VALUE-ADDED MEASUREMENT Services

at American Institutes for Research

Estimating Teacher Contributions to Student Learning

The Role of the School Component

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Estimating Teacher Contributions to Student Learning:

The Role of the School Component

August 2013

Zeyu Xu, Ph.D.
Andrew Swanlund

VALUE-ADDED MEASUREMENT Services

at American Institutes for Research

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202-403-5000 | TTY 877-334-3499
www.educatortalent.org

The research presented in this paper was supported by internal research funds from American Institutes for Research (AIR). The views represented herein are not necessarily representative of AIR.

AIR
AMERICAN INSTITUTES FOR RESEARCH®

wwww.air.org

1978_08/13

Acknowledgments

The authors gratefully acknowledge useful comments received from Juan Copa and Kathy Hebda (Florida Department of Education), Pete Goldschmidt (New Mexico Public Education Department), and Jane Hannaway (National Center for Analysis of Longitudinal Data in Education Research) as well as Dan Sherman at AIR.

Contents

Page

Introduction

Teachers are the most important school factor related to student performance (Aaronson, Barrow, & Sander, 2007; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). Teacher recruitment, retention, and development are hence at the center of education policies and interventions that aim to improve student learning and to close performance gaps between students from different socioeconomic backgrounds. The cornerstone for the success of these human resources policies is the ability to identify high- and low-performing teachers based on performance measures that are fair, accurate, and timely. As a result, teacher labor market research and policy in recent years has shifted focus from measures of qualifications and experience to measures of teacher effectiveness, often quantified by teacher value-added effects on student learning. This process has been expedited by the availability of increasingly sophisticated state longitudinal administrative data as well as federal and state policy changes that push for the incorporation of measures of student learning growth in teacher evaluation systems.

Extensive literature exists on value-added modeling and the properties of teacher value-added estimates (e.g., Aaronson et al., 2007; Hanushek & Rivkin, 2010; McCaffrey, Sass, Lockwood, & Mihaly, 2009; Staiger & Rockoff, 2010). Moving from research to practice, several prominent technical and practical issues emerge. One of the challenging issues is whether and how to attribute student growth between teachers and schools. Many value-added models designed to produce teacher-level estimates of student growth do not decompose the total growth above expectation among a teacher's students into a *common school component* (the portion of growth that is common across all teachers in a school) and a *teacher component* (the portion of student growth associated with a teacher as distinct from the school). As a result, these models implicitly attribute all of the differences in student growth observed across teachers to the teachers themselves. In contrast, alternative model specifications decompose student growth into a teacher component and a school component. In this case, the estimated teacher value-added excludes any contribution to student learning by school factors that are common across teachers within a school.

Accounting for and explicitly estimating the influence of the common school component on student growth (often referred to as the "school effect") is important because students and teachers are not randomly sorted into schools. Part of the correlation between teachers and student growth stems from the matching between teachers and schools. The literature on teacher mobility consistently finds that teachers prefer working in schools with students who are higher achieving and of higher socioeconomic status; because the most experienced teachers are more likely to receive their preferred teaching assignments, teacher assignment policies often result in pairing the most inexperienced teachers with the most educationally needy schools (Hanushek, Kain, O'Brien, & Rivkin, 2005). These policies can introduce a positive correlation between teacher quality and family contribution to student learning (Hanushek, Kain, & Rivkin, 2004). In addition, a school that is well managed and more effective in producing high-achieving students is likely to attract the best teachers (Loeb, Béteille, & Kalogrides, 2012), introducing a positive correlation between teacher quality and school factors that are conducive to student growth.

This sorting of teachers, along with the direct effects of school-level inputs on student achievement (e.g., school safety, leadership quality) will introduce omitted-variables bias into the value-added estimates of teacher effects. Many of those factors are beyond teachers' control, resulting in an unfair evaluation of a teacher's effectiveness. Technically, factors unaccounted for in a regression model are absorbed in the model error term. To the extent that those factors are associated with student test scores and teacher effectiveness (through teacher-school sorting), the requirement that teacher effects are orthogonal (i.e., independent) to the error term is violated, and teacher value-added estimates will be biased. It is typical for value-added models (VAMs) to control for student, classroom, and school characteristics that are available in state or local administrative data. Yet how much of the *unobservable* heterogeneity can be accounted for by these variables is an empirical question with no definitive answers. One way to account for the effect of unobserved school heterogeneity on student test scores—both the direct effect and the indirect effect through its correlation with the teacher effect—is to include either random or fixed school effects in the VAM. By doing so, between-school variation, both observed and unobserved (as long as it is common across teachers within school), is completely removed and each teacher is compared with colleagues in the same school.

An important drawback of this approach in practice is that such teacher value-added estimates must be considered as within-school measures. This designation prevents any cross-school comparisons of teacher effectiveness and makes statewide or districtwide teacher evaluation impossible. In addition, some scholars caution that such an approach probably goes too far (Hanushek et al., 2004) because between-school variation in the common school component can be partially influenced by individual teacher efforts. For instance, the presence of high-performing teachers in a school creates a more productive school environment that makes all teachers in the same school better (a teacher "peer effect"), as reported by Jackson and Bruegmann (2009). The presence of high-performing teachers in a school makes the school more attractive to other high-performing teachers considering a transfer—one of the lessons learned in the Institute of Education Sciences Talent Transfer Initiative (see Glazerman, Protik, Teh, Bruch, & Seftor, 2012). Through these mechanisms, individual teachers contribute to the common school component and, therefore, part of the "school effect" should be attributed to teachers.

Finally, the decision about how much (if any) of the school effect should be attributed to teachers requires more than a technical or statistical consideration. Such decisions can have important policy consequences by creating either positive or adverse incentive structures for teachers. For example, completely removing the common school component from teacher value-added estimates may create adverse incentives against collaboration and the sharing of best teaching practices with colleagues within a school. This problem may be magnified when teacher value-added scores are tied to high-stakes decisions such as differentiated pay for performance.[1]

In contrast, attributing at least some of the common school component to teachers may encourage collaboration. However, attributing too much of the school component to teachers may create a "free-rider" problem, encouraging less effective teachers to work in schools with

[1] The design of pay-for-performance programs varies greatly. Often, the teacher value-added score is used in conjunction with some other individual- and school-level performance measures, thus mitigating potential adverse effects that would have resulted from using value-added scores alone.

high school-value-added scores, and discouraging teachers from teaching in schools that are the neediest.

More empirical investigation is needed on the impact of including or not including the school component in teacher value-added scores and, if included, how much weight to apply to the school component relative to the teacher effect. Educators need to understand how large the between-school variance in student performance is and how much of the variation in teacher value-added is between schools. The following sections will discuss each question in turn, which will be followed by a summary of what has been learned in the value-added work conducted by American Institutes for Research (AIR) in Florida.

Between-School Variation in Student Test Scores

A small but nontrivial portion of the total variation in student test scores can be attributed to between-school differences. Using test scores on state standardized assessment in North Carolina and Florida, researchers Zhu, Jacob, Bloom, and Xu (2012) reported that about 7 percent to 10 percent of the total variation in student test scores at the elementary school level is between schools. An additional 6 percent to 14 percent of the total variation is between classrooms/teachers within the same school. The rest of the variation is attributed to differences between students within classrooms/teachers. The researchers also reported that variance decomposition of student test scores is somewhat different at the secondary school level, where 7 percent to 17 percent of the total variation is between schools and 20 percent to 38 percent is between classrooms/teachers within the school. However, the elementary and secondary school results are not directly comparable because the former is based on state end-of-grade tests in mathematics and reading whereas the latter is based on end-of-course tests in mathematics and science subjects.

These findings are consistent with the landmark publication *Equality of Educational Opportunit*y, also known as the Coleman Report (Coleman et al., 1966), which documented that less than 16 percent of the variance in student test scores can be attributed to between-school differences. Kane and Staiger (2002) similarly concluded that the heterogeneity in student scores within the average school is nearly as large as the heterogeneity in scores overall. Even though school-level variation in student test scores is small relative to the total variation, it is not trivial relative to the teacher-level variation. The school-level variance component is about the same in magnitude as the teacher-level variance component at the elementary school level and about one third in magnitude at the secondary school level.

The between-school variance is even smaller when student score *gains* are used in calculation. Chiang and Schochet (2010) documented the relative size of variance components at the school, classroom/teacher, and student level based on 10 value-added studies. They found that on average, slightly more than 1 percent of the total variation in student score gains is between schools and about 7 percent of the total variation is between teachers and classrooms within each school. Kane and Staiger (2002) reported that the between-school variance in mean student *gains* among schools is only about one fifth as large as the between-school variance in mean fourth-grade scores—that is, there is much more variability in scores than in gain scores.

Researchers have pointed out that a significant portion of the school-level variation is due to sampling errors and other idiosyncratic fluctuations. That is, the variation in school average test score estimates across schools consists of three components: variation in long-term differences among schools, variation due to sampling errors that are driven by school sizes (larger schools having more precise estimates than smaller schools), and variation due to random fluctuation in any particular year. As a result, the long-term (or persistent) between-school performance differentials could be much smaller. Kane and Staiger (2002) calculated that sampling errors alone account for 14 percent to 15 percent of the variation of fourth-grade mathematics and reading test scores at the school level, and they calculated double that when using gain scores for schools. They reported that the between-school variance in mathematics and reading gain scores is roughly three times as large for schools in the smallest quintile in size as for schools in the largest quintile—that is, the smaller the schools, the larger the between-school variance tends to be. Including other nonpersisting factors that may lead to changes in school-level student performance, Kane and Staiger also reported that between 50 percent and 80 percent of the variance in the change in mean fourth-grade scores to be nonpersistent. This finding is an important point to keep in mind when considering attributing some of the school contribution to student growth to teachers.

Teacher Value-Added Estimates With and Without the School Component

Between-School Variation of Teacher Value-Added Estimates

Given the variance distribution of student test scores (either level scores or gain scores) within and between schools, it is not surprising that the empirical literature on teacher value-added finds most of the variation in estimated teacher value-added to be among teachers working in the same school rather than differences across schools (Kane & Staiger, 2008). Between-school variation accounts for roughly 12 percent to 20 percent of the total variation in estimated teacher value-added each year.

The empirical literature finds that including school effects in teacher VAMs has a relatively small impact on the size of teacher value-added estimates. Kane and Staiger (2008) reported that the standard deviation of mathematics teachers' value-added estimates reduces from 0.23 standard deviations *(SD)* in value-added models without school fixed effects to 0.22 *SD* in models with school fixed effects. For English language arts teachers, the standard deviation of teacher value-added estimates reduces from 0.18 to 0.17 *SD*. Other studies (e.g., Koedel & Betts, 2009a, 2009b) reported drops in standard deviation units of similar magnitude (about 5 percent). Some papers report larger decreases (about 20 percent) in standard deviations of teacher value-added estimates when school fixed effects are included in the model. For example, Hanushek et al. (2005) reported that the standard deviation of teacher value-added scores changes from 0.27 *SD* when school fixed effects are excluded from the model to 0.22 *SD* when school fixed effects are included. Aaronson et al. (2007) reported that the standard deviation of eighth- and ninth-grade mathematics teachers' value-added effect changes from 0.24 *SD* to 0.19 *SD* when school fixed effects are added to their model. The largest teacher effect size reduction as a result of the inclusion of school fixed effects was reported in Hanushek and Rivkin (2008), from 0.20 *SD* to 0.13 *SD*, a reduction of about 34 percent. These differences within the literature likely reflect differences among the study samples (e.g., districts having different types of tests), not just inconsistency in estimation methods or modeling specifications.

Teacher Mobility

Changes in the productivity of mobile teachers (teachers who change schools) provide another perspective on the impact of school components on teacher value-added estimates. Mobile teachers provide an opportunity to investigate *within-teacher* productivity variations that could be explained by school-level differences. If a teacher's productivity changes substantially before and after a school move, it could imply that either school factors play an important role in the students' learning even though the teacher's own productivity remains unchanged or the teacher's productivity has improved (or deteriorated) because of contextual differences between schools. If, conversely, teacher productivity changes very little before and after a school change, especially when the sending and receiving schools have very different characteristics (such as student poverty rates and school performance levels), it would imply that whether to include or exclude the school component in calculating teacher value-added is less of a concern.

Jackson (2010) reported that on average, North Carolina teachers who switched schools increased their value-added by 0.025 *SD* for mathematics and 0.014 *SD* for reading. When school-by-year effects are included in the model, the results are virtually unchanged, indicating that the estimated outcome differences before and after a move are not spuriously driven by teachers moving from low- to high-achievement schools and are not driven by schoolwide events that would affect both teacher mobility and teacher performance. Xu, Ozek, and Corritore (2012) similarly reported small positive changes or no changes in teacher value-added estimates among teachers who switched schools. More important, the authors found that this pattern holds regardless whether teachers are moving from high-performing (or high-poverty) schools to low-performing (or low-poverty) schools or vice versa. These findings on productivity change for mobile teachers provide additional exploratory evidence that school components probably play a small role in estimating teacher value-added.

Teacher Value-Added Rankings With and Without the School Component

Most studies that employ VAMs, both with and without school effects, look only at how the size of the estimated teacher effect (as measured in standard deviations of student score gains) changes. More pertinent in practice and to individual teachers, however, is how relative teacher rankings may be affected by whether school effects are specified in the model. Aaronson et al. (2007) reported the rank correlations of teacher value-added estimated under various model specifications. The authors had three key findings. First, teacher value-added rankings are sensitive to the inclusion or exclusion of the common school component. The rank correlation (after correcting for estimation errors) is between 0.65 and 0.86 (both specifications include the usual student, peer, and neighborhood variables). Second, when the school component is accounted for, teacher value-added rankings appear to be robust to modeling choices. Third, value-added rankings become very sensitive to the model specification when the school component is not included.

Beyond a “Common” School Component

So far, this paper has considered only whether or how the “school effect” may affect estimates of individual teacher value-added. Conceptually, such “school effects” are the common components that apply to all teachers equally within a school. But in addition to these common components, school effects also can manifest themselves as indicators of school-teacher matching quality. In other words, the school component may have a heterogeneous impact on the performance of individual teachers within the same school. Teachers with certain characteristics and backgrounds may work better in one type of school culture (types of colleagues, school leadership styles, and so on) than in another. Should this teacher-school-specific component be attributed to the teachers or to the schools? The answer to this question depends on what a teacher evaluation system is trying to measure. From a school or students’ perspective, the teacher-school-specific component should be attributed to the teachers because it represents a real benefit to the students. For a district or state, because this part of the teacher performance is “school specific,” it is not fully transferrable to other types of schools or students and, therefore, its value may not be fully recognized in the general district/statewide teacher labor market.

A match effect is anything that makes a teacher more or less productive at one school versus another that is not caused by a school characteristic that affects all teachers equally (Jackson, 2010). Anything that affects all teachers at a school equally would be part of a school effect (i.e., the “common” school component), and only those combinations of characteristics that vary at the teacher-by-school level are part of a match effect. By some estimates, 10 percent (for mathematics teachers) to 50 percent (for reading teachers) of what typically is called a “teacher effect” is actually a match effect (Jackson, 2010).

In summary, from a research perspective and at the aggregate level, it seems that the amount of variation in student learning that can be attributed to teachers is marginally affected by whether school fixed or random effects are included in a model. However, to individual teachers in a state or local performance evaluation system, value-added-based performance rankings are sensitive to this decision.

The next section presents evidence from AIR’s Florida value-added work on how much difference the inclusion or exclusion of the common school component makes in estimating teacher annual value-added. Specifically, it presents empirical results as related to the following questions:

1. What are the relative magnitudes of the school- and teacher-level variance components?
2. What are the rank correlations between teacher value-added estimates that incorporate varying portions of school effects?
3. What are the percentages of teachers in the top and the bottom 10 percent categories (those most likely to be affected by high-stakes decisions) that change performance categories when varying portions of school effects are attributed to teachers?
4. Are teachers with certain characteristics or working in particular school types more likely to be affected by whether school effects are added to teacher effects?

Lessons Learned From AIR's Florida Value-Added Work

Through a contract with the Florida Department of Education, AIR is responsible for designing and implementing the teacher value-added component of Florida's evaluation system. The *Florida Value-Added Model: Technical Report* (American Institutes for Research, 2011) summarizes the exploratory analysis that formed the basis of the state's value-added modeling decisions. It shows, among other things, that the variance between students within a class is the largest of all variance components, that the variance between schools tends to be smaller than the variance between teachers within a school, and that the relative size of variance components varies by grade.

Figures 1 and 2 show the root mean squared error (residual error) and standard deviation of the teacher and school components for mathematics and reading, respectively, in the Florida model.

Figure 1. Magnitude of Teacher and School Variance Components: Mathematics

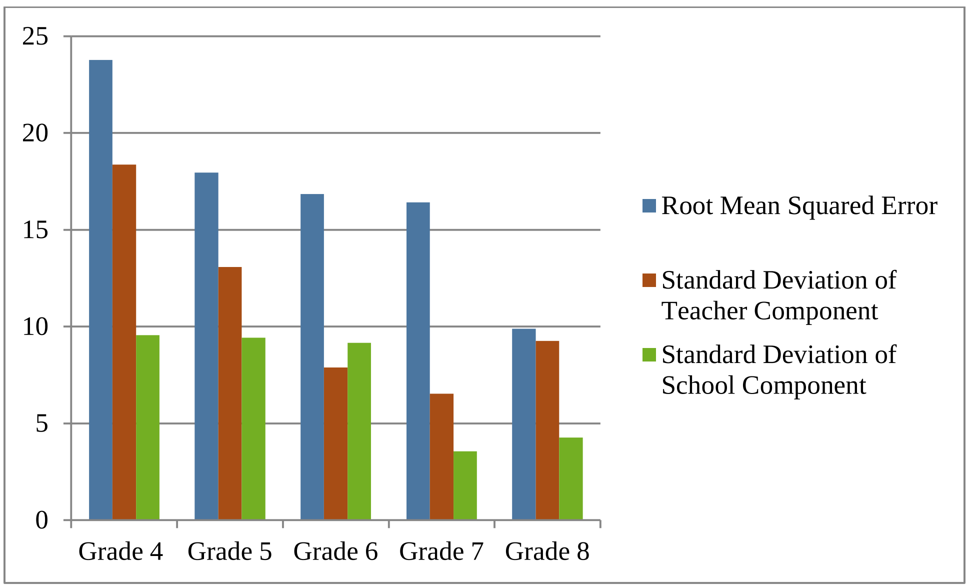
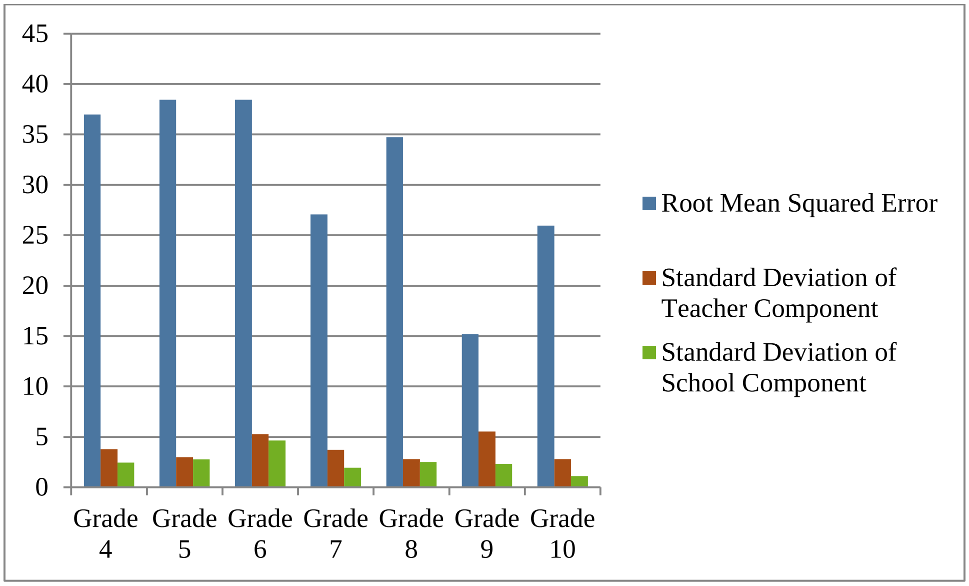


Figure 2. Magnitude of Teacher and School Variance Components: Reading



The Florida value-added model decomposes total variation in student achievement on the state's standardized Florida Comprehensive Assessment Test assessments (which measure student success with the Sunshine State Standards) into three orthogonal components: variation between schools, variation between teachers within a school, and variance between students within a classroom. The Florida Student Growth Implementation Committee (SGIC) held significant discussion on the topic of whether to model both teacher and school effects and on how much of the school component (if any) to include in the teacher value-added score. Committee members concluded that teachers are responsible in part for what goes on in the school beyond their classroom. Not including the school component in the model would attribute all differences across classrooms to the teachers (including the differences that are common to all teachers in a school). Adding the school component to the Florida value-added model would remove differences common to the school, thereby giving teachers no credit (or demerit) for the common performance of the school. With no school effect in the model, teachers would be compared with all teachers in the state. Including the school effect would result in teachers being compared with the teachers within their own school. After significant discussion, the SGIC determined that some of the school effect should be attributed back to teachers. Therefore, the Florida model estimates both teacher and school effects, which allows a proportion of the school effect to be included in the teacher value-added score. The SGIC agreed to a 50 percent proportion.

The remaining figures and tables in this paper examine the impact of that decision. In particular, they examine the relationship between teacher scores that include 0 percent, 25 percent, 50 percent, 75 percent, and 100 percent of the school component. Understanding the difference between adding a school component into a teacher value-added score and including a school component in the model is key to examining the effect of a particular approach. Fitting a value-added model without a school component is analogous to including 100 percent of the school component in the teacher's score. In contrast, fitting a model that includes a school component is analogous to including 0 percent of the school component in the teacher's score. As stated earlier, the Florida model includes both a teacher and a school effect. Therefore, the teacher effects that are drawn from the model include 0 percent of the school component. To include 50 percent of the school component in the value-added score (as is done in Florida) means that the teacher value-added score is equal to the teacher effect plus half the value of the school effect.

The equation used for calculating the final teacher value-added score is as follows:

$$Teacher\ VAM\ Score = teacher\ effect + (X \times school\ component)$$

The X in the equation represents the percentage of the school component that is added to the teacher effect (in Florida, this was set to 0.5 for 50 percent). Changing the value of X would add more or less of the school component to the teacher effect. For example, if X were 0, then the teacher value-added score would simply be equal to the teacher effect from the model. If X were 1, the value-added score would be the sum of the teacher effect and the school component (and thus similar to the results from a model that did not include a school component at all).

The results presented here show the impact on teacher value-added scores of adding different proportions of that school component to the teacher effect (beyond the 50 percent selected by SGIC). Adding some of the school component to teachers' scores will necessarily change the rank order of the value-added estimates. The following analyses examine the extent to which teachers' scores change based on the amount of school component included in the score. These analyses also examine the demographic characteristics of the teachers and schools for which the change occurs.

Table 1 and Table 2 show the rank correlation between the model-based teacher effects (within-school effects) and teacher value-added scores that include 25 percent, 50 percent, 75 percent, or 100 percent of the school component for mathematics and reading, respectively, in the Florida model. In general, the correlation among the scores is high. However, as more of the school component is added to the teacher effect, the correlation declines to as low as 0.677 (for Grade 6 mathematics) and 0.648 (for Grade 6 reading). These results indicate that adding some (or all) of the school component to the teacher scores does result in a change in the ordering of teachers in terms of effectiveness. The appendix contains additional tables showing how top and bottom decile teacher ratings change when some or all of the school component is added to teacher scores.

Table 1. Rank Correlation Between Teacher Effects and Value-Added Estimates, Including a Portion of the School Component—Mathematics

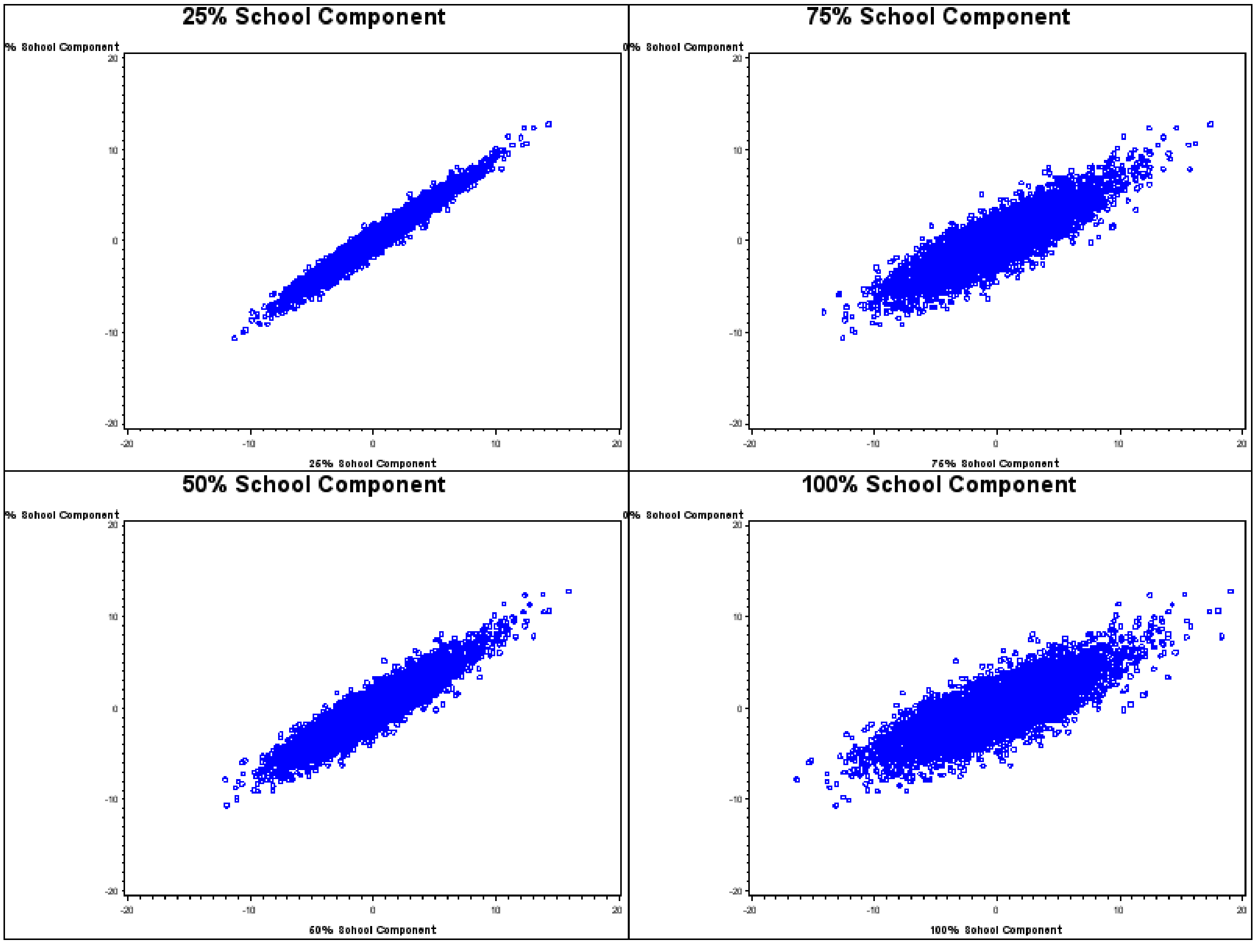
Grade	Percentage of School Component Added			
	25%	50%	75%	100%
4	0.981	0.938	0.885	0.831
5	0.971	0.907	0.838	0.774
6	0.943	0.841	0.750	0.677
7	0.975	0.920	0.858	0.799
8	0.984	0.944	0.895	0.845

Table 2. Rank Correlation Between Teacher Effects and Value-Added Estimates, Including a Portion of the School Component—Reading

Grade	Percentage of School Component Added			
	25%	50%	75%	100%
4	0.955	0.874	0.797	0.731
5	0.923	0.813	0.724	0.656
6	0.933	0.821	0.724	0.648
7	0.957	0.877	0.798	0.731
8	0.915	0.796	0.701	0.631
9	0.971	0.909	0.839	0.773
10	0.949	0.865	0.786	0.720

Figure 3 provides a visual representation of the changing of the rank ordering of teachers. It shows four scatterplots of teacher effects that include no school component (the vertical axes) against teacher scores that include 25 percent, 50 percent, 75 percent, and 100 percent of the school component (the horizontal axes). The results are for Grade 5 mathematics but generalize to other grades and subjects. As with the rank-order correlations, the ordering of teachers changes in a nontrivial manner as larger proportions of the school component are added to the teacher effect. The findings are consistent with those reported in Aaronson et al. (2007). Comparing teacher value-added scores estimated with and without the common school component, the rank correlation ranges between 0.65 and 0.85.

Figure 3. Plot of Teacher Value-Added Estimate (Including a Portion of the School Component) Versus Teacher Effect



Figures 4 through 7 are designed to show the impact of different policy decisions on individual teachers—in particular, the impact of including larger proportions of the school components on the teachers with the highest and lowest teacher effects. Figures 4 and 5 show (for reading and mathematics, respectively) the percentage of teachers who were classified in the bottom 10 percent of teachers (based on a teacher effect that includes 0 percent of the school component) who would move out of the bottom 10 percent if a different proportion of the school component were added to their scores. In general, the more that the school component is added to the scores, the more the teachers shift out of the bottom 10 percent (and other teachers shift downward into the bottom 10 percent). Figures 6 and 7 show similar results but for teachers classified in the top 10 percent of teachers (based on a teacher effect that includes 0 percent of the school component). The more that the school component is added to the teacher effect, the more teachers shift out of the top decile. The effect of the different percentages varies across grades and subjects. Grade 6 mathematics and Grade 8 reading appear to be most affected by the addition of the school component.

Figure 4. Percentage of Teachers Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect by Grade: Mathematics

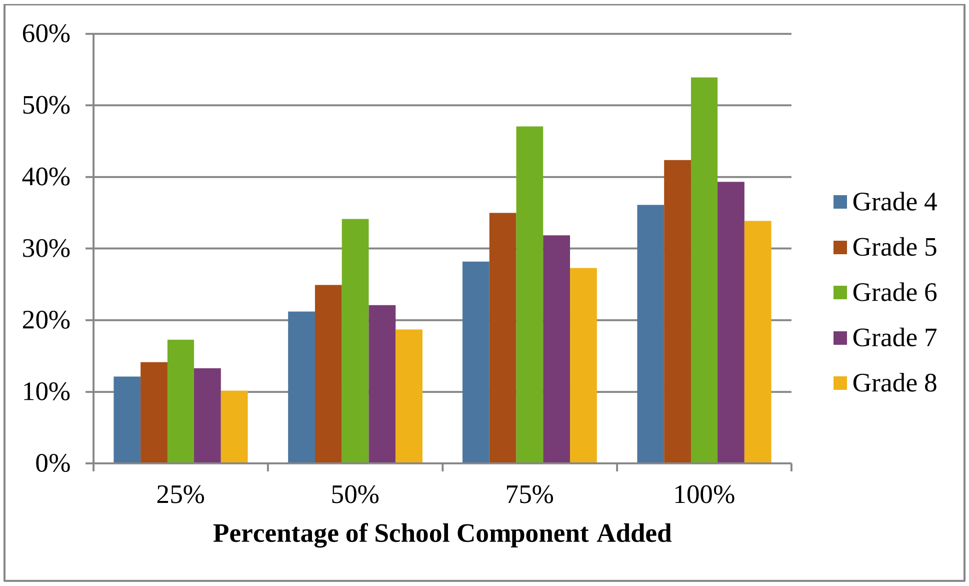


Figure 5. Percentage of Teachers Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect by Grade: Reading

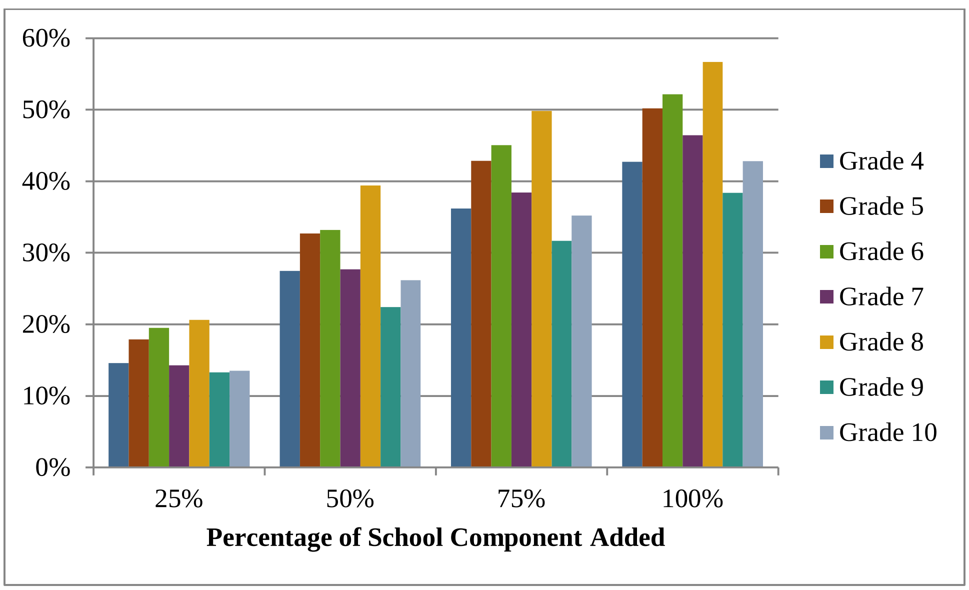


Figure 6. Percentage of Teachers Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect by Grade: Mathematics

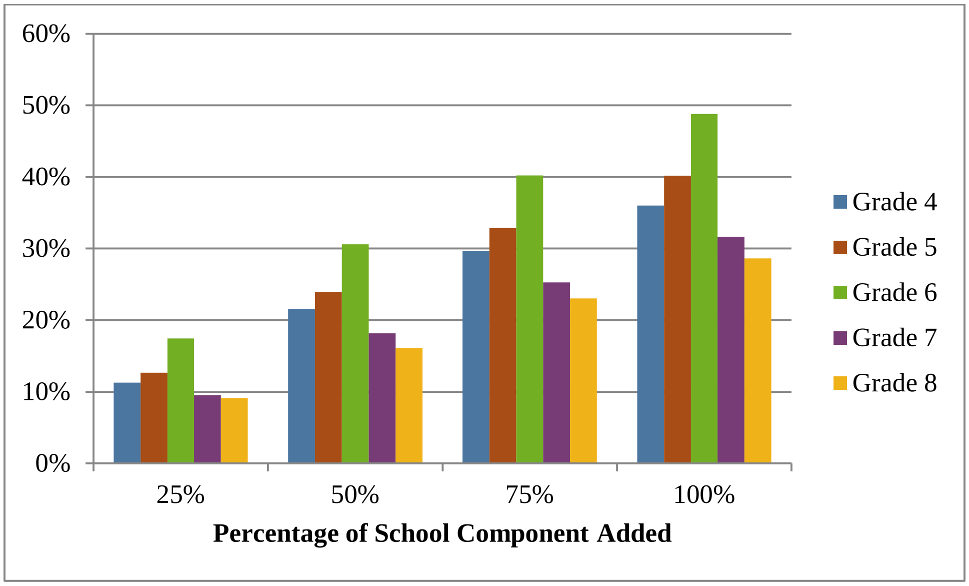
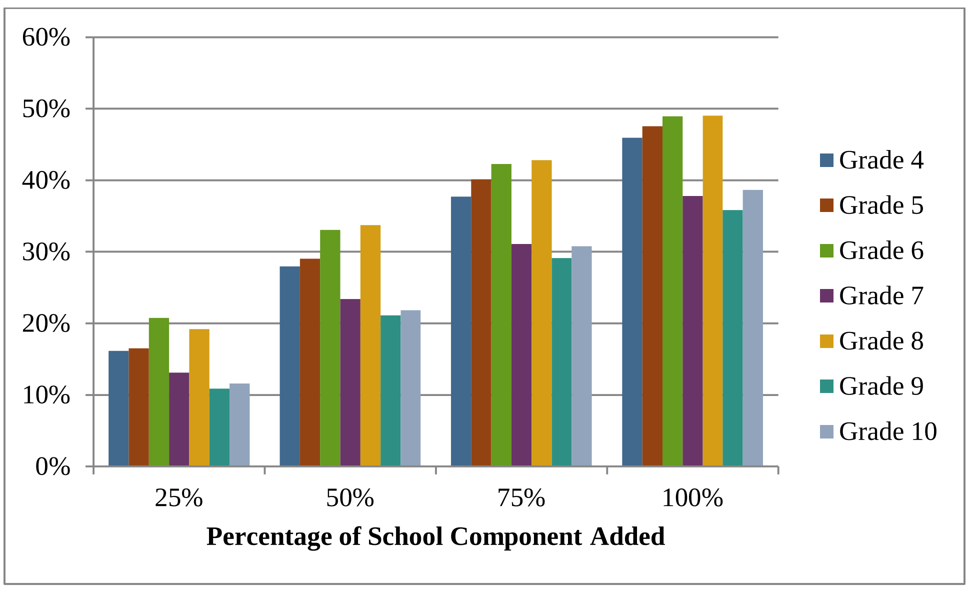


Figure 7. Percentage of Teachers Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect by Grade: Reading



To take this analysis further, it is worth examining which types of teachers are affected by different decisions about the school component. The next section presents results by several demographic characteristics for the teachers who stayed in the top or bottom 10 percent and teachers who moved out of the top or 10 percent. These characteristics are as follows:

- Percentage of students receiving free or reduced-price lunch in the school
- Percentage of minority students in the school

- Title I status of the teacher's school
- Teacher's years of experience

The following results show where differences exist in the types of teachers who are moving out of the top or bottom 10 percent of teachers—that is, which teachers are advantaged or disadvantaged under different policy options.

Percentage of Students Receiving Free or Reduced-Price Lunch (FRPL)

Figures 8 through 11 show the percentage of FRPL students in the school for teachers staying in the top and bottom deciles. For teachers in the bottom decile—the least effective teachers when the school component is not added—there is a greater likelihood that teachers at schools with lower percentages of FRPL students will improve their scores as more of the school component is added. Similarly, teachers at schools with higher percentages of FRPL students are more likely to see a decrease in their scores as more of the school component is added. These results appear to show that the effect of including the school component in teacher's scores reduces (or fails to improve) the scores more often for teachers in higher poverty schools.

Figure 8. Average Percentage of FRPL Students (School Level) for Teachers Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Mathematics

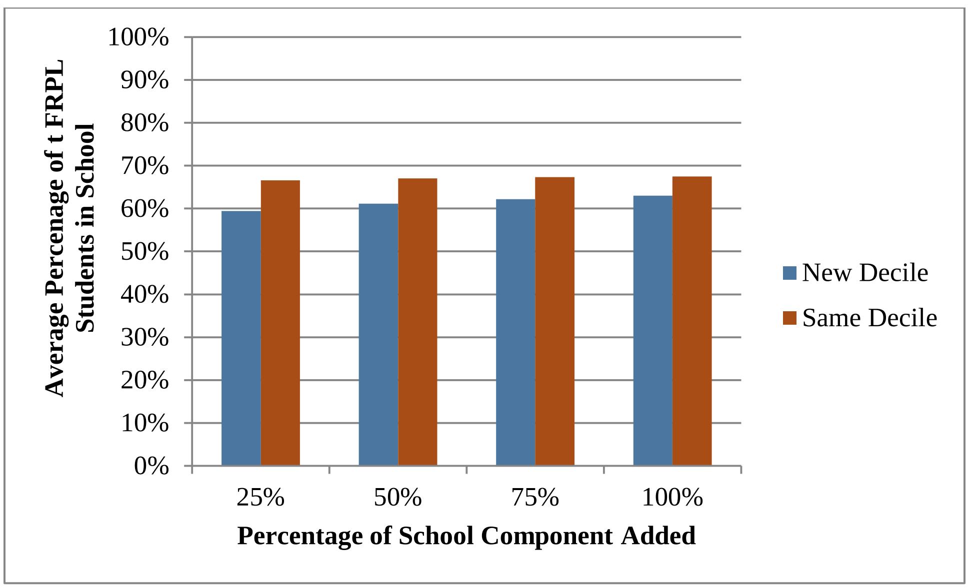


Figure 9. Average Percentage of FRPL Students (School Level) for Teachers Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Mathematics

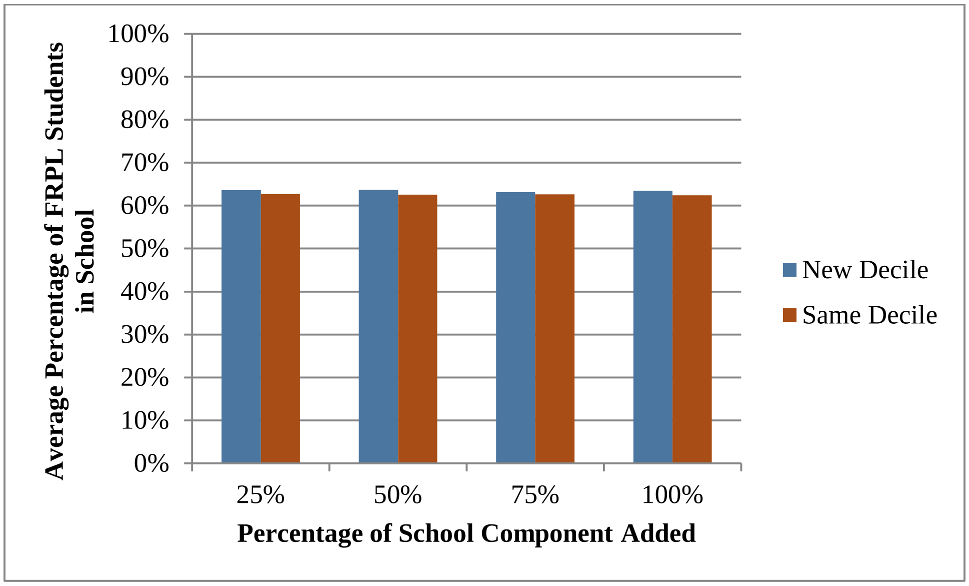


Figure 10. Average Percentage of FRPL Students (School-Level) for Teachers Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Reading

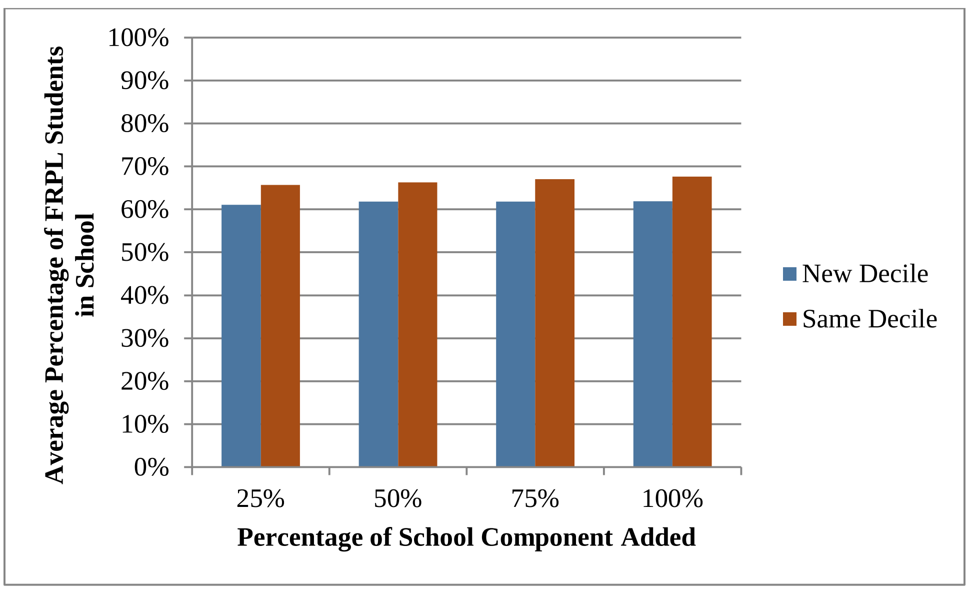
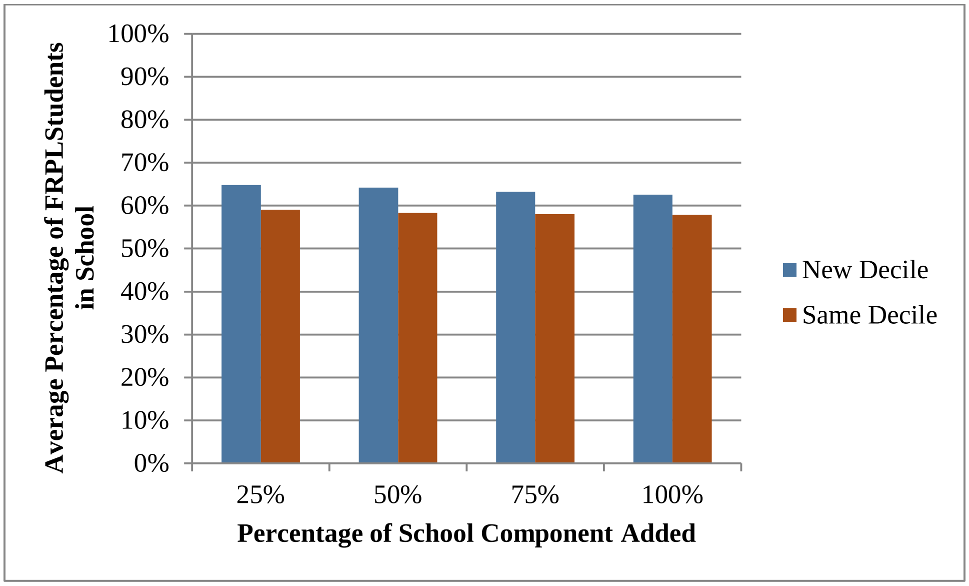


Figure 11. Average Percentage of FRPL Students (School Level) for Teachers Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Reading



Percentage of Minority Students

Figures 12 through 15 show the percentage of minority students in the school for teachers staying in the top and bottom deciles. For teachers in the bottom decile—the least effective teachers when the school component is not added—there is a greater likelihood that teachers at schools with lower percentages of minority students will improve their scores as more of the school component is added. Similarly, teachers at schools with higher percentages of minority students are more likely to see a decrease in their scores as more of the school component is added. These results appear to show that the effect of including the school component in teachers' scores reduces (or fails to improve) the scores more often for teachers in schools with a higher concentration of minority students.

Figure 12. Average Percentage of Minority Students (School Level) for Teachers Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Mathematics

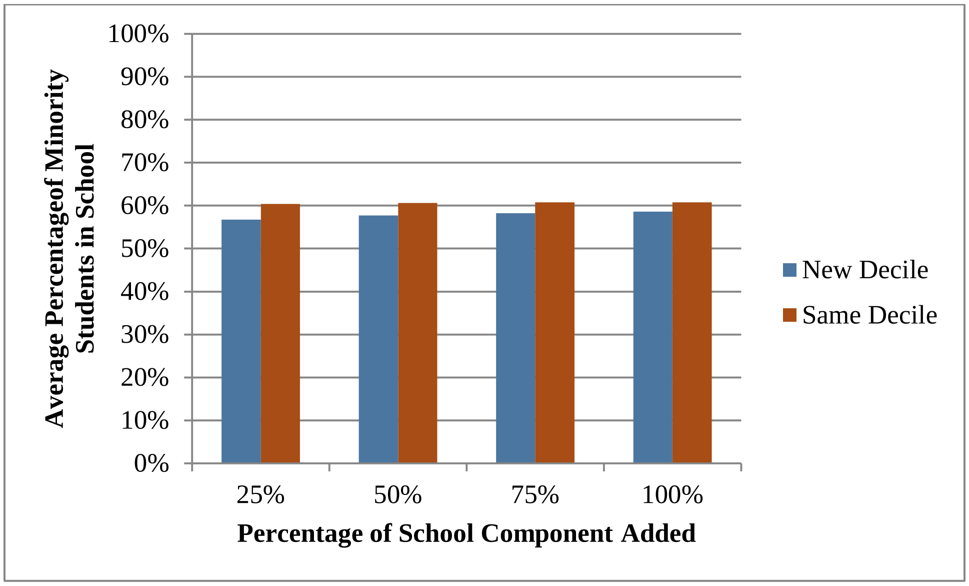


Figure 13. Average Percentage of Minority Students (School Level) for Teachers Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Mathematics

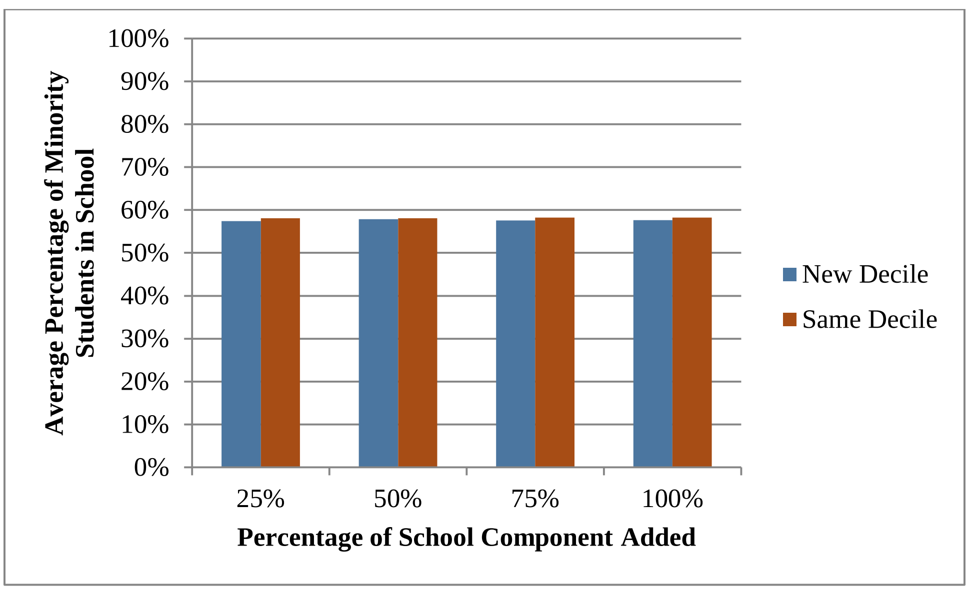


Figure 14. Average Percentage of Minority Students (School Level) for Teachers Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Reading

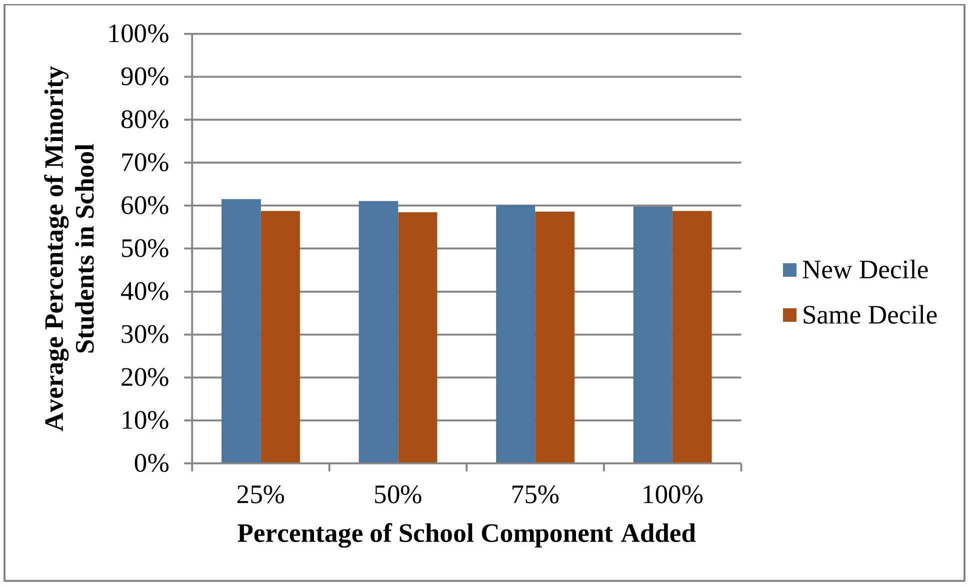
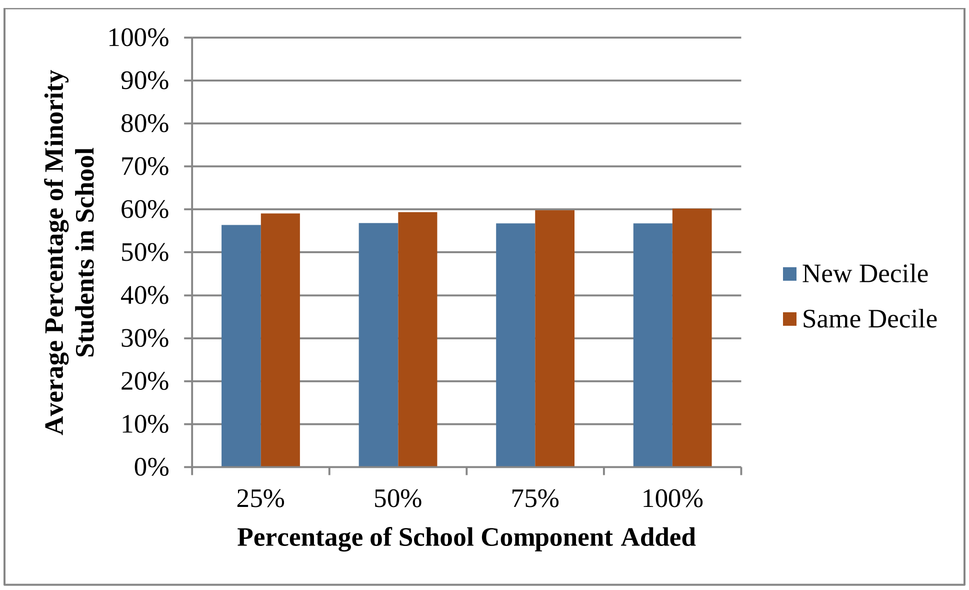


Figure 15. Average Percentage of (School Level) for Teachers Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Reading



School Title I Status

Figures 14 through 17 show the percentage of teachers who teach at Title I schools for teachers staying in the top and bottom deciles. For teachers in the bottom decile—the least effective teachers when the school component is not added—there is a greater likelihood that teachers at non–Title I schools will improve their scores as more of the school component is added. Similarly, teachers at Title I schools are more likely to see a decrease in their scores as more of the school component is added. These results appear to show that the effect of including the school component in teachers' scores reduces (or fails to improve) the scores more often for teachers in Title I schools.

Figure 16. Proportion of Teachers at Title I Schools Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Mathematics

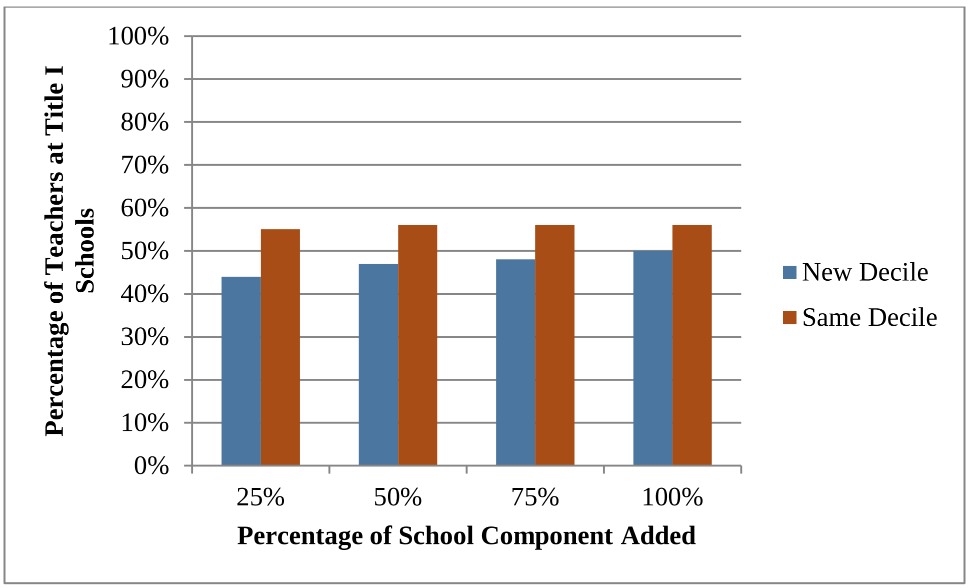


Figure 17. Proportion of Teachers at Title I Schools Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Mathematics

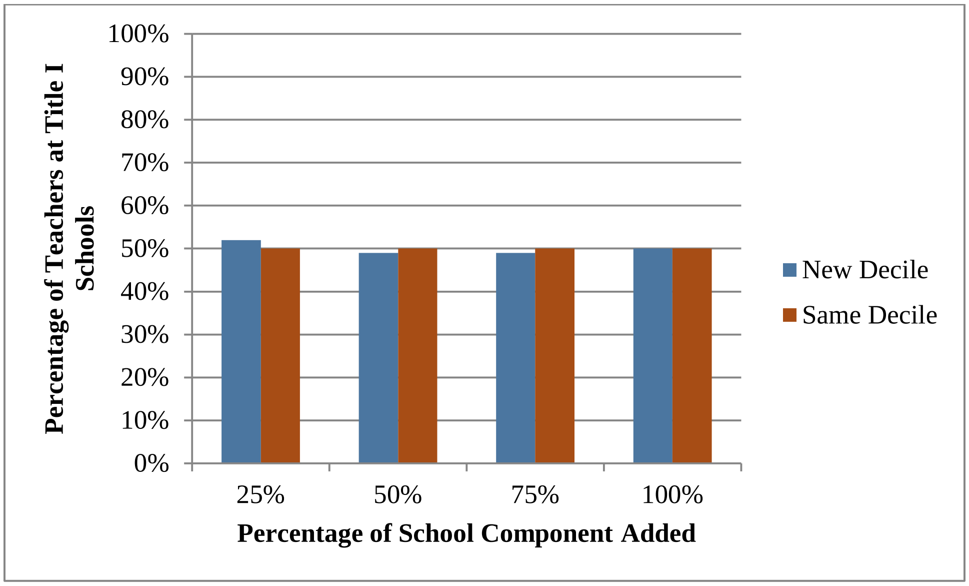


Figure 18. Proportion of Teachers at Title I Schools Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Reading

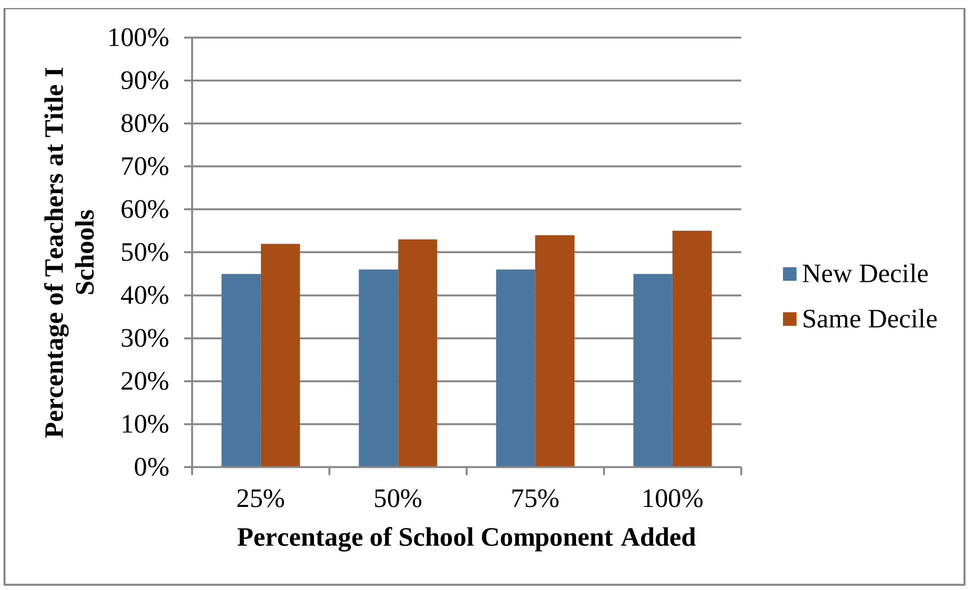
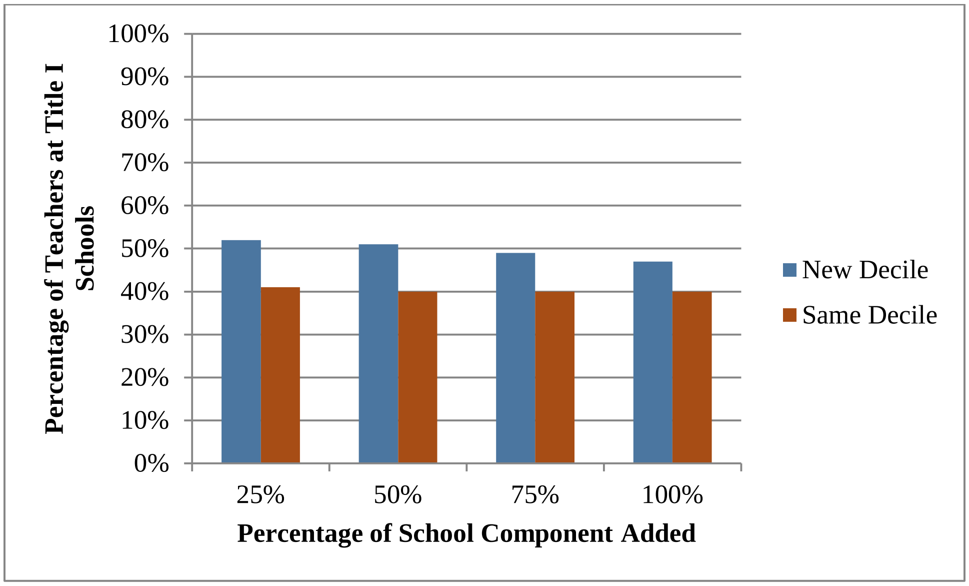


Figure 19. Proportion of Teachers at Title I Schools Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Reading



Teacher Experience

Figures 20 through 23 show the average years of teaching experience for teachers staying in the top and bottom deciles. For teachers in the bottom decile—the least effective teachers when the school component is not added—there is a greater likelihood that teachers with more experience will improve their scores as more of the school component is added. However, at the top end of the spectrum, teachers with more experience more often see their scores reduced when more of the school component is added. This relationship is different from the other three demographic variables considered. With teacher experience, it appears that—in general—adding more of the school component moves teachers with more experience toward the center of the distribution of scores.

Figure 20. Average Years of Experience for Teachers Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Mathematics

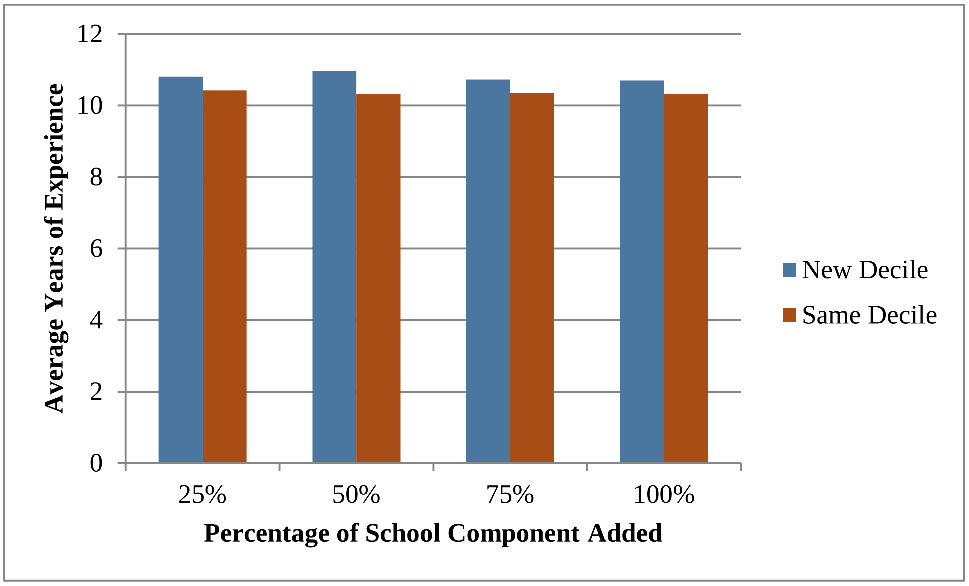


Figure 21. Average Years of Experience for Teachers Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Mathematics

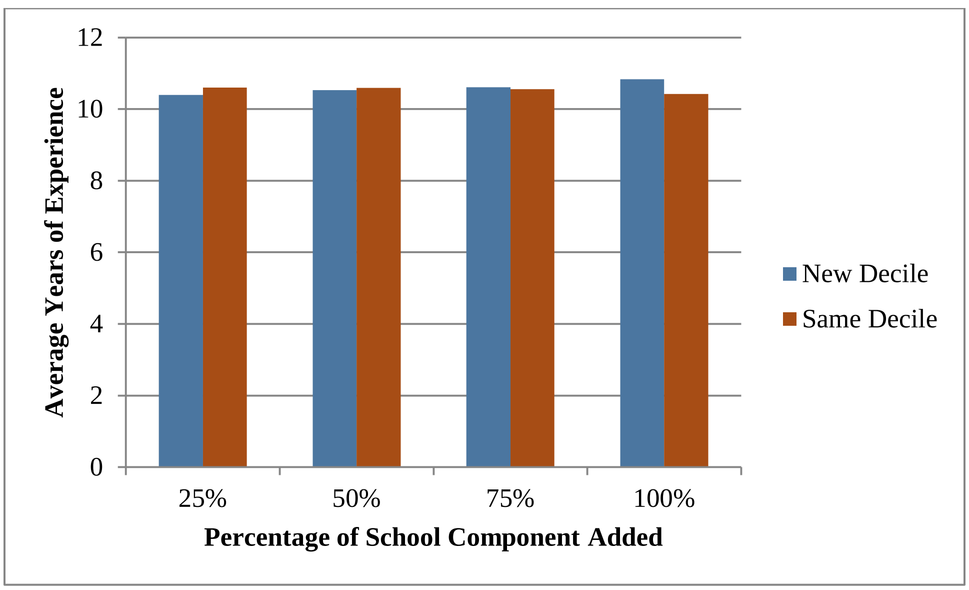


Figure 22. Average Years of Experience for Teachers Moving Out of the Bottom 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Reading

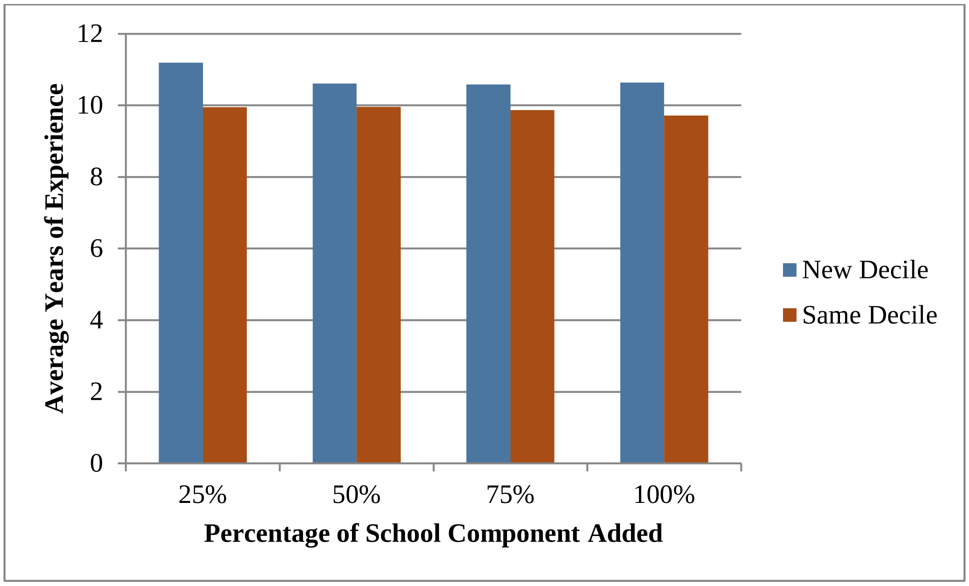
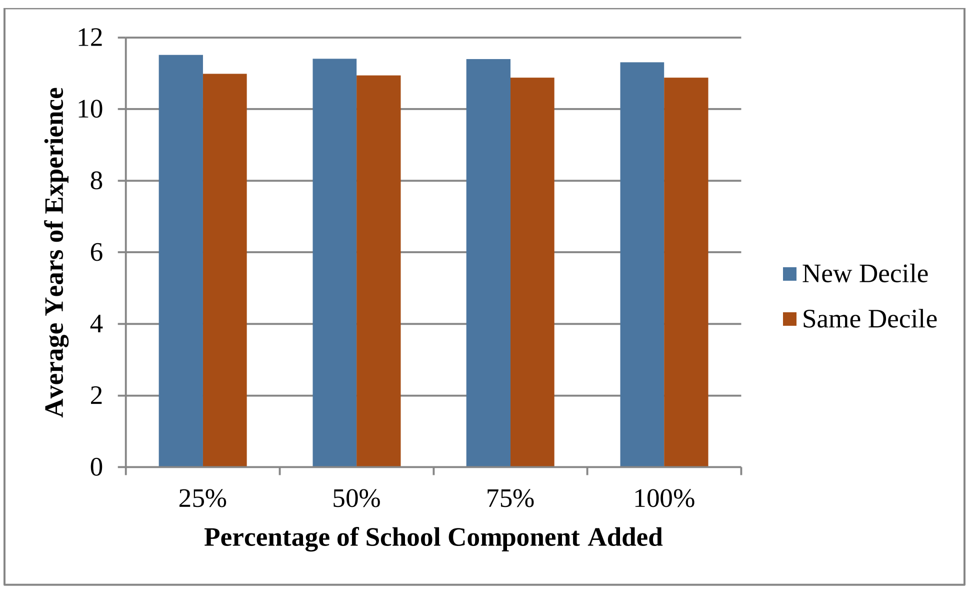


Figure 23. Average Years of Experience for Teachers Moving Out of the Top 10 Percent When Varying Proportions of the School Component Are Added to the Teacher Effect: Reading



When examining the results for percentage of FRPL students, percentage of minority students, Title I status, and teacher experience, a general trend becomes apparent. For these data, teachers who teach in high-poverty, high-minority schools will benefit less from adding larger percentages of the school component into the teacher value-added score than teachers from low-poverty, lower minority concentration schools.

Conclusion

This paper has provided a discussion of the research on the contributions of teachers and schools to student learning. Although teachers tend to contribute more to the variation in student scores, a small but nontrivial portion of the variance also can be attributed to between-school differences. Policymakers, therefore, face important decisions in determining how to isolate and use the information produced from value-added models regarding teachers and the schools in which they teach. Removing all the school effect from the teacher value-added scores results in teacher effectiveness scores in which teachers are compared only with teachers within their same school and are not held accountable for any systematic changes in student learning at the school level. Including the full school effect in the teacher value-added scores holds teachers accountable for everything that happens at their schools. In reality, the "truth" of the situation is probably somewhere in the middle.

Although empirical evidence (such as that presented here for Florida) can help policymakers decide how much of the school component to factor in to their educator effectiveness metrics, there is no statistical answer to the problem (as the "truth" is always unknown). The results from Florida presented here indicate that although school effects are somewhat small compared to teacher effects, the decision about how much to include in the final teacher score has an important effect on the rankings of the teachers[2]. The validity of any decision regarding the inclusion (or exclusion) of the school component hinges on the ability of policymakers and stakeholders to understand the rationale for and implications of that decision.

One possible practice is to report more than one score for each teacher.[3] For example, one score could be estimated without controlling for the school component to indicate how the teacher ranks against all other teachers in the state. Another score could be estimated within schools or within school types, such as defined by student demographics or by school performance levels; this score could indicate how a teacher compares with colleagues in the same school or in the same type of schools. The decision on which score receives more weight in a teacher evaluation can be left to school and district leaders, depending on local priorities. If the school or district culture emphasizes collegiality and coordinated efforts in making a school better, the value-added score that includes the common school component should receive more weight. By comparison, if the school has very large variation in its teachers' productivity and focuses on bringing up the lower end of its teacher effectiveness distribution, the within-school value-added scores are probably more pertinent to the school's immediate goal and should receive more weight in its teacher evaluation. The decision depends on the intended use of teacher value-added scores and what goal a teacher evaluation system is striving to achieve—both of which are likely to vary across schools and districts. Therefore, presenting more than one version of the value-added score allows flexibility to address varying local priorities.

---

[2] It is important to note that in our analysis, single-year teacher value-added scores were used. In practice, aggregate value-added scores based on multiple years of performance are typically used. The impact of school effects on teacher rankings based on multiple-year value-added scores may be significantly smaller than those reported here.
[3] Additional scores may add more confusion for teachers, and therefore the most likely target users may be school and district leadership.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95–135. Retrieved from http://faculty.smu.edu/millimet/classes/eco7321/papers/aaronson%20et%20al.pdf

American Institutes for Research. (2011). *Florida's value-added model: Technical report*. Washington, DC: Author. Retrieved from http://www.fldoe.org/committees/doc/Value-Added-Model-Technical-Report.docx

Chiang, H., & Schochet, P. (2010). *Error rates in measuring teacher and school performance based on student test score gains.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf

Coleman, J. S. Campbell, E. Q., Hobson, C. J., McPartland, J. , Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Department of Health, Education, and Welfare; Office of Education.

Glazerman, S., Protik, A., Teh, B., Bruch, J., & Seftor, N. (2012). *Moving high-performing teachers: Implementation of transfer incentives in seven districts* (NCEE 2012-4501). U.S. Department of Education, Institute of Education Sciences, National Center on Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/pubs/20124051/pdf/20124051.pdf

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (NBER Working Paper 11154). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w11154.pdf

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources, 39*(2), 326–354.

Hanushek, E. A., & Rivkin, S. G. (2008). *Do disadvantaged urban schools lose their best teachers?* (Brief 7). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. Retrieved from http://www.eric.ed.gov/PDFS/ED508277.pdf

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100*(2), 267–271.

Jackson, C. K. (2010). *Match quality, worker productivity, and worker mobility: Direct evidence from teachers* (NBER Working Paper 15990). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w15990.pdf

Jackson, C. K., & Bruegmann, E. (2009). *Teaching students and teaching each other: The importance of peer learning for teachers* (NBER Working Paper 15202). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w15202.pdf

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives, 16*(4), 91–114. Retrieved from http://pubs.aeaweb.org/doi/pdfplus/10.1257/089533002320950993

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (NBER Working Paper 14607). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14607.pdf

Koedel, C., & Betts, J. (2009a). *Does student sorting invalidate value-added models of teacher effectiveness?* Columbia: University of Missouri. Retrieved from http://economics.missouri.edu/working-papers/2009/WP0902_koedel.pdf

Koedel, C., & Betts, J. (2009b). *Value-added to what? How a ceiling in the testing instrument* (NBER Working Paper 14778). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14778.pdf

Loeb, S., Béteille, T., & Kalogrides, D. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy, 7*(3), 269–304.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy, 4*(4), 572–606.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–457.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252.

Staiger, D., & Rockoff, J. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives, 24*(3), 97–118.

Xu, Z., Ozek, U., & Corritore, M. (2012). *Portability of teacher value-added across school settings.* Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. Retrieved from http://www.caldercenter.org/publications/upload/wp77.pdf

Zhu, P., Jacob, R., Bloom, H., & Xu, Z. (2012). Designing and analyzing studies that randomize schools to estimate intervention effects on student academic outcomes without classroom-level information. *Educational Evaluation and Policy Analysis, 34*(1), 45–68.

Appendix. Additional Tables

The tables included in this appendix show the percentage of teachers in Florida who would shift out of the bottom or top 10 percent of teachers given different proportions of the school component being included in the teacher value-added score. Tables A1and A3 show the percentage of teachers no longer in the bottom 10 percent when varying proportions of the school component are added, for mathematics and reading, respectively. Tables A2 and A4 show a similar pattern of movement for teachers classified in the top 10 percent (before adding a proportion of the school component).

Table A1. Percentage of Bottom Decile Teachers Effects No Longer in Bottom Decile—Mathematics: By Percentage of School Component Added

Grade	Percentage of School Component Added			
	25%	50%	75%	100%
4	12.1%	21.2%	28.2%	36.1%
5	14.1%	24.9%	35.0%	42.4%
6	17.3%	34.1%	47.1%	53.9%
7	13.3%	22.1%	31.8%	39.3%
8	10.2%	18.7%	27.3%	33.9%

Table A2. Percentage of Top Decile Teachers Effects No Longer in Top Decile—Mathematics: By Percentage of School Component Added

Grade	Percentage of School Component Added			
	25%	50%	75%	100%
4	11.3%	21.6%	29.7%	36.0%
5	12.7%	23.9%	32.9%	40.2%
6	17.5%	30.6%	40.2%	48.8%
7	9.6%	18.2%	25.3%	31.6%
8	9.1%	16.1%	23.1%	28.6%

Table A3. Percentage of Bottom Decile Teachers Effects No Longer in Bottom Decile—Reading: By Percentage of School Component Added

Grade	Percentage of School Component Added			
	25%	50%	75%	100%
4	14.6%	27.5%	36.2%	42.7%
5	17.9%	32.7%	42.9%	50.2%
6	19.5%	33.2%	45.0%	52.2%
7	14.3%	27.7%	38.4%	46.4%
8	20.6%	39.4%	49.8%	56.7%
9	13.3%	22.4%	31.7%	38.4%
10	13.5%	26.2%	35.2%	42.8%

Table A4. Percentage of Top Decile Teachers Effects No Longer in Top Decile—Reading: By Percentage of School Component Added

Grade	Percentage of School Component Added			
	25%	50%	75%	100%
4	16.1%	28.0%	37.7%	46.0%
5	16.5%	29.1%	40.1%	47.6%
6	20.8%	33.1%	42.3%	49.0%
7	13.1%	23.4%	31.1%	37.8%
8	19.2%	33.7%	42.8%	49.0%
9	10.9%	21.1%	29.1%	35.8%
10	11.6%	21.8%	30.8%	38.6%

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

Value-Added Measurement Services, part of AIR's Educator Talent Management services and resources, can assist states and districts in finding the right approach to measuring how educators contribute to growth in student achievement.

1978_08/13