

Working PAPER

BY ERIC ISENBERG AND ELIAS WALSH

Accounting for Co-Teaching: A Guide for Policymakers and Developers of Value-Added Models

October 2013

Abstract

We outline the options available to policymakers for addressing co-teaching in a value-added model. Building on earlier work, we propose an improvement to a method of accounting for co-teaching that treats co-teachers as teams, with each teacher receiving equal credit for co-taught students. Hock and Isenberg (2012) described a method known as the Full Roster Method (FRM) that is feasible and practical, but it effectively counts co-taught students more than once—these students receive a full weight with each of their teachers, so such students receive extra weight when calculating the relationship between student characteristics and achievement. The improvement, known as the Full Roster-Plus Method, allows co-taught students to receive full weight with their teachers, but all students contribute equally to the calculation of the relationship between student characteristics and achievement. To investigate how the application of this method empirically changes value-added estimates, we use data from District of Columbia Public Schools, which uses a roster confirmation process that allows teachers to verify which of the students listed on their administrative rosters they actually taught. We find that there are very small empirical differences between the two methods.

I. THERE NEEDS TO BE A METHOD TO ANALYZE CO-TEACHING

Value-added models of teacher effectiveness have evolved from a statistical methodology employed by quantitative educational researchers to become a part of the tool kit used by district policymakers seeking rigorous methods to evaluate teacher effectiveness (Kane et al. 2012). In a value-added model, teachers are evaluated based on the achievement of their students, accounting for baseline student achievement and other measurable student characteristics. To estimate this kind of statistical model for such high-stakes applications as teacher evaluations requires high quality data on student test scores, other background characteristics, and teacher-student links.

To create high quality teacher-student links, many districts and some states have asked teachers to confirm the data available on administrative rosters of teacher-student links (Battelle for Kids 2013). This involves verifying which subjects teachers taught, which of the students listed on their administrative rosters they actually taught and for how long, and which students who are not on the rosters need to be added (Isenberg and Walsh 2013). In addition to revealing a high degree of departmentalization of instruction in upper elementary grades (Isenberg et al. 2013), roster-confirmed data have also revealed a level of co-teaching not previously documented in administrative rosters. For example, in District of Columbia Public Schools (DCPS), which has conducted roster confirmation for a high-stakes teacher evaluation system known as IMPACT, 29 percent of math teachers and 40 percent of reading teachers shared students with another teacher receiving a value-added score, and 9 percent of math teachers and 13 percent of reading teachers shared all of their students with another teacher in the 2010–2011 school year (Isenberg and Hock 2011). This high level of co-teaching has made it necessary for developers of value-added models to grapple with how to estimate the effectiveness of teachers who share some or all of their students with other teachers.

II. THREE METHODS, BUT ONLY ONE IS PRACTICAL

Hock and Isenberg (2012) discuss three methods of handling co-teaching within a value-added model: the Partial Credit Method, Teacher Team Method, and Full Roster Method. In addition to incorporating co-teaching into value added, all three methods allow for incorporating dosage (the proportion of the year a student spent with a teacher). These methods were defined and/or developed on behalf of DCPS, and, as Table II.1 shows, all three have been used for estimating value added—for teachers or for schools—in DCPS under the IMPACT evaluation system. All three are variants of a “teacher fixed effects” strategy, whereby a value-added model estimates the effectiveness of individual teachers by creating an indicator for each teacher that links that teacher to the students he or she taught, then estimating an effect for that indicator that can be interpreted as the relative effectiveness of this teacher compared to the average teacher in the district or state. A value-added model with teacher fixed effects can be a robust method of estimating teacher effectiveness even if some of the underlying assumptions of the model are not true (Guarino et al. 2012).

Table II.1. Methods of Accounting for Co-Teaching in DCPS Value-Added Models for IMPACT

School Year	Teacher Value Added	School Value Added
2009–2010	Teacher Team Method	Partial Credit Method
2010–2011	Full Roster Method	Partial Credit Method
2011–2012	Full Roster Method	Partial Credit Method
2012–2013	Full Roster-Plus Method	n.a.
2013–2014*	Full Roster-Plus Method	n.a.

Notes: DCPS discontinued the school value-added model as part of the IMPACT evaluation system for the 2012–2013 school year.

n.a. = not applicable

*For 2013–2014, DCPS plans to use the Full Roster-Plus Method.

The Partial Credit Method attempts to assign individual responsibility to teachers of shared students. Analytically, this is accomplished by assigning a dosage to each teacher in place of an indicator for each teacher-student link. For example, if a student switches from one teacher to another halfway through the year, both teachers would receive a dosage of 0.5. This may be theoretically appealing if a district believes teachers have different levels of effectiveness, and that the combination of two teachers in a team equals “the sum of the parts.”

As appealing as this may be in theory, the Partial Credit Method is rarely feasible in practice. If two teachers share all their students, it is not possible to assign separate estimates to both teachers. A similar situation arises when teachers share many students with each other and also have a few they teach individually. Situations like this can arise when there are complicated patterns of co-teaching, or due to data errors in roster confirmation (because it is costly to move toward 100 percent correspondence between the amount of time students spend in teachers’ classrooms and the student-teacher links that emerge from a roster-confirmation process). Estimates obtained in this situation tend to be unstable and unreliable. The Partial Credit Method is feasible when estimating value added at the school level, as it would be rare that two schools would share almost all of their students, but is not an option at the teacher level, given the patterns of co-teaching typically seen in roster-confirmed data.

One alternative to the Partial Credit Method is known as the Teacher Team Method. Rather than assuming that co-teachers have individual levels of effectiveness, in the Teacher Team Method, co-teachers are assumed to have a single, shared level of effectiveness, regardless of how effective each is when teaching students solo. In other words, the effect of the team may be more or less than the sum of its parts. In the value-added model, this is achieved by adding extra indicators for teams. The effectiveness of the teams is estimated along with the effectiveness of individual teachers. Teachers who have taught students solo and have also co-taught students receive a single estimate that is an average of their individual and team value-added estimates.

Unlike with the Partial Credit Method, it is analytically possible to estimate teacher value added using the Teacher Team Method, but it can be impractical. Because team effectiveness is estimated as a separate, single input into student achievement, there is no problem of trying to use unshared students to estimate the effectiveness of two teachers with a considerable amount of overlap. Rather, the problems with the Teacher Team model arise from the complexities of roster confirmation data, which result in the need to decide when to create a team indicator, and which can cause some students to be delinked from a teacher’s roster of claimed students (Isenberg and Hock 2010). In addition, a set of rules for how to handle teams of three or more teachers is required, especially when teachers share few students and complex patterns of sharing exist among them (for

example Teacher A shares some students with Teacher B, Teacher B shares with Teacher C, Teacher C shares with Teacher A, and all three share some students). Hock and Isenberg (2012) outline six specific decision rules that must be created in order to implement the Teacher Team model. Because of all the possible combinations of teaming that can arise from different data configurations as defined under these rules, it can be resource intensive and complicated to write programming code that is exhaustive.

A third method, known as the Full Roster Method (FRM), produces results that are nearly identical to the Teacher Team Method, but the FRM is analytically much simpler to estimate. Like the Teacher Team Method, the FRM assumes that co-teachers have a single, shared level of effectiveness. Under this method, students are linked to each of their teachers. Analytically, this involves creating unique records for each teacher-student combination in cases where students are co-taught by two or more teachers. For example, two records would be created for a student team taught by two teachers—one for each teacher-student combination. The value-added model produces one estimate per teacher, but some students contribute directly to the value-added estimate of multiple teachers. Because of its simplicity, no students are delinked from teachers (which is why this carries the “full roster” name). In the regression analysis, each teacher-student combination is weighted according the dosage for that teacher-student pair.

III. IMPLEMENTING THE FULL ROSTER METHOD LEADS TO A DILEMMA

Within the FRM, there are various options for how to assign weights to student-teacher combinations when there is co-teaching. We discuss three options, which involve setting the dosage weights so each fulltime, full-year student contributes either:

- The same dosage, no matter how many courses and teachers they have
- The same dosage per teacher, regardless of how many courses they take
- The same dosage per course-teacher combination

Equalizing weights regardless of the number of courses or teachers (option 1) is accomplished by subdividing the total student dosage among co-teachers so the sum of the dosage for every student is equal to his or her total dosage. For example, a student co-taught by two teachers for the whole school year would have a weight of 50 percent for each teacher. The advantage of this approach in the FRM is that data from all students contribute equally to the estimation of the relationship between student background characteristics and achievement.

The second option (used by DCPS in the 2010–2011 school year) is to have students count equally toward each of their teachers’ value-added estimates regardless of the number of courses taken. This option preserves the incentive for teachers to work equally hard to raise the achievement of all their students, regardless of the presence of a co-teacher. Compared to the first method, assigning equal weights within a teacher also protects against the impact of errors in the roster-confirmation procedure, in which one teacher would correctly add a missing student to her roster but another teacher would neglect to remove the student from hers. With this option, the error made by the second teacher does not affect the relative weight of students claimed by the first teacher. The drawback of this method is that each co-taught student is counted multiple times when determining how student characteristics contribute to value-added scores.

Option 3 is a variant of the second option. In option 2, each student could contribute at most 100 percent of their dosage to a given teacher regardless of the number of courses they take. Under option 3, a student may be assigned a dosage for a teacher in excess of 100 percent if the teacher claims him or her for more than one course during roster confirmation. For example, if a teacher has a student in a regular class and also in a second class designed for students who need extra help, the teacher would claim that student twice, leading to 200 percent dosage for that teacher-student link. This student would then count twice as much toward a teacher's value-added estimate as a student enrolled in only the regular class. This option moves more in the direction of allowing a district to set rules that flexibly weight the students within a teacher, but moves away from having all students contribute equally to the estimation of student characteristics.

IV. THE FULL ROSTER-PLUS METHOD RESOLVES THE DILEMMA OF THE FULL ROSTER METHOD

The FRM+ ensures equal weighting for students when contributing to the estimation of the student covariates, while allowing for any weighting of students within a teacher. It does this by creating extra student observations for nonduplicated students: each student has observations that sum to the total dosage of the student with the maximum dosage under the FRM. The new observations are linked to artificial teacher indicators so each teacher in the data set receives a "shadow teacher" who absorbs the extra dosage for each student required to weight up each student to the same level of dosage. The observations linked to these artificial teachers are not included in the teacher value-added estimates. Each student thereby contributes equally to the estimates of student characteristics without affecting the proportional contributions of co-taught students to teachers' scores. Details are given in the appendix.

If one believes students should be weighted equally when estimating how student characteristics contribute to value-added scores, the FRM+ should slightly decrease the bias in teacher estimates relative to the FRM. More than its effects on the estimates themselves, however, the gain to adopting the FRM+ may be in the increase in the face validity of the value-added model. Under FRM+, a school district or state is able to claim that all students contribute equally to the determination of the association between student background characteristics and student achievement, which is not true using the options 2 or 3 of the FRM.

V. EMPIRICALLY, THE FULL ROSTER-PLUS METHOD DOES NOT DIFFER MUCH FROM THE FULL ROSTER METHOD

We tested how different the results for teachers would be if we were to implement the FRM+ in place of the FRM, using roster-confirmed data from DCPS from the 2010–2011 school year. We estimated a value-added model under the FRM and again using the FRM+, and compared the results. For both methods, we used option 2, which gives each student the same weight per teacher, regardless of the number of courses taken. The value-added model accounted for a set of student background characteristics (including same-subject and opposite-subject pre-tests from the prior year), applied an errors-in-variables measurement error correction for pre-test scores, standardized value-added estimates to produce a similar distribution of teacher value added across grades, and applied empirical Bayes shrinkage to reduce the probability that teachers with imprecise estimates would receive extreme estimates. Details of the value-added model can be found in Isenberg and Hock (2011).

The relationships between student characteristics and achievement were not significantly affected by adopting the FRM+ in place of FRM, resulting in few changes to value-added estimates or the consequences for teachers in the IMPACT system. The overall results were similar to a value-added model that used FRM. The correlation in teacher value added from the two value-added models was above 0.9999 for both math and reading. We also examined a variety of other statistics to gauge the magnitude of the changes within the context of the DCPS IMPACT evaluation system. The other statistical tests are described more fully in Walsh and Isenberg (2013). For example, we measured the percentage of teachers who would have had a different Individual Value Added (IVA) score under IMPACT. IVA is a score that ranges from 1.0 to 4.0 in increments of 0.1; consequently, there are 31 possible IVA scores. As a result of switching from FRM to FRM+, only 8.5 percent of teachers' IVA scores would have been affected. All affected teachers would have moved just 0.1 points, or from one score to an adjacent score on the IVA scale. Using data on teachers' other IMPACT components, such as classroom observations, we examined what percentage of teachers would have changed from one IMPACT performance category to another, and found that only 0.6 percent (that is, six-tenths of 1 percent) of teachers would have changed IMPACT performance categories. Finally, there was almost no change in the average precision of the teacher value-added estimates when using the FRM+ in place of the FRM. The width of the average confidence interval of teacher value-added estimates was 0.1 percent smaller for math and 0.3 percent larger for reading compared to the value-added model that used FRM. These results are summarized in Table V.1.

Table V.1. How Much Value-Added Estimates Change Using the Full Roster-Plus Method Instead of the Full Roster Method

	Math	Reading	Combined
Correlation with FRM value-added model	0.99994	0.99999	--
Percentage of teachers who changed IVA scores	--	--	8.5%
Percentage of teachers who changed IMPACT effectiveness categories	--	--	0.6%
Percent increase in confidence intervals from FRM value-added model	-0.1%	0.3%	--

Note: Correlation is calculated as a Pearson correlation.

Although the magnitude of the changes in the results using these data was small, the results may be larger in different contexts. For example, when the DC value-added model was expanded to include high school teachers of English/language arts in grades 9 and 10 in the 2012–2013 school year, teachers were allowed to claim students multiple times if those students were enrolled with the teacher for several sections throughout the day (option 3 for handling dosage). As a result, some students had dosage in excess of 100 percent; one student had a dosage of 400 percent. More generally, for value-added models estimated across an entire state, if the data contributed from individual districts has discrepancies in the way dosage is handled for a teacher, the FRM+ would offer a means by which each student would contribute equally to the estimates of student covariates across the state, regardless of how the data are contributed from individual school districts. In these contexts, we might expect that the magnitude of the changes might be a bit larger.

VI. CONCLUSION: IT IS UP TO THE SCHOOL DISTRICT

1. In a value-added model, it is challenging to attribute different levels of effectiveness to teachers who share students in roster-confirmed data sets.
2. The Full Roster Method offers a practical solution to the problem of incorporating co-teaching into value added, as long as policymakers are willing to accept that teachers who share students will be treated as teams, with all members of the team receiving equal credit.
3. Policymakers must choose how to weight students who are taught by multiple teachers or in multiple courses by the same teacher.
4. Under the Full Roster Method, the choice of weights will indirectly affect all teachers by affecting the regression coefficients for student characteristics in the value-added model.
5. The Full Roster-Plus Method allows districts to choose any weighting method while allowing each student to contribute equally to the calculation of the regression coefficients for student characteristics.
6. In theory, the Full Roster-Plus Method offers an advantage over the Full Roster method, but empirically it makes little difference.

Given these facts about the modeling of co-teaching, we conclude that it should be left to district policymakers to decide how to handle the weighting of co-taught students and whether they add the extra level of complexity and resources needed to model the FRM+ in place of the FRM. The FRM+ solves a theoretical dilemma posed by the implementation of the FRM, but seems to make little practical difference in the value-added estimates. So the validity of a value-added model is changed little; the main advantage of moving from FRM to FRM+ is the increase in face validity.

REFERENCES

- Battelle for Kids. "Roster Verification." Columbus, OH: Battelle for Kids, 2013.
- Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge. "Can Value-Added Measures of Teacher Education Performance Be Trusted?" East Lansing, MI: Education Policy Center at Michigan State University, December, 2012.
- Hock, Heinrich, and Eric Isenberg. "Methods for Accounting for Co-Teaching in Value-Added Models." Washington, DC: Mathematica Policy Research, June 2012.
- Isenberg, Eric, and Heinrich Hock. "Measuring School and Teacher Value Added for IMPACT and TEAM in DC Public Schools." Washington, DC: Mathematica Policy Research, August 2010.
- Isenberg, Eric, and Heinrich Hock. "Design of Value Added Models for IMPACT and TEAM in DC Public Schools, 2010–2011 School Year." Washington, DC: Mathematica Policy Research, May 2011.
- Isenberg, Eric, Bing-ru Teh, and Elias Walsh. "Elementary School Data Issues: Implications for Research." Chicago, IL: Mathematica Policy Research, October 2013.
- Kane, Thomas, Douglas Staiger, Steve Cantrell, Jeff Archer, Sarah Buhayar, Kerri Kerr, Todd Kawakita, and David Parker. "Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains." Seattle, WA: Bill & Melinda Gates Foundation, 2012.
- Walsh, Elias, and Eric Isenberg. "How Does a Value-Added Model Compare to the Colorado Growth Model?" Chicago, IL: Mathematica Policy Research, October 2013.

APPENDIX: STATISTICAL DETAILS

In this appendix, we describe the statistical details of the Full Roster Method (FRM) and the Full Roster-Plus Method (FRM+). In general, teacher value added using the FRM can be calculated as follows:

$$(1) \quad Y_i = \lambda_1' P_i + \beta_1' X_i + \delta_1' T_{1ii} + \varepsilon_{1ii}$$

where Y_i is the post-test score for student i and P is a vector of pre-test scores for student i from previous years. The pre-test scores capture prior inputs into student achievement. The vector X_i denotes control variables for other individual student background characteristics. The vector T_{1ii} contains one indicator variable for each teacher. In each record, only one indicator variable has a one; the rest are zeros. A student contributes one observation to the model for each teacher to whom the student is linked. Students were weighted in the regression according to their dosage—the amount of time the teacher taught the student. The vector δ_1 includes the value-added estimates, one coefficient for each teacher. Finally, ε_{1ii} is the random error term. We account for clustering at the student level so that the standard errors of teacher estimates are not artificially small due to student records being duplicated in the analysis file.

To equalize the contribution of students to the estimation of the coefficients on student background characteristics, the FRM+ variation changes the analysis file that feeds into the value-added regression model. To do this, observations are replicated in the data set and assigned a dosage weight so that all students have the same amount of total dosage in the analysis file. The new records are linked to artificial teacher indicators so each teacher in the data set receives a “shadow teacher” who absorbs the extra dosage for each student required to assign each student the same total dosage.¹ The shadow teacher links are recorded in T_{2ii} , distinct from T_{1ii} . This yields the following regression for the FRM+:

$$(2) \quad Y_i = \lambda_2' P_i + \beta_2' X_i + \delta_2' T_{1ii} + \theta_2' T_{2ii} + \varepsilon_{2ii}$$

The dosage for the original observations does not change in this process. Each student thereby contributes equally to the estimates of student characteristics without affecting the proportional contributions of co-taught students to measures of teachers’ effectiveness.

The original teacher indicators, T_{1ii} , are set to zero for all of the duplicate observations, and the new set of teacher indicators, T_{2ii} , are set to zero for all of the original observations. Including the duplicate but distinct teacher links as indicators in the regression allows the duplicate observations to contribute to the estimation of coefficients on student background characteristics but not to the

¹ For computational reasons, it is a good practice to add one to the dosage of the student with the maximum dosage to set the total dosage for all students. For example, if the greatest student dosage is three (hundred percent), set the total dosage at four. For a student who is claimed by only one teacher for the whole year, the associated shadow teacher would then have a dosage of three for a total dosage of four for this student. This step ensures that all teacher-student records are replicated. Otherwise, students who are at the maximum total dosage value would not be replicated, which can lead to computational problems if some shadow teachers are linked to few students.

estimation of the teacher effects used to estimate the value added scores. The vector δ_2 includes one coefficient for each teacher-grade combination. The coefficients in the vector θ_2 are not used to calculate teachers' value-added estimates. In the FRM+, we account for clustering at the student level so the standard errors are robust to repeated observations for individual students who have multiple teachers and also to including the shadow teacher records.

Authors' Note

We thank the Office of the State Superintendent of Education of the District of Columbia (OSSE) and the District of Columbia Public Schools (DCPS) for providing the data for this study. We are grateful to Duncan Chaplin for helpful comments. Juha Sohlberg provided excellent programming support. The paper was edited by Sharon Peters and produced by Jackie McGee. The text reflects the views and analyses of the authors alone and does not necessarily reflect views of Mathematica Policy Research, OSSE, or DCPS. All errors are the responsibility of the authors.

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 about this paper, contact Eric Isenberg, senior researcher, at ejisenberg@mathematica-mpr.com.

www.mathematica-mpr.com

**Improving public well-being by conducting high-quality,
objective research and surveys**

PRINCETON, NJ - ANN ARBOR, MI - CAMBRIDGE, MA - CHICAGO, IL - OAKLAND, CA - WASHINGTON, DC

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.