual is wholly responsible for all of the student's instructional time during the school year. Hence, a single element of $\mathbf{c}_i$ is equal to one, and the rest are zero. In other words, team teachers are considered to be *jointly responsible* for the achievement of their students. With this modeling strategy, it is not possible to recover the distinct contribution that individual teachers make to the team.

Unlike the Partial Credit Method, the coefficient estimates are not vulnerable to instability due to teaming arrangements. Because the teacher and team variables are binary indicators, the estimated effects of teachers and teams are only indirectly correlated with one another due to commonalities in student characteristics. This insulates teachers and teams from one another, resulting in more stable coefficient estimates.

The Teacher Team Method may, however, be adapted so that team variables are only formed for selected groups of teachers. For example, teams may be limited to teachers within the same school, with the Partial Credit Method used to model the cumulative effect of teachers of students who switch schools mid-year. Alternatively, team variables can be formed only for teachers who have some minimum number of students in common, with Partial Credit applied to teachers sharing fewer students.

Granularity of Teams. Teacher teams may be defined with varying levels of detail. *Fully interacted* teams represent unique combinations of (1) a set of teachers with shared students and (2) a dosage contribution for each teacher. Thus, if Teacher A taught a group of students for one-half of the year and Teacher B taught them for the other half, this would constitute a different fully interacted team than one in which Teacher A taught students for one-third of the year and Teacher B taught them for the remaining two-thirds. *Aggregated* teams, on the other hand, simply represent unique sets of co-teachers. Thus the two fully interacted teams in the previous example would compose a single aggregated team. This approach implicitly assumes that team effects do not vary according to the mix of dosage contributions from each teacher.

Combining Effects for Teachers with Multiple Estimates. We create a single "overall" estimate of effectiveness for teachers who receive multiple estimates because they belong to multiple teams or because they both belong to a team and teach students individually. This overall teacher effect, $\hat{\omega}_k$, is the weighted average of the relevant regression coefficients using student-equivalents as weights. Specifically:

(3) $\hat{\omega}_k = \sum_{m \in M_k} p_{km} \times \hat{\gamma}_m$,

where M_k denotes the set of teams to which teacher *k* belongs (including the solo "team" consisting of just teacher *k*), p_{km} is the fraction of teacher *k*'s student-equivalents who are taught as part of teacher team *m*, and $\hat{\gamma}_m$ is the estimated combined achievement effect of teacher group *m*. For example, if Teacher A individually taught 15 students and shared another 10 students equally with Teacher B, the regression would produce a Teacher A-alone effect and a Team-AB effect. Assuming both teachers receive half credit for each shared student, the total number of student-equivalents from this group for Teacher A would be 0.5 × 10 = 5. As a result, the teacher's overall effectiveness measure would be estimated as 15/20 × [Teacher-A-alone estimate] + 5/20 × [Team AB estimate].

C. Full Roster Method

The Full Roster Method accounts for co-teaching by expanding the row space but not changing the column space of the data. This method does not require all possible teams of teachers to be characterized based on the data. Instead, the Full Roster Method produces one regression estimate for each teacher, which represents an overall estimate of effectiveness. The regression is based on a series of variables for each of the *K* teachers (but no team variables). Student records are replicated so that each teacher-student combination is a unique observation. Thus, each teacher is associated with a set of records that covers all of his or her students, regardless of whether or not they are taught by other teachers.

The estimating equation for the Full Roster Method is:

$$(4) \quad y_{ik} = \boldsymbol{\eta}'\mathbf{z}_i + \boldsymbol{\beta}'\mathbf{t}_{ik} + \zeta_{ik},$$

where *k* denotes a teacher to whom student *i* is linked. Each replicated observation has only one teacher dummy variable set to one, and is weighted according to the fraction of the year the student spent with that teacher. Thus, the elements of the (*K* × 1) vector of teacher indicators, $\mathbf{t}_{ik}$, are uniformly zero except for the *k*th element (i.e., $t_{ikk} = 1$ and $t_{ikj} = 0 \;\forall\, j \neq k$). Single effects for each teacher, $\hat{\beta}_k$, are estimated by applying weighted least squares (WLS) to equation (4), with the weights equal to the dosages w_{ik} as defined above. In the example above, interaction variables are not created for the students whom Teachers A and B shared. Instead, those student records would be duplicated in the regression—once for each teacher. One observation would have an entry of zero in the column for one teacher and an entry of one for the other. In the next row, the zero and one would be reversed. Each duplicated observation would receive a weight of 0.5.

The Full Roster Method is similar to the Teacher Team Method in that the teacher variables included in the regression are uncorrelated, which reduces the potential for an outlier estimate for one teacher to skew the estimates of other teachers. In fact, as shown in the next section, the Full Roster Method produces results that are, under certain circumstances, identical to estimates from the fully interacted Teacher Team Method. Thus, the Full Roster Method also embodies the notion that teachers are jointly responsibly for the achievement of their shared students. Unlike the Teacher Team Method, there is not a straightforward modification of the Full Roster Method that would allow it to incorporate Partial Credit for subgroups of teachers.

D. Numerical Comparison of Estimates from the Teacher Team and Full Roster Methods

The teacher regression coefficients obtained using the Full Roster Method ($\hat{\beta}_k$) and the overall effectiveness estimates based on the Teacher Team Method ($\hat{\omega}_k$) are identical under certain conditions. As demonstrated in the appendix, $\hat{\beta}_k = \hat{\omega}_k$ if (1) the Teacher Team Method is implemented using fully interacted teams, and (2) student covariates are not included in the model. Thus, in the example above, the Full Roster Method would produce a regression coefficient for Teacher A that is precisely equal to 15/20 × [Teacher-A-alone estimate] + 5/20 × [Team AB estimate].

However, the two sets of estimates will generally differ if (1) aggregated teams are specified or (2) covariates are included in the regression. If aggregated teams are specified, the differences occur due to weighting. The Full Roster Method always implicitly weights multiple estimates as if they were produced from a series of fully interacted teams, but the Teacher Team Method would in this case explicitly weight the effects of a smaller number of aggregated teams. When covariates are included, the estimated coefficients of the covariates differ across the two methods. This occurs because the two regression equations include different sets of intercepts. These differences in covariate coefficients, in turn, result in different estimates of teacher effects. See the appendix for details.

III. IMPLEMENTATION OF METHODS

In practice, one must make a series of decisions to implement any value-added method to account for co-teaching. In this section, we first review the roster validation procedure that was used by the study school district, an important precursor to understanding our implementation decisions. We then explain the decisions we made for each method.

Unlike the data available in many longitudinal administrative databases, information on teacher-student links in the study school district was based on a roster validation process covering each instructional term. Teachers validated the presence of students indicated on their subject-specific administrative rosters, adding students if necessary. Teachers could claim responsibility for each of their students for the entire term, or for a fraction of the term.[5] Intermediate values were intended mainly for students in pullout programs taught by teachers who were not eligible for a value-added estimate, typically special education teachers. These percentages determined the fractional dosage attributed to each teacher. In cases in which the total percentage of instructional time added up to more than 100 percent, we rescaled the dosages so that they summed to one. For students who were claimed by one teacher for part of a term and not claimed for the remainder of the time by another teacher in the same subject, we attributed the student's remaining dosage to a catchall "unspecified teacher" for each grade level. That is, for each grade we estimated an extra teacher variable in addition to the variables representing teachers and teacher teams.

A. Implementation of the Teacher Team Method

Implementing the Teacher Team Method involves making the follow six key decisions:

- Setting the minimum number of students needed for an individual teacher to be included in the value-added model
- Setting the minimum number of students for a teacher team to be included in the model
- Determining other criteria necessary for forming a team
- Deciding how to account for students in teams that do not have a sufficient total number of students to receive an estimate
- Establishing whether to define fully interacted teams or aggregated teams
- Setting the relative weights on multiple estimates for teachers who taught students individually and as part of a team

Minimum Student Counts. For the first two decisions, we chose seven as the minimum number of students for both individual and team estimates. In preliminary work in which we analyzed a variety of approaches, it became apparent that these two thresholds needed to be identical. Otherwise, the number of special cases in the formation of individual and team variables could become unmanageable. In deciding on a threshold value, we had to balance competing

[5] Although the roster validation process was necessary to establish credible teacher-student links, it was also imperfect. Some cases in which students were double-claimed or left unclaimed likely resulted from mistakes made by teachers during the roster validation process. Although a rigorous analysis has not been conducted, there is no evidence of teachers consciously trying to game the system.

objectives. On one hand, we aimed to incorporate the greatest amount of information in the value-added model, including as many valid teacher-student links as possible. On the other hand, we were attentive to data errors in the roster validation process: a math teacher mistakenly claiming students in reading could appear in the data as a reading teacher, and vice versa. A minimum threshold helps to eradicate potentially errant teacher-student links. In balancing these objectives, we decided to create variables only if teachers or teams could be linked to at least seven students.

Other Criteria for Forming Teams. We limited teams to teachers who shared students within the same school. For students shared by teachers across schools (four percent of students), we used the Partial Credit Method, believing that this would more accurately reflect the underlying production function. We foresaw little risk of unstable estimates for teachers in different schools, since it would be very unlikely that they shared many students. We formed teams not only between teachers participating in the roster validation process, but also between those teachers and the catchall "unspecified teacher" when there were at least seven students in common. Without this step, we found that some of the value-added estimates for teachers diverged sharply from the average covariate-adjusted gain score of their students.[6] This discrepancy occurred because estimated effects for the unspecified teacher were negative at all grade levels in both subjects, often substantially so. As a result, estimates for teachers who shared students with an unspecified teacher tended to be higher when this student sharing was modeled using the Partial Credit Method. To eliminate this "windfall" to teachers with many pullout students, we formed teams when seven or more students were shared, which brought the value-added estimates closer to the average covariate-adjusted gain score. However, this approach did not eliminate the possibility of a benefit to teachers sharing fewer than seven students with an unspecified teacher.

Teams with Few Students. We also used the Partial Credit Method for cases in which students were taught by multiple teachers who had fewer than seven students in common. For two-person teams, these students were distributed back to the individual teachers, assuming that they had at least seven solo-taught students. For "broken teams" composed of three or more members, we devised a series of rules to distribute these students among two-person teams or individual teachers. In implementing these rules, we manually checked the teaming arrangements to ensure that the students were distributed as intended because it was not possible to program the full set of rules to cover all possible cases. In some cases, teachers received estimates that were informed by some, but not all, of their students. For example, if a two-person team formed because there were seven or more shared students, but an individual estimate for one of the teachers in the team did not form because there were not at least seven solo-taught students, these students were delinked from the teacher and allocated to the unspecified teacher.[7]

Granularity of Teams. To avoid dropping students, we opted to use aggregated teams, rather than fully interacted teams. An attempt to create multiple teams among the same set of teachers

[6] The average covariate-adjusted gain for a teacher is calculated by (1) estimating a value-added regression; (2) calculating each student's adjusted gain as the difference between the actual posttest and the posttest that is predicted from the student's pretest and background characteristics (but not teacher and team variables); and (3) taking a dosage-weighted average of the adjusted gains of the teacher's students.

[7] Chiang and Chaplin (2010) developed a method for addressing co-teaching in a network of public charter schools using an algorithm that determines when to create teacher teams. Their method forms teams only for teachers who have sole responsibility for fewer than 10 students, while our method creates every possible team, as long as the teachers share at least 7 students.

based on different relative student dosage weights would have resulted in more teams with fewer than seven students. These teams would "break," resulting in additional students becoming delinked from their teachers.

B. Implementation of the Full Roster Method

The Full Roster Method requires only one implementation decision: establishing the minimum number of students for an individual teacher to be included in the value-added model. For the purpose of comparing results generated from the Full Roster Method with those generated from the Teacher Team Method, we set this minimum threshold to seven students, the same number used for the Teacher Team Method.

IV. OVERVIEW OF VALUE–ADDED ESTIMATION

Estimates of teacher value added were based on one year of teacher-student links data covering grades 4 through 8 in the study school district. Student posttests were based on end-of-year results from the state-wide assessment test, and pretest data were based on their results at the end of the prior school year. These test scores were both standardized in each subject so that they were on a common scale across grades within the district.

Initial value-added estimates were calculated using regressions like equation (2) for the Teacher Team Method and equation (4) for the Full Roster Method based on a two-step regression process. All regressions included teacher variables and team variables (if applicable), as well as student characteristics available from the school district's administrative database. In the first step, we estimated a grade-specific value-added regression to obtain the decay parameter on the pretest in each grade using an errors-in-variables technique (Buonaccorsi 2010). Grade-specific test-retest reliability information was based on the technical manual obtained from the test publisher. We then calculated a pretest-adjusted gain score by netting out the estimated contribution from the pretest. In the second step, the pretest-adjusted gain score was used as the dependent variable in a value-added regression that pooled data from all grades.[8] To account for heteroskedasticity, we calculated robust standard errors using the Huber-White sandwich estimator. To address the replication of student observations when applying the Full Roster Method, we also allowed standard errors to be clustered at the student level.

We calculated "final" teacher estimates by applying the following process to the initial regression estimates obtained from both the Teacher Team Method and the Full Roster Method:

1. We adjusted the coefficient estimates so that the mean and standard deviation of effects in each grade were identical. This adjustment was intended to account for any differences in the alignment of the pretest and posttest across grades.
2. For teachers with both solo-taught and co-taught students, we combined multiple grade-specific estimates by taking a dosage-weighted average within the grade, as explained in Section II. (This step was only needed when applying the Teacher Team Method.)
3. For teachers with students in more than one grade, estimates were combined using a dosage weighted average across grades.
4. We applied an empirical Bayes shrinkage procedure based on Morris (1983) to offset the possibility that teachers with few students will be distributed in the tails of the distribution due to chance.
5. To reduce the prevalence of imprecise estimates, we excluded teachers who taught fewer than 15 students, which we call the "reporting threshold" for teachers.

[8] The preliminary set of grade-specific regressions was necessary because it was not numerically possible to apply the errors-in-variables technique to all grades simultaneously—the error-adjusted matrix of regressors was singular.

V. EMPIRICAL COMPARISON

A major difference between the methods we examined was that the Teacher Team Method maintained fewer links between teachers and their students due to the limitations of the team-forming rules. When considering teachers who met the reporting threshold of 15 students (before any reallocation to the unspecified teacher), the Full Roster Method resulted in 5.3 percent of math teachers and 3.8 percent of reading teachers being matched to additional students. Among those teachers, the average increase was 4.5 students in math and 3.9 students in reading. In addition, of the teachers with 15 or more students one more teacher received an estimate based on the Full Roster Method.[9]

The increase in the number of teacher-student links naturally resulted in differences in the value-added estimates produced by the two methods. In order to better evaluate differences based solely on the methodology, we made two adjustments. First, we re-estimated the value-added regression using the Full Roster Method with the teacher-student links limited to those that would have been maintained using the Teacher Team Method. Second, we re-standardized the post-shrinkage estimates obtained from the two methods to a mean of zero and a standard deviation of one among the teachers who met the reporting threshold and received estimates based on both methods.

As shown in Figures 1 and 2, the two methods produced value-added estimates that are very similar for both math and reading. Both figures display a tight fit, with most observations concentrated near the 45-degree line. The correlations between the two sets of estimates are 0.995 in math and 0.994 in reading. However, some observations are noticeably away from the diagonal.

Based on the discussion in Section II, one of the sources of differences between the two sets of estimates could be a difference in the estimated coefficients on the covariates. To explore this possibility, we ran the following regression for each subject:

$$(5) \quad DVA_k = \boldsymbol{\lambda}'\overline{\mathbf{z}}_k^w + \varepsilon_k ,$$

where DVA_k is the difference between the teacher *k*'s value-added estimate on the Teacher Team Method and the value-added estimate based on the Full Roster Method. The vector $\overline{\mathbf{z}}_k^w$ denotes dosage-weighted averages of the student covariates included in the value-added regression. Although a few of the estimated coefficients were significant, there did not appear to be a systematic relationship between students commonly associated with lower achievement gains and the sign of the difference in value-added obtained from the two methods.[10]

---

[9] The team-forming rules are such that a teacher who has enough students to clear the threshold might not receive an estimate based on the Teacher Team Method. This could occur if, for example, all of a teacher's students were shared with other teachers in three teams of five students each, in which case all the students would be reallocated to the unspecified teacher.

[10] At the request of the participating local education agency, we do not describe the specifics of these characteristics.

Figure 1. Scatter Plot of Standardized Value–Added Estimates for Teachers of Math

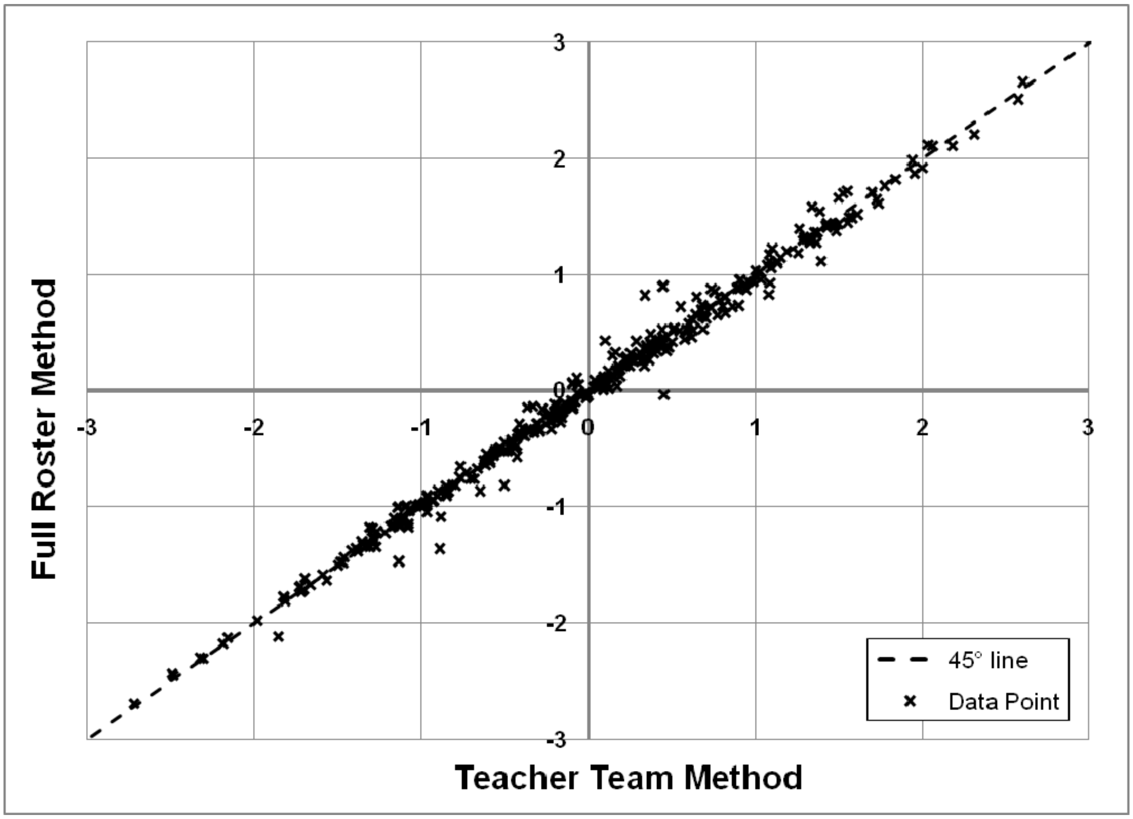
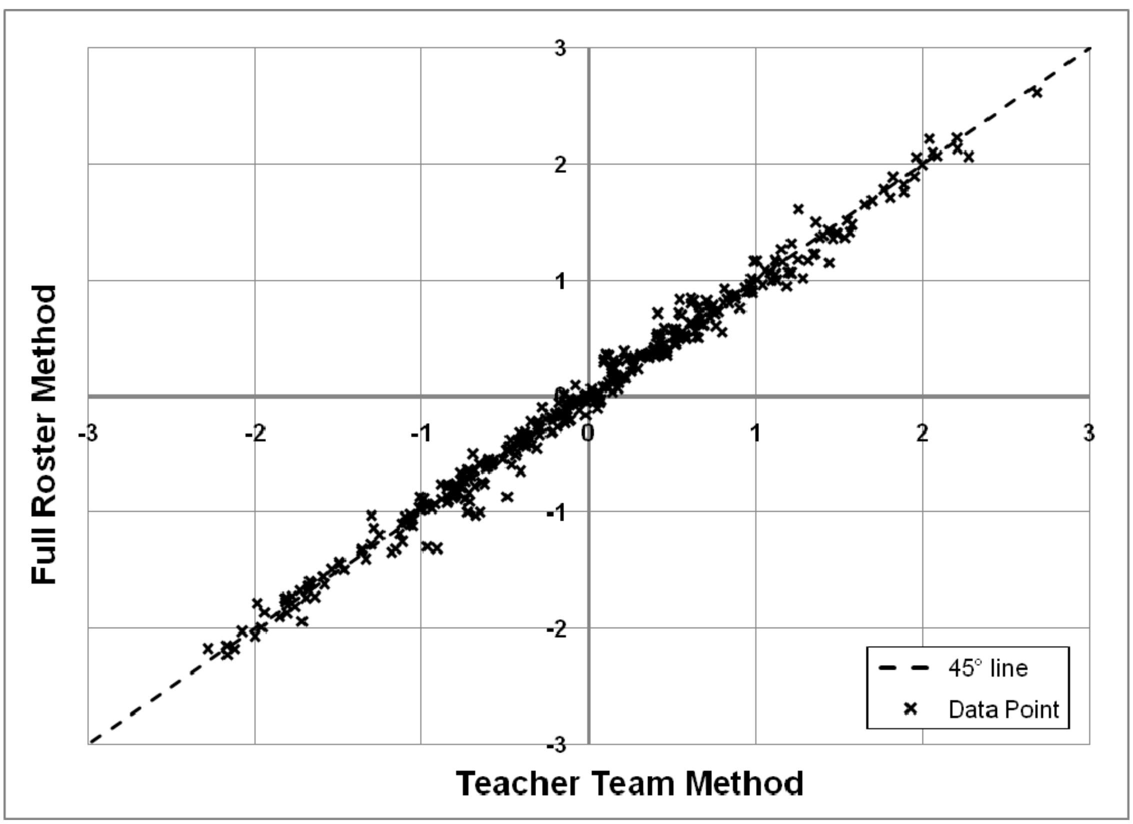


Figure 2. Scatter Plot of Standardized Value–Added Estimates for Teachers of Reading



VI. CONCLUSION

We have shown that the two methods of modeling co-teaching—the Teacher Team Method and the Full Roster Method—are theoretically related and empirically similar. Both models implicitly assume that teachers have an interactive effect on student achievement. From a policy perspective, this implies that co-teachers are held jointly responsible for the outcomes of the shared students. By contrast, the Partial Credit Method assumes that teachers have distinct and separable effects on student outcomes and is premised on the idea of individual responsibility. We take no stance on whether joint responsibility or individual responsibility is preferable, noting only that it is not empirically tenable to fully implement the Partial Credit Method when teachers have many or all of their students in common.

One advantage of the Teacher Team Method over the Full Roster Method is that it is more adaptable. Specifically, it allows for a hybrid approach in which the Partial Credit Method is used to model individual responsibility in some cases. In our implementation, we applied the Partial Credit Method to model student sharing among teachers at different schools and those with fewer than seven students in common. In principle, it is possible to expand the extent to which Partial Credit is used, but this increases the likelihood of statistically unreliable estimates arising from collinearity.

The Full Roster Method has two key advantages for a policymaker who prefers, or is at least comfortable with, the assumption of joint responsibility. First, it allows for a greater number of students to contribute directly to the calculation of a value-added score for a teacher because it does not rely on a complex mechanism for forming team variables, as the Teacher Team Method does. Increasing the sample size for a given teacher will, in turn, increase the accuracy and precision of the value-added estimate.[11] It also adds to the face validity of value-added methods among teachers to know that all of the students they claim during a roster validation process directly contribute to their performance score.

The second advantage of the Full Roster Method is that it relies on a simpler, more transparent set of rules, which may make it better suited to high-stakes applications. The team variables used in the Teacher Team Method are formed based on counts of students claimed by each combination of teachers. It is not possible to pre-specify an algorithm that covers every possible scenario of "broken" teams—that is, combinations of teachers sharing fewer students than the minimum number needed to reliably estimate a team effect. Consequently, it may be necessary to adapt the algorithm on the fly to account for special cases. The Full Roster Method relies on a much simpler set of rules that avoids the concern about special cases. As such, this method may be preferable for policymakers who wish to provide a clearer *a priori* explanation of which students contribute to a teacher's value-added score and how they do so.

Based on our data from a large urban school district, the two methods produced very similar estimates of teacher value added. Given the high correlation in the results, both the Full Roster Method and the Teacher Team Method appear to provide valid estimates of teacher value added under the assumption of joint responsibility. Consequently, we recommend basing decisions about

[11] Even if the hybrid features of the Teacher Team Method are assumed to embody the "true" model of education production or responsibility for shared students, linking more students to their teachers using the Full Roster Method may produce an estimate with a lower mean squared error.

which method to use on considerations related to modeling responsibility for shared students, computational robustness, and face validity.

APPENDIX

A. Comparing Estimates from Teacher Team and Full Roster Methods

The teacher regression coefficients, $\hat{\beta}_k$, obtained from the Full Roster Method (FRM) potentially represent alternative estimates of the overall effectiveness measures obtained from the Teacher Team Method (TTM), $\tilde{\hat{\omega}}_k$ described by equation (3) in the main text. We demonstrate that these two estimates are numerically identical if the TTM is specified using fully interacted teams and if student covariates are not included in the model. We then show that the FRM estimates will generally differ from the TTM estimates if aggregated teams are specified or if covariates are included in the value-added models. For notational purposes, we will use an umlaut to denote elements of the fully interacted TTM (e.g., $\ddot{\mathbf{c}}_i$ and $\ddot{M}_k$) and a tilde for elements of the aggregated TTM (e.g., $\tilde{\mathbf{c}}_i$ and $\tilde{M}_k$).

1. No Covariates and Fully Interacted Teams

Without any covariates, the stacked regression equation used for the fully interacted TTM is

$$(6) \quad \mathbf{y} = \ddot{\mathbf{C}}\boldsymbol{\gamma} + \boldsymbol{\xi},$$

where $\mathbf{y} = [y_1, \ldots, y_N]$ is the vector of gain scores for each of the N students and $\ddot{\mathbf{C}} = [\ddot{\mathbf{c}}_1', \ldots, \ddot{\mathbf{c}}_N']$ is the matrix of links between students and teachers/teams. Applying OLS yields

$$(7) \quad \hat{\gamma}_m = \frac{1}{N_m}\sum_{\ddot{c}_{im}=1} y_i \equiv \bar{y}_m .$$

Hence, the effect for team m is simply the average achievement gain of students taught by the teacher team. Equations (3) and (7) imply that

$$(8) \quad \hat{\ddot{\omega}}_k = \sum_{m \in \ddot{M}_k}\left\{\frac{N_{km}}{N_k} \times \bar{y}_m\right\},$$

where N_k represents the total number of student-equivalents for whom teacher k has responsibility and N_{km} is the number of those students taught as part of teacher team m. Hence, teacher k's overall estimate is a student-equivalent-weighted average of the mean gain scores of students in all teams of which k was a member.

The stacked regression equation for the FRM is:

$$(9) \quad \underline{\mathbf{y}} = \mathbf{T}\boldsymbol{\beta} + \boldsymbol{\zeta}.$$

We will assume that the data are ordered so that student records associated with a teacher are contiguous. The vector $\underline{\mathbf{y}} = [\mathbf{y}_1, \ldots, \mathbf{y}_K]$ represents a series of ($N_k \times 1$) subvectors $\mathbf{y}$, the elements of which correspond to the gain scores of the students assigned to teacher k. The matrix $\mathbf{T}$ is partitioned similarly: $\mathbf{T} = [\mathbf{T}_1, \ldots, \mathbf{T}_K]$, where $\mathbf{T}_k$ is an ($N_k \times K$) matrix of stacked teacher match variables, $\mathbf{t}_{ik}$. Applying weighted least squares (WLS) to equation (9) using the w_{ik} as weights results in

$$(10)\quad \hat{\beta}_k = \frac{1}{N_k}\sum_{t_{ikk}=1} w_{ik} y_{ik} \ ,$$

where $t_{ikk} = 1$ indicates students matched to teacher *k*. Hence, teacher *k*'s regression estimate is the dosage-weighted average of the gain scores of all the students that *k* taught.

Subdividing teacher *k*'s students according to the set of fully interacted teams of which *k* was a member, $\ddot{M}_k$ we can rewrite equation (10) as

$$\hat{\beta}_k = \frac{1}{N_k}\sum_{m\in\ddot{M}_k}\sum_{\ddot{c}_{im}=1} w_{ik} y_{ik} \ ,$$

where $\ddot{c}_{im}$ is the *m*th element of $\ddot{\mathbf{c}}_i$ and indicates whether student *i* was taught by team *m*. With fully interacted teaming, each team *m* is unique according to the dosage contributed by the constituent teachers. This implies that w_{ik} is constant across students taught by team *m* and equal to N_{km} / N_m. Additionally, actual gains do not vary across replicated student observations, and so $y_{ik} = y_i$. As a result,

$$(11)\quad \hat{\beta}_k = \frac{1}{N_k}\sum_{m\in\ddot{M}_k}\left\{\sum_{\ddot{\mathbf{c}}_{i[m]}=1}\frac{N_{km}}{N_m} y_i\right\} = \sum_{m\in\ddot{M}_k}\left\{\frac{N_{km}}{N_k}\times\bar{y}_m\right\},$$

which is precisely the estimated overall teacher effect obtained in equation (8) using the fully interacted TTM.

2. No Covariates and Aggregated Teams

Using the same logic as above, the estimated overall teacher effect for teacher *k* based on the aggregated TTM is

$$(12)\quad \hat{\tilde{\omega}}_k = \sum_{m\in\tilde{M}_k}\left\{\frac{N_{km}}{N_k}\times\bar{y}_m\right\}.$$

This estimate differs from that in equation (8) because it specifies a smaller set of aggregated teams ($\tilde{M}_k$). However, the estimated teacher effect from the FRM remains defined over the set of fully interacted teams ($\ddot{M}_k$), as in equation (11). As such, the weighted averages will differ and the overall teacher effect in equation (12) will differ from $\hat{\beta}_k$.

If practical limitations necessitate forming aggregated teams when implementing the TTM, but fully interacted teams are believed to better reflect the underlying data-generating process, then the estimate in equation (12) will be incorrect. However, the FRM regression coefficient will correctly estimate teacher effects.

3. Student Covariates Included in the Regression

Adding covariates, the stacked equation for the fully interacted TTM is

(13) $\mathbf{y} = \mathbf{Z}\boldsymbol{\pi} + \ddot{\mathbf{C}}\boldsymbol{\gamma} + \boldsymbol{\xi}$,

where $\mathbf{y}$ and $\ddot{\mathbf{C}}$ are defined as above and the $(N \times L)$ matrix $\mathbf{Z} = [\mathbf{z}'_1, \ldots, \mathbf{z}'_N]$. The stacked equation for the FRM can be represented as

(14) $\underline{\mathbf{y}} = \underline{\mathbf{Z}}\boldsymbol{\eta} + \mathbf{T}\boldsymbol{\beta} + \boldsymbol{\zeta}$,

where $\underline{\mathbf{y}}$ and $\mathbf{T}$ are as above. The matrix $\underline{\mathbf{Z}} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_J]$ denotes a stacked set of *K* matrices, with the rows of the $(R_k \times L)$ matrix $\mathbf{Z}_j$ corresponding to vectors of covariate for teacher *k*'s students.

Standard regression algebra, as in Goldberger (1991), provides the following results:

(15) $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}^s - \left(\left[\ddot{\mathbf{C}}'\ddot{\mathbf{C}}\right]^{-1}\ddot{\mathbf{C}}'\mathbf{Z}\right)\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^s - \left(\left[\mathbf{T}'\mathbf{W}\mathbf{T}\right]^{-1}\mathbf{T}'\mathbf{W}\mathbf{Z}\right)\hat{\boldsymbol{\eta}}$,

where $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\pi}})$ and $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\eta}})$ are the coefficients from the "long" regressions in equations (13) and (14). The estimates $\hat{\boldsymbol{\gamma}}^s$ and $\hat{\boldsymbol{\beta}}^s$ are the coefficients from the "short" regressions that ignore student covariates, i.e., equations (6) and (9). Finally, $\mathbf{W}$ is an $(R \times R)$ diagonal matrix of dosage weights. It is straightforward to show that

(16) $\hat{\gamma}_m = \overline{y}_m - \hat{\boldsymbol{\pi}}'\overline{\mathbf{z}}_m$ and $\hat{\beta}_k = \overline{y}_k^w - \hat{\boldsymbol{\eta}}'\overline{\mathbf{z}}_k^w$.

To simplify the notation, $\overline{y}_k^w = (N_k)^{-1}\sum_{t_{ikk}=1} w_{ik} y_i$ represents the weighted average of gain scores among the students of teacher *k* and $\overline{\mathbf{z}}_k^w$ and may be interpreted similarly.

Based on equation (16) and the derivation above, the FRM and the fully interacted TTM will yield identical point estimates of overall teacher effects if and only if $\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\eta}}$. That is, the two methods must produce identical estimates of the coefficients on the covariates. However, it is possible to show that

$$\hat{\boldsymbol{\pi}} = \left[\sum_m \sum_{\ddot{c}_{im}=1} \{\mathbf{z}_i\mathbf{z}'_i - \overline{\mathbf{z}}_m\overline{\mathbf{z}}'_m\}\right]^{-1} \sum_m \sum_{\ddot{c}_{im}=1} \{\mathbf{z}_i y_i - \overline{\mathbf{z}}_m \overline{y}_m\},$$

whereas,

$$\hat{\boldsymbol{\eta}} = \left[\sum_k \sum_{t_{ikk}=1} w_{ik} \left\{\mathbf{z}_i\mathbf{z}'_i - (\overline{\mathbf{z}}_k^w)(\overline{\mathbf{z}}_k'^w)\right\}\right]^{-1} \left[\sum_k \sum_{t_{ikk}=1} w_{ik} \left\{\mathbf{z}_i y_i - (\overline{\mathbf{z}}_k^w)(\overline{y}_k^w)\right\}\right]$$

Thus, the coefficients on $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\eta}}$ will generally differ because the estimated coefficients implicitly re-center the data on a different basis. In the TTM, groups of teachers and individual teachers serve as the basis, while in the FRM only individual teachers serve as the basis. This implies that, when covariates are included, teacher regression coefficients from the FRM will not, in general, be numerically equal to the overall teacher effect obtained from the fully interacted TTM. As shown above, additional differences between $\hat{\beta}_k$ and $\hat{\gamma}_k$ will arise if an aggregated TTM is specified.

REFERENCES

Baker, Eva L., Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard. "Problems with the Use of Student Test Scores to Evaluate Teachers." Briefing Paper #278. Washington, DC: The Economic Policy Institute, 2010.

Buonaccorsi, John P. *Measurement Error: Models, Methods, and Applications.* Boca Raton, FL: Chapman & Hall/CRC, 2010.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood." NBER Working Paper No. 17699. Cambridge, MA: National Bureau of Economic Research, 2011.

Chiang, Hanley, and Duncan Chaplin. "Recommended Design of Value-Added Measures for Achievement First." Washington, DC: Mathematica Policy Research, 2010.

Goldhaber, Dan, and Duncan Chaplin. "Assessing the Rothstein Test: Does It Really Show Teacher Value-Added Models Are Biased?" Working Paper. Washington, DC: Mathematica Policy Research, 2012.

Glazerman, Steven, Susanna Loeb, Dan Goldhaber, Douglas Staiger, Stephen Raudenbush, and Grover J. Whitehurst. "Evaluating Teachers: The Important Role of Value-Added." Washington, DC: The Brookings Institution, 2010.

Goldberger, Arthur S. *A Course in Econometrics.* Cambridge, MA: Harvard University Press, 1991.

Hock, Heinrich, and Eric Isenberg. "Teacher and Student Misclassification in Value-Added Models." Manuscript. Washington, DC: Mathematica Policy Research, 2010.

Kane, Thomas J., and Douglas O. Staiger. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives,* vol. 16, no. 4, fall 2002, pp. 91–114.

Koedel, Cory, and Julian R. Betts. "Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique." *Education Finance and Policy,* vol. 6, no. 1, 2011, pp. 18–42.

McCaffrey, Daniel F., Daniel Koretz, J. R. Lockwood, and Laura S. Hamilton. *Evaluating Value-Added Models for Teacher Accountability.* Santa Monica, CA: The RAND Corporation, 2004.

Morris, Carl N. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association*, vol. 78, no. 381, March 1983, pp. 47–55.

Rothstein, Jesse. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *The Quarterly Journal of Economics,* vol. 125, no. 1, 2010, pp. 175–214.

Steele, Jennifer, Laura S. Hamilton, and Brian M. Stecher. *Incorporating Student Performance Measures into Teacher Evaluation Systems.* Santa Monica, CA: The RAND Corporation, 2010.

About the Series

Policymakers require timely, accurate, evidence-based research as soon as it's available. Further, statistical agencies need information about statistical techniques and survey practices that yield valid and reliable data. To meet these needs, Mathematica's working paper series offers policymakers and researchers access to our most current work.

For more information, contact Heinrich Hock, research economist, at hhock@mathematica-mpr.com or Eric Isenberg, senior researcher, at eisenberg@mathematica-mpr.com.