education policy analysis archives

A peer-reviewed, independent,
open access, multilingual journal

epaa aape

Arizona State University

Volume 23 Number 38 March 30th, 2015 ISSN 1068-2341

A Value-Added Study of Teacher Spillover Effects Across Four Core Subjects in Middle Schools

Kun Yuan
RAND Corporation
United States

Citation: Yuan, K. (2015). A value-added study of teacher spillover effects across four core subjects in middle schools. *Education Policy Analysis Archives, 23*(38). http://dx.doi.org/10.14507/epaa.v23.1761

Abstract: This study examined the existence, magnitude, and impact of teacher spillover effects (TSEs) across teachers of four subject areas (i.e., mathematics, English language arts [ELA], science, and social studies) on student achievement in each of the four subjects at the middle school level. The author conducted a series of value-added (VA) analyses, using multiple years of state achievement test scores in the four tested subjects for students at grades 7 and 8 from an urban school district in the Southern US. Results showed evidence that mathematics and ELA teachers jointly contributed to student achievement in mathematics and ELA. ELA teachers also showed TSEs on student achievement in science (at grade 8 only) and social studies (at both grade levels), with ELA teachers' effect sizes close to or even greater than those of the own-subject teachers at grade 8. Results also showed that controlling for TSEs slightly decreased the variation and precision of teachers' VA scores and changed the quartile rankings of individual teachers' VA scores for a non-negligible group of teachers (11%–25%). On average, the percentage of teachers whose VA rankings were affected due to controlling for TSEs was greater for test subjects with TSEs than subjects without TSEs. Results challenge the current practice of ignoring teachers' TSEs when estimating teachers' VA scores. Results also support the use of group-based incentive plan when rewarding secondary mathematics and ELA teachers based on student achievement growth in these two subjects.

Journal website: http://epaa.asu.edu/ojs/
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 6/9/2014
Revisions received: 9/30/2014
Accepted: 10/2/2014

Keywords: value-added modeling; teacher spillover effects; teacher evaluation; middle school

Un Estudio Sobre el Valor Agregado y la Influencia Indirecta de los Docentes en Cuatro Materias Básicas de Escuelas Medias

Resumen: El presente estudio examinó la existencia, magnitud y el impacto de la influencia de los docentes (EET) a través de profesores de cuatro áreas temáticas (por ejemplo, matemáticas, artes del lenguaje Inglés [ELA], ciencias y estudios sociales) en el rendimiento de los estudiantes en cada uno de los cuatro materias en escuelas secundarias. El autor realizó una serie de valor agregado (VA), utilizando varios años de resultados de las pruebas estatales de logro en las cuatro materias que evalúan estudiantes en los grados 7 y 8 de un distrito escolar urbano en el Sur de los EE.UU. Los resultados proveen evidencia de que maestros de matemáticas y ELA contribuyeron conjuntamente al logro de los estudiantes en matemáticas y ELA. Maestros de ELA también mostraron las EET en el rendimiento de los estudiantes en la ciencia (solamente en grado 8) y los estudios sociales (tanto a nivel de grado), con efecto de tamaño cercanos o incluso superiores a los de los maestros de grado 8. Los resultados también muestran que el control de las EET se redujo ligeramente según la variación y precisión de los VA de las puntuaciones docentes y cambió la clasificación de VA de los cuartiles individuales de maestros para un grupo no despreciable de los docentes (11% -25%). En promedio, el porcentaje de maestros cuyo VA resultaron afectados por el control de las EET fue mayor para los sujetos con casos de EET que para los sujetos sin EET. Estos resultados desafían la práctica actual de ignorar los EET al estimar las puntuaciones de los VA docentes. Los resultados también apoyan el uso de incentivos basado en los grupo cuando premiar maestros de secundaria de matemáticas y ELA de acuerdo al aumento del rendimiento de los estudiantes en estos dos temas.
Palabras clave: modelos de valor añadido; efectos indirectos maestro; evaluación de los maestros; la escuela media

Um Estudo Sobre o Valor Agregado e a Influência Indireta de Professores em Quatro Disciplinas Centrais nas Escolas de Ensino Médio

Resumo: O presente estudo analisou a existência, magnitude e impacto da influência dos professores (EET) de quatro áreas temáticas (por exemplo, matemática, artes da linguagem Inglês [ELA], ciências e estudos sociais) no o desempenho dos alunos em cada uma das quatro disciplinas em escolas secundárias. O autor realizou uma série de valor agregado (VA), utilizando-se de vários anos de resultados de testes de desempenho que avaliam os alunos nas séries 7 e 8 em um distrito escolar urbano no sul dos EUA. Os resultados fornecem evidências de que os professores de matemática e ELA contribuíram para o desempenho dos alunos em matemática e ELA. Professores ELA também mostraram EET no desempenho dos alunos em ciência (apenas no grau 8) Estudos sociais com efeito perto ou até mesmo superiores aos dos professores no grau 8. Os resultados também mostram que controlando do TSE da uma ligeira redução de variação e precisão dos ratings VA de ensino e alterou a classificação dos quartis de VA de professores para um grupo considerável de professores (11% - 25%). Em média, a porcentagem de professores cujo VA foi afetadas pelo controlo do TSE foi maior para indivíduos com casos de EET que para indivíduos sem EET. Estes resultados desafiam a prática corrente de ignorar a EET para estimar a pontuação VA dos professores. Os resultados também suporta o uso de grupo recompensa baseada em incentivos, quando os professores secundários de matemática e ELA acordo com o aumento o desempenho dos alunos sobre estas duas questões.
Palavras-chave: modelos de valor agregado; efeitos indiretos; avaliação de professores; ensino médio

Introduction

Value-added modeling (VAM), one class of statistical models used to estimate an individual teacher's or school's contribution to student achievement based on student test score growth between consecutive years, has become increasingly popular in the past decade (Amrein-Beardsley, 2008; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004b; Sanders & Horn, 1994). Moreover, teacher value-added (VA) scores have been used to evaluate teaching performance and make high-stakes decisions about teachers' compensation, bonus, and tenure (Harris, Sass, & Semykina, 2010; Winters, 2012).

Although VAM has been widely used to assess teaching in practice, many researchers are concerned about the quality of VA scores as a measure of teaching and the consequences of using VA scores for key decision-making (Amrein-Beardsley, 2008; Amrein-Beardsley, Collins, Polasky, & Sloat, 2013; Braun, Chudowsky, & Koenig, 2010). For example, VA models that do not properly control for student background characteristics might yield biased teacher VA estimates and make teachers who teach low-performing students more likely to receive low VA estimates than those who teach high-performing students (Braun, 2005; Harris, 2011; McCaffrey, Koretz, Lockwood, & Hamilton, 2004a). Some scholars are also concerned about the adequacy and quality of student assessment data and teacher-student linkage data used for VAM (McCaffrey, Sass, Lockwood, & Mihaly, 2009b; McCaffrey et al, 2004a). In addition, other researchers worry that teacher VA measures are not precise and stable enough to be used for key decision-making about educators such as bonus (Hill, 2009; McCaffrey et al., 2009b; Newton, Darling-Hammond, Haertel, & Thomas, 2010). Indeed, more research on VAM is needed to fully understand the assumptions of VAM, the properties of VAM estimates, various conditions of tests and data and decisions made during the modeling process that may affect VA estimates, especially when these results are used for high-stakes decision-making (Amrein-Beardsley et al., 2013; Harris et al., 2010; McCaffrey, Han, & Lockwood, 2009a; Reardon & Raudenbush, 2009).

One area of VAM that needs more research is about teacher spillover effects (TSEs). Prior studies have examined two types of TSEs. The first type of TSE refers to a teacher's influence on another teacher's students through peer interactions between these two teachers (Jackson & Breugmann, 2009). For example, two teachers teach mathematics at two different grade levels in the same school. Teacher A teaches at grade 3 and teacher B teaches at grade 4. Teacher A is less experienced than teacher B. They often plan lessons together. Teacher A always seeks advices from teacher B on how to design and improve mathematics instruction. Thus, teacher B may indirectly affect teacher A's students on their mathematics learning through coaching teacher A on instruction. Teacher B's influence on the mathematics achievement of teacher A's students is the first type of TSE.

The second type of TSE, which is the focus of this study, refers to a teacher's influence on his/her students' achievement in another subject taught by another teacher (Koedel, 2009). For instance, suppose four middle school teachers teach the same group of students on four subjects, including mathematics, English language arts (ELA), science, and social studies, with one teacher for each subject. Mathematics teachers directly affect students' mathematics achievement through teaching. ELA, science, and social studies teachers may also indirectly affect students' mathematics achievement through their teaching of their own subjects. ELA, science, and social studies teachers' effects on student mathematics achievement are the second type of TSE.[1] In the remainder of this paper, I refer to teachers of the same subject area as the test subject as "own-subject" teachers (i.e., mathematics teachers in this example) and teachers of subject areas different from the test subject as "cross-subject" teachers (i.e., ELA, science, and social studies teachers in this example).

Observing the second type of TSE requires that students receive instructions from different teachers on different subjects. Elementary school teachers often teach the same group of students on all subjects. Teachers in secondary schools tend to specialize in teaching one subject or closely related subjects such as mathematics and science to different groups of students (Jacob & Rockoff, 2011). In addition, tracking and other scheduling issues in secondary schools may also create imbalanced groupings of students so that teachers of other subjects may contribute to students' achievement in one subject area (Goldhaber, Goldschmidt, Sylling, & Tseng, 2011). Thus, TSEs[2] are more likely to exist in secondary schools than in elementary schools.

Multiple reasons may lead to findings of TSEs when students receive instruction from different teachers on different subjects. For instance, collaboration among teachers during prep time and instruction may result in overlaps in the knowledge and skills students learn from teachers of different subjects (Strauss, 2013). Overlaps in the curricula for two closely related subjects may also lead to common knowledge and skills students learn from different teachers. For instance, if a science curriculum requires students to extensively practice certain mathematics knowledge and skills that are tested in a mathematics test, science teachers using this curriculum may show TSEs on students' mathematics achievement. Moreover, TSEs may happen because different subject teachers contribute to developing the same set of cognitive skills that are important for students' performance on a test. For example, ELA teachers may have TSEs on students' achievement in other subjects because students' reading and language skills are important for learning in almost all subjects (Abedi & Lord, 2001; Chang, Singh, & Filer, 2009; O'Reilly & McNamara, 2007). Meanwhile, teachers of mathematics and other subjects may also have TSEs on students' ELA achievement because all teachers may affect a common set of knowledge and skills such as working memory and visual perception that are important for students' performance on any test (Hecht, Torgesen, Wagner, & Rashotte, 2001). In addition, test design may also play a role in the findings of TSEs in VAM results. For example, when a science test requires a certain level of reading skills to be able to answer its questions, VA analysis may show TSEs for ELA teachers on student science achievement. In reality, these possible contributors of TSEs are not necessarily exclusive of each other, which makes it difficult to identify the main reason of TSEs.

Several studies have examined TSEs at the high school level. For example, Aaronson, Barrow, and Sander (2007) analyzed 9th graders' mathematics and ELA test scores on the state achievement tests in the Chicago Public Schools and found TSEs for both mathematics and ELA teachers. They reported that mathematics and ELA teachers had an effect of 0.17 and 0.08 standard deviations[3] on students' mathematics achievement, and an effect of 0.15 and 0.12 standard deviations on students' reading achievement, respectively. Buddin and Zamarro (2009) examined state achievement test scores of students at grades 9–11 from the Los Angeles Unified

[1] Please note that both types of TSEs examined in this study are different from the teacher persistent effects examined in prior studies (McCaffrey et al., 2004a; Lockwood, McCaffrey, Mariano, & Setodji, 2007), which refer to a teacher's continuing effect on his/her students as these students move on to another grade level and taught by other teachers.
[2] From now on, the term TSE refers to the second type of TSE. Analyses of such TSEs require there is adequate variation in the class roles for teachers across subject areas, which may not be applicable for small secondary schools.
[3] Scores used in this study were grade equivalents.

School District and also found evidence of TSEs for both mathematics and ELA teachers at the high school level. Specifically, they found that mathematics and ELA teachers had an effect of 0.25 and 0.24 standard deviations on students' mathematics achievement, and an effect of 0.14 and 0.17 on students' ELA achievement, respectively. Moreover, Koedel (2009) studied mathematics, ELA, science, and social studies teachers' effects on 9th–11th graders' reading test scores on the Stanford 9 reading achievement test in the San Diego Unified School District. He reported that ELA and mathematics teachers had an effect of 0.07 and 0.06 standard deviations, respectively, on students' reading achievement. Science and social studies teachers did not show TSEs on students' reading achievement in his study. In addition, Jackson (2012) analyzed mathematics and ELA teachers' effects on 9th graders' Algebra I and English I end-of-course exam scores from over 600 secondary schools in North Carolina. He did not find any TSEs for teachers of either subject.

Two studies that found evidence of TSEs also examined the impact of controlling for TSEs on the variation of own-subject teachers' effects. Aaronson, Barrow, and Sander (2007) found that, for both test subjects (i.e., mathematics and ELA), controlling for TSEs reduced the standard deviations of own-subject teachers by 0.02 standard deviations. Buddin and Zamarro (2009) reported that controlling for TSEs reduced 0.01 standard deviations for mathematics teachers' effects and 0.04 standard deviations for ELA teachers.

These studies provided important findings regarding the existence and magnitude of TSEs, with most studies finding evidence of TSEs that were meaningfully large. All studies were conducted at the high school level. Most of them focused on mathematics teachers' TSEs on student ELA achievement and ELA teachers' TSEs on student mathematics achievement. Two studies examined the influence of controlling for TSEs on own-subject teachers' effects and found small impact of controlling for TSEs on the variation of own-subject teachers' effects.

Although the situations that may lead to TSEs are common in secondary schools and prior research has shown evidence of TSEs at the high school level, current practices in VAM ignore TSEs and attribute students' achievement growth on a test subject only to the own-subject teachers. Such practices may be acceptable for elementary teachers as the student-teacher assignments are mainly one-to-one in elementary schools. However, ignoring TSEs at the secondary level may lead to biased teacher VA estimates, which invalidate both within- and across-school comparisons of teachers' VA scores and the key decisions made based on these estimates, such as teacher compensation and bonuses.

With the implementation of the Common Core State Standards (CCSS; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), the magnitude of TSEs might increase as the CCSS asks for more collaboration among teachers across subjects. Moreover, the practices of evaluating teaching performance and rewarding high-performing teachers with monetary bonuses based on teachers' VA scores have also become increasingly popular, such as programs supported by the Race to the Top Fund (U.S. Department of Education, 2014). With the potential increase in the magnitude of TSEs and the use of teachers' VA scores in high-stakes decisions about educators, ignoring TSEs when estimating teachers' VA scores has greater potential to threaten the validity of estimated VA scores and the decisions made based on these VA scores. All these factors make it necessary to conduct more research on the existence and impact of TSEs on teachers' VA scores.

In this study, I examined the existence of TSEs across teachers of four subject areas, including mathematics, ELA, science, and social studies, on student achievement in each of the four subjects at the middle school level and the impact of controlling for TSEs on own-subject teachers' VA measures. Specifically, I addressed the following three research questions: (1) Do

TSEs exist across teachers of mathematics, ELA, science, and social studies on any of the four core test subjects at the middle school level? (2) If TSEs exist, what are the effect sizes of own- and cross-subject teachers on each test subject? And (3) if TSEs exist, how does controlling for TSEs affect the variation, precision, and relative stance of own-subject teachers' VA scores?

When estimating teachers' VA scores, I use models similar to what is commonly used in practice, which estimate teachers' annual effects and control for teacher or classroom aggregates of student demographic and achievement variables (Bill & Melinda Gates Foundation, 2010). By using models that are commonly used in practice, I expect to understand the magnitude of TSEs and the consequence of ignoring TSEs when estimating own-subject teachers' VA scores in the common practice. Results from this study contribute to better understanding of different subject teachers' joint contributions to student achievement and may help decision-makers develop better teaching evaluation and pay-for-performance programs for teachers at the secondary school level.

Data

Data used in this study came from an urban school district in the Southern United States. This district served a student population of 70,000 to 80,000 students annually, which had about 50% African American, 36% White, and 11% Hispanic students. Ten percent of the students were English language learners (ELL). Over 60% of the student population was eligible for free and reduced-price lunch (FRPL). The district's performances on the statewide mathematics and ELA tests were below the state averages.

Data used for analyses followed students at grades 7 and 8 and their mathematics, ELA, science, and social studies teachers from 2006–07 to 2008–09. This data set included students' demographic characteristics such as gender, race, and FRPL; students' test scores on the state mathematics, ELA, science, and social studies tests from 2003–04 to 2008–09; and teacher-student linkage data on each of the four subjects each year from 2006–07 to 2008–09.

Students' test scores on the state mathematics and ELA tests were presented on a developmental scale with scores linked across grades and years from the 2003–04 school year forward. Scores on the science and social studies assessments were not vertically linked or linked across school years but were scaled to have the same mean and variance at each grade level and year. To make results comparable across subjects, I converted students' scale scores on the four state achievement tests to rank-based z scores and used them in the analysis (McCaffrey et al., 2009a).

I restricted the student analytical sample to those who met a set of criteria. Specifically, On each of the four subjects, students had to be taught by the same single teacher for 90% or more of the target school year and had five or more peers taught by the same teacher on the same subject. Students had to have test scores on all four subjects in the target year and the immediate previous school year. Each teacher included in the analysis had to be linked to at least five students[4] on the subject(s) they taught. In total, the analytical sample included 13,663 students at grades 7–8 and 636 linked teachers.

Restricting the sample to teachers with at least five students per subject and students in classes with at least five students on each of the four subjects did not reduce the student sample substantially. It led to a change of five to seven percent for the student sample. The final student analytical sample was slightly more advantaged than the population of students at grades 7–8 in

[4] I conducted sensitivity analyses with other choices of the threshold number, ranging from five to ten. Results showed that the overall findings about TSEs did not change with the choice of the threshold.

the district but was still representative in terms of demographic characteristics. These restrictions reduced the teacher sample by 2–20%. Most of the teachers excluded from the analysis might be special education or ELL teachers. Overall, this set of restrictions might have led to more homogeneous samples than their respective populations and have underestimated the variation of teacher effects in middle schools.

I analyzed data by grade-year groups. This is consistent with the common practice that focuses on estimating teachers' contributions to student achievement gains in a single year rather than estimating the same teacher's contributions to student achievement gains using multiple years of data. In total, I analyzed six grade-year groups (see Table 1). The number of students included in each grade-year group was about 3,000. The average number of students used to estimate an individual teacher's VA score ranged from 38–57 across four subjects at two grade levels.

The majority of students (90%) in the analytical sample were taught by single-subject teachers on each of the four tested subjects. The percentage of students who were taught by any particular type of multi-subject teachers was no greater than 3% within each grade-year group. Given the small percentage of students taught by multi-subject teachers, I treated multi-subject teachers as single teachers in the analysis.[5]

Analysis Models

To answer Research Question 1, I used fixed teacher effect VA models to test whether teachers of each subject area had significant contributions to students' achievement in any of the four test subjects. If results showed significant teacher effects for any cross-subject teachers on a test subject, that was considered evidence of TSEs. To answer Research Questions 2 and 3, I used random teacher effect VA models to estimate the variation of teacher effects for teachers of each subject area on each test subject and examined changes in the variation, precision, and relative stance of own-subject teachers' VA scores due to controlling for TSEs.

[5] I also conducted the analyses without students taught by multi-subject teachers. Findings remained the same.

Table 1
Descriptive Statistics for Teachers and Students Included in Each Grade-Year Group

		Student Demographics				Mean (SD) of Test Scores and [Number of Teachers] for Each Test Subject			
Grade	**Year**	**N**	**% White**	**% Black**	**% FRPL**	**Math**	**ELA**	**Sci.**	**Soc.**
7	1	2,822	35%	50%	58%	.18	.18	.1	.13
						(.91)	(.91)	(.94)	(.94)
						[72]	[67]	[58]	[65]
	2	2,840	35%	48%	57%	.26	.25	.24	.26
						(.97)	(.97)	(.98)	(.98)
						[71]	[64]	[52]	[63]
	3	3,080	35%	46%	63%	.17	.15	.14	.15
						(.94)	(.93)	(.94)	(.94)
						[65]	[82]	[49]	[63]
8	1	2,842	38%	51%	57%	.18	.22	.13	.15
						(.93)	(.93)	(.93)	(.96)
						[69]	[64]	[50]	[69]
	2	2,860	36%	48%	57%	.16	.19	.16	.16
						(.9)	(.92)	(.94)	(.95)
						[65]	[59]	[51]	[62]
	3	2,966	34%	47%	61%	.11	.1	.12	.11
						(.94)	(.92)	(.93)	(.92)
						[58]	[78]	[46]	[55]

Notes.
1. Year 1 = 2006–07; Year 2 = 2007–08; Year 3 = 2008–09.
2. FRPL = free and reduced-price lunch; ELA = English Language Arts.
3. The number of schools included in the analysis ranged from 32 to 36.
4. The last four columns show the means and standard deviations (in parentheses) of the rank-based z scores on four test subjects for the analytical sample and the corresponding number of teachers of four subjects for each grade-year group (in brackets).

Fixed Teacher Effect Model

Model 1 shows the fixed teacher effect model that includes teachers of all subject areas:

$$Y_{ijt} = X_{ijt}^{T}\lambda + \sum_{p=1}^{3} Y_{ij(t-p)}^{T}\beta_p + D_{ijt}^{school}\delta_J + D_{ijt}^{M(mathematics)}\theta_M + D_{ijt}^{L(language)}\theta_L + D_{ijt}^{Q(science)}\theta_Q + D_{ijt}^{U(social\ studies)}\theta_U + \varepsilon_{ijt} \quad (1)$$

where *i* index student, *j* index school, and *t* index year (t = 2007, 2008, 2009). Y_{ijt} represents the rank-based z score on one of the test subjects for student *i* in school *j* in year *t*. X_{ijt}^{T} is a vector of student demographic characteristics, including gender, race, eligibility for FRPL, and status on ELL and special education. $Y_{ij(t-p)}^{T}$ $(p = 1, 2, 3)$ is a vector of rank-based z scores on four subjects for student *i* in school *j* in *p* years prior to the target school year. λ and β_p $(p = 1, 2, 3)$

are vectors of parameters to be estimated for student demographic variables and prior achievement. D_{ijt}^{school}, $D_{ijt}^{M(mathematics)}$, $D_{ijt}^{L(language)}$, $D_{ijt}^{Q(science)}$, and $D_{ijt}^{U(social\ studies)}$ are indicator variables for the school and teachers of four subject areas for student *i* in school *j* in year *t*. δ_J represents fixed school effect for school j. $\theta_L, \theta_M, \theta_Q$, and θ_U are fixed teacher effects to be estimated from the model. ε_{ijt} is residual error assumed to be mean zero, independently and identically distributed across students.

I applied the fixed teacher effect model on students' rank-based z scores on each of the four test subjects in each grade-year group. For each test subject in each group, I first fit the model with teachers of all four subject areas (referred to as full model). Then I excluded teachers of one subject area and reran the model (referred to as reduced model). For each full model, I fit four reduced models. Next, I compared the results of the full model with those of each reduced model and used F-test to examine the significance of effects for the group of teachers excluded from each reduced model. The results of these analyses showed whether teachers of a particular subject area had significant effects on students' achievement in a test subject in a particular school year. Finally, I pooled the F-test results across years for teachers of each subject area on each test subject at each grade level using Fisher's combined probability test (Fisher, 1925).

The reason to aggregate results across years for each grade level was that although the fixed teacher effect model was fit to allow for flexibility in the modeled relationships among test scores and other variables in each grade-year combination, the year-to-year variance in these relationships was not really of interest because the relationships among teacher effects were not expected to have systematic annual variation. Conducting the analyses by grade was also sensible because the degree of TSEs may vary by grade in ways that were persistent across years and might be worth understanding. Thus, results were aggregated across years for each grade.

I applied the Benjamini and Hochberg (1985) method to control for false discovery rate (FDR) at 5% across all tests to adjust for multiple comparisons. If cross-subject teachers of any subject area showed significant effects on students' achievement in a particular test subject after the adjustment for multiple comparisons, that was considered as evidence of TSEs.

Random Teacher Effect Model

Although results from the fixed teacher model showed whether teachers of a particular subject area had significant effects on student achievement in a test subject, they did not provide information about the effect sizes of the own-subject and cross-subject teachers. Such information is important for understanding the relative magnitude of contributions teachers of different subject areas made to student achievement in each test subject. The random teacher effect model naturally lends itself to estimate the variation of teacher effects. Although it is also possible to estimate the variation of teacher effects based on results from the fixed teacher effect model, there are various pitfalls to mis-estimate these variance components under different decisions of dealing with collinearity in the fixed teacher effect model (McCaffrey, Lockwood, Mihaly, & Sass, 2012). Thus, I implemented the random teacher effect model to estimate the variation of teacher effects for teachers of each subject area on student achievement in each test subject.

As peer characteristics may be associated with the potential achievement growth observed for individual students, many VA models used in practice control for peer characteristics (Bill & Melinda Gates Foundation, 2010). Given that this study focused on examining TSEs in the context of common practice of VAM, it is necessary to control for peer characteristics when estimating variation of teacher effects. However, in the annual model that has no repeated measures on teachers, including both fixed teacher effects and teacher-level aggregates makes the model un-identified. Including peer characteristics in the random teacher effect model does not pose a problem for the estimation of teacher effects. Therefore, I included peer characteristics in the random teacher effect model.[6] Model 2 shows the random teacher effect model used to estimate the variance of teacher effects:

$$Y_{ijt} = X_{ijt}^{T}\lambda + \sum_{p=1}^{3} Y_{ij(t-p)}^{T}\beta_p + C_{ij(t-1)}^{T}\gamma + D_{ijt}^{school}\delta_J + D_{ijt}^{L(language)}\zeta_L + D_{ijt}^{M(math)}\zeta_M + D_{ijt}^{Q(science)}\zeta_Q + D_{ijt}^{U(social\ studies)}\zeta_U + \varepsilon_{ijt} \quad (2)$$

The notations for most of the model components remain the same as those in the fixed teacher effect model, including the dependent variable (Y_{ijt}), student demographic characteristics (X_{ijt}^{T}) and prior test scores $(Y_{ij(t-p)}^{T}\ (p = 1, 2, 3))$ and their associated coefficients (λ and β_p); indicator variables for schools (D_{ijt}^{school}) and fixed school effects (δ_J); indicator variables for teachers ($D_{ijt}^{M(mathematics)}, D_{ijt}^{L(language)}$, $D_{ijt}^{Q(science)}$, and $D_{ijt}^{U(social\ studies)}$); and residual error (ε_{ijt}). $C_{ij(t-1)}^{T}$ is a vector of teacher-level achievement and socio-economic status variables, including teacher-level average rank-based z scores on four subjects and the percentage of students eligible for FRPL in the year prior to the target school year. γ is a vector of parameters for the teacher-level aggregated achievement and socio-economic status variables. $\zeta_L,\ \zeta_M, \zeta_Q,$ and ζ_U are random teacher effects for teachers of each subject area. The variations of these teacher effects are the parameters of interest to be estimated from the model.

I applied the random teacher effect model to each test subject for each grade-year group and collected estimated variances of teacher effects for teachers of each subject area. Then I calculated the average variance of teacher effects for teachers of each subject area across years for each test subject at each grade level. The square root of the average variance of teacher effects was used as the effect size for teachers of each subject area on each test subject at each grade level.

To answer Research Question 3, I examined changes in the estimated standard deviations of own-subject teachers' effects, the standard error of individual own-subject teachers' VA scores, and the quartile rankings of own-subject teachers' VA scores before and after controlling for TSEs.[7] As quartile rankings of teachers' VA scores are often used in the practice to make decisions regarding teachers' compensation and bonus (Aaronson et al., 2007), studying changes in teachers' quartile rankings is useful to gauge the potential impact of TSEs on the input for high-stakes decisions made for teachers. To do this, I first collected the estimated VA scores for individual own-subject teachers from the random teacher effect modeling results. Then I pooled teachers' VA scores for the same subject at the same grade level across years and obtained their quartile rankings. I conducted the first and second steps with and without controlling for the effects of all three cross-subject teachers for each test subject.

---

[6] I also implemented random teacher effect models without teacher-level aggregates. Results showed similar relative sizes of teacher effects for four types of teachers on each test subject at both grade levels.

[7] Sensitive analyses using a random teacher effect model without the fixed school effect showed the same overall findings about the changes in the variation, precision, and relative stance of teachers' VA scores.

Finally, I examined the percentages of teachers who changed their quartile rankings due to the control of TSEs and the number of quartiles changed.

Student sorting poses a potential threat to the validity of teachers' VA scores. Prior research showed that controlling for students' test scores in the same subject in the immediate prior year substantially reduced the bias in teachers' VA estimates (Chetty, Friedman, & Rockoff, 2013). Both types of VA models used in this study controlled for a rich set of student demographic characteristics and test scores in four test subject areas in up to three prior years. In addition, controlling for peer effects helps mitigate the potential influence of possible student sorting on teacher VA estimates (Sass, Semykina, & Harris, 2014). These model specification strategies helped minimize the potential influence of student sorting on the validity of teacher VA estimates and findings about TSEs.

Results

Research Question 1: Do TSEs exist across teachers of mathematics, ELA, science, and social studies on any of the four core test subjects at the middle school level?

Table 2 shows significant effects for the own- and cross-subject teachers found in the results pooled across years for each test subject at each grade level.[8][9] Dark cells represent significant effects for own-subject teachers. Grey cells represent TSEs found (i.e., significant effects of associated cross-subject teachers on students' achievement in the corresponding test subject). For instance, the dark cell corresponding to mathematics teachers' effects on students' mathematics achievement at grade 8 indicates mathematics teachers had significant effects on eighth graders' mathematics achievement. The grey cell corresponding to ELA teachers on students' mathematics achievement at grade 8 indicates ELA teachers had significant TSEs on eighth graders' mathematics achievement.

Results in Table 2 show that own-subject teachers had significant effects on student achievement on all four test subjects at both grade levels. Meanwhile, TSEs were found at both grade levels, although the specific TSEs found varied by grade and subject. At grade 7, mathematics teachers showed TSEs on students' ELA achievement. ELA teachers had TSEs on student achievement in social studies. At grade 8, ELA teachers had TSEs on student achievement in the other three subjects. Social studies teachers also showed TSEs on student science achievement.

---

[8] When I applied the Benjamini and Hochberg method to adjust for multiple comparisons, the critical p-values used to compare with the observed p-values change with the rankings of the observed p-values among all the tests conducted. Therefore, providing the specific p-values found for each cell does not help readers understand the results. Thus, I used color schemes to indicate whether significant teacher effects were found for the own- or cross-subject teachers on a test subject. The largest p-value that was still significant after the adjustment for multiple comparisons was 0.008.

[9] The presentation of findings from the fixed teacher effect model focused on the significance of the teacher effects for teachers of four subject areas, as this is the parameter of interest. Estimated parameters for other covariates and their significance were not presented in the paper, but are available from the author upon request.

Table 2
Significant Teacher Effects Found by the Fixed Teacher Effect Model

Grade	Teacher Subject	Test Subject: Mathematics	ELA	Science	Social Studies
7	Mathematics	■	▒		
7	ELA		■		▒
7	Science			■	
7	Social Studies				■
8	Mathematics	■			
8	ELA	▒	■	▒	▒
8	Science			■	
8	Social Studies			▒	■

Notes.
■ represents significant effects found for the own-subject teachers.
▒ represents TSEs (i.e., significant effects found for the corresponding cross-subject teachers on the associated test subject).

Research Question 2: If TSEs exist, what are the effect sizes of own- and cross-subject teachers on each test subject?

Table 3 shows the estimated standard deviations of the effects for teachers of each subject area on student achievement in each test subject. Results in Table 3 show that own-subject teachers had larger effect sizes than cross-subject teachers on most test subjects at two grade levels. The effect sizes of own-subject teachers ranged from 0.1 to 0.2 standard deviations. The effect sizes of most cross-subject teachers ranged from 0.03 to 0.08 standard deviations, except that ELA teachers' TSEs on student achievement in the other three subjects at grade 8 were in the range of 0.12–0.17. The relative effect sizes between own-subject and cross-subject teachers who showed TSEs varied by grade. At grade 7, the effect sizes of cross-subject teachers with TSEs were about one-half of those of own-subject teachers. For instance, mathematics teachers had an effect of 0.06 standard deviations on students' ELA achievement, compared with 0.1 standard deviations for ELA teachers. ELA teachers had an effect of 0.08 standard deviations on student achievement in social studies, compared with 0.17 standard deviations for social studies teachers. At grade 8, ELA teachers had an effect size that was close to or even greater than those of own-subject teachers on student achievement in the other three subjects. ELA teachers' effects on student achievement in mathematics, science, and social studies was 0.12, 0.12, and 0.17 standard deviations, respectively, compared with 0.16, 0.13, and 0.12 standard deviations for the own-subject teachers on each test subject. The effect size of social studies teachers' TSEs on student science achievement (i.e., 0.07 standard deviations) was one-half of that of the own-subject teachers (i.e., 0.13 standard deviations).

Research Question 3: If TSEs exist, how does controlling for TSEs affect the variation, precision, and relative stance of own-subject teachers' VA scores?

Results showed that controlling for TSEs had small impact on the variation and precision of own-subject teachers' VA scores. Specifically, controlling for TSEs reduced the variation of own-subject teachers' VA scores by less than 0.01 standard deviations for teachers at grade 7 and 0.01–0.05 standard deviations for teachers at grade 8. The standard errors of own-subject teachers' VA scores increased by 0.01–0.03 across subjects at two grade levels after controlling for TSEs.

Table 3
Standard Deviations of Teacher Effects Estimated from the Random Effect Model

	Teacher	**Test Subject**			
Grade	**Subject**	**Mathematics**	**ELA**	**Science**	**Social Studies**
7	Mathematics	0.16	0.06	0.07	0.08
7	ELA	0.05	0.1	0.05	0.08
7	Science	0.03	0.04	0.19	0.05
7	Social Studies	0.04	0.04	0.04	0.17
8	Mathematics	0.16	0.06	0.04	0.06
8	ELA	0.12	0.16	0.12	0.17
8	Science	0.03	0.03	0.13	0.05
8	Social Studies	0.05	0.04	0.07	0.12

Note. Shaded cells indicate TSEs found in the fixed teacher effect model for the corresponding cross-subject teachers.

Results also showed that controlling for TSEs affected the quartile rankings of teachers' VA scores for a non-negligible proportion of teachers. Table 4 shows the total number of own-subject teachers included in the analysis for each test subject at each grade level, the percentage of total affected teachers, and the proportion of teachers in each type of quartile ranking change after controlling for TSEs. For instance, for mathematics at grade 7, 208 VA scores were ranked across three years. Among them, 24.5%, 20.7%, 20.2%, and 24% remained in Quartiles 1–4, respectively, after controlling for TSEs. The remaining 11% changed their quartile rankings as the result of controlling for TSEs, with 1% moving between Quartiles 1 and 2, 2% moving between Quartiles 3 and 4, and 8% moving between Quartiles 2 and 3.

Across four subjects at two grade levels, controlling for TSEs changed the quartile rankings of VA scores for 11%–25% of teachers, with the percentage of affected teachers varying by subject and grade level. Test subjects on which TSEs were detected by the fixed teacher effect model tend to have a higher percentage of affected teachers than those on which no TSEs were detected. Social studies at grade 8, on which ELA teachers showed TSEs and had a greater effect size than that of social studies teachers, had the highest percentage of affected teachers (25%). Science at grade 8, on which ELA and social studies teachers showed TSEs and the effects of ELA teachers were close to those of science teachers, had the second highest percentage of affected teachers (21%). Three other subjects on which TSEs were found, including ELA and social studies at grade 7 and mathematics at grade 8, had 16%, 11%, and 18% of affected teachers, respectively. The three subjects on which no TSEs were found by the fixed teacher effect model, including mathematics and science at grade 7 and ELA at grade 8, had 11%–14% of teachers who changed their quartile rankings due to controlling for TSEs. On average, test subjects without TSEs had lower percentages of affected teachers than those with TSEs.

The majority of affected teachers changed one quartile. A small percentage of affected teachers changed two quartiles, most of which happened at grade 8. Most of the affected teachers (70%–90%) were ranked in the second and third quartiles before controlling for TSEs. The remaining affected teachers were roughly equally distributed between the lowest and highest quartiles.

Table 4
Proportion of Teachers in Each Quartile Before and After Controlling for TSEs

Before Controlling for TSEs									
		After Controlling for TSEs							
		Mathematics (N=208, Affected=11%)				ELA (N=213, Affected=16%)			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Grade 7	Q1	0.245	0.005	0.000	0.000	0.230	0.019	0.000	0.000
	Q2	0.005	0.207	0.038	0.000	0.019	0.183	0.047	0.000
	Q3	0.000	0.038	0.202	0.010	0.000	0.047	0.188	0.014
	Q4	0.000	0.000	0.010	0.240	0.000	0.000	0.014	0.239
		Science (N=159, Affected=13%)				Social Studies (N=191, Affected=11%)			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	Q1	0.226	0.019	0.000	0.000	0.225	0.021	0.000	0.000
	Q2	0.019	0.195	0.038	0.000	0.021	0.209	0.021	0.000
	Q3	0.000	0.031	0.208	0.013	0.000	0.021	0.220	0.010
	Q4	0.000	0.006	0.006	0.239	0.000	0.000	0.010	0.241
		Mathematics (N=192, Affected=18%)				ELA (N=201, Affected=14%)			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Grade 8	Q1	0.229	0.021	0.000	0.000	0.234	0.015	0.000	0.000
	Q2	0.016	0.182	0.052	0.000	0.015	0.194	0.040	0.000
	Q3	0.005	0.042	0.182	0.021	0.000	0.040	0.194	0.015
	Q4	0.000	0.005	0.016	0.229	0.000	0.000	0.015	0.239
		Science (N=147, Affected=21%)				Social Studies (N=186, Affected=25%)			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	Q1	0.238	0.007	0.000	0.000	0.210	0.032	0.005	0.000
	Q2	0.007	0.156	0.082	0.007	0.038	0.156	0.059	0.000
	Q3	0.000	0.088	0.156	0.007	0.000	0.065	0.156	0.027
	Q4	0.000	0.000	0.014	0.238	0.000	0.000	0.027	0.226

Discussion

This study examined whether TSEs existed among mathematics, ELA, science, and social studies teachers on student achievement in these four test subjects at grades 7 and 8 when analyzing teachers' contributions to student learning, based on student test scores on the state achievement tests. Results showed mathematics teachers affected students' ELA achievement at grade 7 and ELA teachers affected students' mathematics achievement at grade 8. Mathematics teachers did not show TSEs on student achievement in science or social studies. ELA teachers showed TSEs on student achievement in social studies at both grade levels and TSEs on student achievement in both science and social studies at grade 8. The effect sizes of ELA teachers on student achievement in science and social studies were close to or greater than those of the own-subject teachers at grade 8. In addition, neither science nor social studies teachers showed TSEs on student achievement in mathematics or ELA. Social studies teachers showed TSEs on students' science achievement only at grade 8.

Findings of mathematics teachers' TSEs on students' ELA achievement and ELA teachers' TSEs on students' mathematics achievement are consistent with results from most previous studies on TSEs at the high school level (Aaronson et al., 2007; Buddin & Zamarro, 2009; Koedel, 2009). These studies used test scores on state and district-level standardized achievement tests from different research sites and applied VA models that varied in the number of prior test scores to examine TSEs of mathematics and ELA teachers on student ELA and mathematics achievement, respectively. The fact that multiple different studies have found evidence of TSEs for mathematics and ELA teachers on student achievement in these two subjects suggests that mathematics and ELA teachers jointly contribute to student learning in both subjects.

ELA teachers' contributions to students' achievement in the other three subjects may have resulted from the important role of reading and language skills in the learning of other subject areas (Abedi & Lord, 2001; O'Reilly & McNamara, 2007; Chang et al., 2009). Mathematics teachers' contributions to developing students' basic cognitive skills such as working memory may be associated with their TSEs on student ELA achievement (Hecht et al., 2001; Jordan, Kaplan, & Hanich, 2002). Other factors, such as collaboration, overlap in the curricula used for different subjects, test design factors, and the length of students' exposure to teachers of different subjects, may also have contributed to the findings of TSEs in this study. For instance, the extent to which reading and language skills affect students' performance on the tests of the other three subjects may have contributed to the varying effects of ELA teachers on student achievement in those subjects. It is also possible that ELA and mathematics teachers effect student achievement in other subjects because they have more exposure to students than teachers of other subjects. Unfortunately, data needed to further examine the potential contributions of these factors to the findings of TSEs, such as state achievement test forms and student course enrollment data, were unavailable for analysis in this study.

Findings that science or social studies did not have TSEs on students' mathematics or ELA achievement are consistent with results from Koedel (2009), which did not find TSEs of science or social studies teachers on students' ELA achievement. None of the previous studies on TSEs has examined mathematics or ELA teachers' TSEs on student achievement in science or social studies, or TSEs of science and social studies teachers on student achievement in these two subjects. Future research is needed to test the robustness of these findings using data from other sites and test scores from other types of student achievement assessments.

Findings of TSEs varied between two middle school grade levels in this study. Multiple reasons might have contributed to the differences in the findings of TSEs between two grade levels. For instance, teachers at two grade levels may have different levels of collaboration during prep time and instruction. The extent to which different curricula overlapped with each other may also vary by grade. In addition, the degree to which the mathematics, science, and social studies tests rely on students' reading and language skills may be different between two grade levels. However, this study cannot analyze reasons for differential findings between grade levels due to lack of data on potential contributors.

This study also investigated how controlling for TSEs affected the variation, precision, and relative stance of own-subject teachers' VA scores. Results showed that the impact of controlling for TSEs varied by the measures examined, with small impact on the variation and precision of teachers' VA scores and non-negligible impact on the quartile rankings of teachers' VA scores. The small impact on the variation of teachers' VA scores is consistent with findings from two prior studies (Aaronson, Barrow, & Sanders, 2007; Buddin & Zamarro, 2009). This finding, together with the slight decrease in the precision of teachers' VA scores, may partly justify why TSEs are not taken into consideration in the current practice of estimating teachers' contributions to student learning using value-added modeling. Moreover, challenges in collecting longitudinal student achievement data and comprehensive and accurate student-teacher linkage data may also have contributed to the difficulty in accounting for TSEs in the common practice of value-added modeling (McCaffrey et al., 2009b).

However, findings that controlling for TSEs led to changed quartile rankings for 11%–25% of teachers' VA scores indicates that TSEs may warrant more attention when it comes to using teachers' VA scores for key decisions, such as performance evaluation and bonus decisions. Although test subjects with no TSEs also had certain percentages of affected teachers after controlling for TSEs, this may have resulted from the known instability of VA estimates and associated rankings, which suggests that the true percentage of affected teachers might be smaller than what was observed in this study. The small percentage of teachers who changed for two quartiles may also justify the argument that there is no need to worry about TSEs in the current practice of teacher value-added modeling. However, findings that, on average, the percentages of affected teachers were greater for subjects with TSEs than those without TSEs suggest that TSEs had some extra influence on teachers' quartile rankings in addition to that resulted from the known instability of VA estimates and associated rankings. Such influence warrants careful examination when teachers' VA scores are used for high-stakes decision-making.

Overall, findings of TSEs in this study suggest that the contributions of ELA and mathematics teachers at the middle school level may go beyond the specific subject they teach, especially for ELA teachers. Results also suggest careful analysis of TSEs when using teachers' VA scores for important decisions about educators. Results from this study have several implications for the design of teacher evaluation and pay-for-performance programs based on teachers' VA measures.

First, findings from this study provide support for, yet also challenge, the current practice of how to estimate teachers' VA scores. Findings of significant effects for own-subject teachers on all four subjects at both grade levels suggest that it is reasonable to attribute student achievement growth in a particular subject to the own-subject teachers when using VA scores to evaluate teaching. However, findings of TSEs challenge the current practice of evaluating teaching performance based on student achievement growth only in subjects a teacher teaches and not controlling for the potential influence of teachers of other subject areas. Both types of practices have the potential to produce biased teacher VA estimates and invalidate the decisions made based on such estimates. This study provided evidence of the existence, magnitude, and impact of TSEs at the middle school level. However, results from this study are still insufficient to draw solid conclusions about the magnitude and impact of TSEs on teachers' VA scores on a large scale. As future studies on TSEs provide more evidence about the prevalence of TSEs and the impact of TSEs on individual teachers' VA scores, decision-makers need to consider whether it is necessary to account for TSEs when estimating teachers' VA scores, especially for mathematics and ELA teachers.

Findings of TSEs in this study also provide support for group-based incentive pay programs for mathematics and ELA teachers. Koedel (2009) did not find significant effects of science and social studies teachers on student reading achievement and did not think group-based incentive was strongly supported by his findings. This study also did not find any evidence that science or social studies teachers had significant TSEs on student mathematics or ELA achievement. However, results from this and prior studies showed mathematics and ELA teachers jointly contribute to student achievement in both subjects. These results suggest group-based incentive pay programs might be reasonable for mathematics and ELA teachers when rewarding teachers based on their students' achievement in these two subjects.

This study has several unique contributions to the field of research on TSEs. First, this is the first study that examines TSEs at the middle school level. It fills in a hole in the existing literature about TSEs in secondary schools. Second, it goes beyond the subject areas commonly studied in prior research (i.e., mathematics and ELA) and, compared with other studies of TSEs, provides the most comprehensive picture of TSEs across four core subject teachers on student achievement in all four subjects at the secondary school level. Third, its findings about the impact of controlling for TSEs on individual teachers' VA scores and associated rankings contribute to raising decision-makers' awareness about the importance of accounting for teachers' joint contributions to student learning in key decisions such as performance evaluation.

It is necessary to note the limitations of this study. Such limitations ask for caution when interpreting the findings of this study. For instance, this study drew on student test scores on the state achievement tests to study TSEs. State achievement tests, as any other standardized achievement tests, have their limitations to fully and accurately capture student learning and measure teachers' contributions to student learning (Koretz, 2002). On one hand, analysis results based on state achievement tests may overestimate the joint contributions that teachers of different subjects may have on student achievement, as test design factors may affect findings of TSEs in value-added analysis. On the other hand, analysis results based on state achievement tests may also underestimate the joint contributions of different subject teachers as standardized achievement tests are limited in their capacity to measure teachers' impact on factors such as motivation, learning effort, and persistency, which have been found to be closely related to student outcomes (Duckworth, Peterson, Matthews, & Kelly, 2007).

Results from this and previous studies on TSEs suggest three areas of research for future studies. First, future research may continue studying the existence and prevalence of TSEs by using data on multiple test subjects and from various sites. Although existing studies found evidence of TSEs, these results vary by site, subject, and grade level. It is important to understand the extent to which the current findings of TSEs can be generalized to other districts in the country. More studies on TSEs using data from different sites across the country will be helpful to understand how widely TSEs exist, the test subject areas that are likely to have TSEs, subject areas of teachers that are likely to demonstrate TSEs, and the magnitude of TSEs commonly observed for different subject teachers on different test subject areas.

Second, it is important for future research to investigate the impact of controlling for TSEs on individual teachers' VA estimates when such scores are used for high-stakes decision-making. If TSEs exist in VAM results but barely affect teachers' VA estimates or any associated measures used for high-stakes decision-making about teachers, there is not much to worry about TSEs. However, results from this study suggest that TSEs have some influence on teachers' VA score rankings, although the true magnitude of influence might be small. Given the potential increase in the magnitude and impact of TSEs due to the implementation of the Common Core State Standards and the increasing use of teachers' VA scores in decision-making about teachers' compensation and tenure, it is important to conduct more research to examine the impact of controlling for TSEs on individual teachers' VA scores in the future.

Third, although most existing studies found evidence of TSEs, prior research has provided little knowledge about the major causes of TSEs. It is important to understand the mechanisms of TSEs to assess whether the observed TSEs represent teachers' true joint contributions to student achievement growth in a certain subject or are just noises due to poor test design. If TSEs are mainly driven by test design factors, test developers need to improve the test design so that student performance on a test truly represents students' knowledge and skills in the subject tested.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago Public High Schools. *Journal of Labor Economics*, *25*, 95–135. http://dx.doi.org/10.1086/508733

Abedi, J. & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219–234. http://dx.doi.org/10.1207/S15324818AME1403_2

Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System. *Educational Researcher, 37*(2), 65–75. http://dx.doi.org/10.3102/0013189X08316420

Amrein-Beardsley, A., Collins, C., Polsky, S. A., & Sloat, E. F. (2013). Value-added model (VAM) research for educational policy: Framing the issue. *Education Policy Analysis Archives, 21*. Retrieved from http://epaa.asu.edu/ojs/article/view/1311

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *Series B, 57*, 289–300. Retrieved from http://www.jstor.org/stable/2346101

Bill & Melinda Gates Foundation (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project*. Retrieved from http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf

Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Retrieved from http://www.ets.org/Media/Research/pdf/PICVAM.pdf

Braun, H., Chudowsky, N., & Koenig, J. A. (2010*). Getting value out of value-added: Report of a workshop*. Washington, DC: National Academies Press. Retrieved from http://www.nap.edu/openbook.php?record_id=12820&page=1

Buddin, R. & Zamarro, G. (2009). *Teacher effectiveness in urban high schools* (WR-693-IES). Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/working_papers/2009/RAND_WR693.pdf

Chang, M., Singh, K., & Filer, K. (2009). Language factors associated with achievement grouping in math classrooms: A cross-sectional and longitudinal study. *School Effectiveness and School Improvement, 20*(1), 27–45. http://dx.doi.org/10.1080/09243450802605704

Chetty, R., Friedman, J., & Rockoff, J. (2013). *Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates* (Working Paper 19423). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w19423

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087–1101. http://dx.doi.org/10.1037/0022-3514.92.6.1087

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, UK: Oliver and Boyd. Retrieved from http://psychclassics.yorku.ca/Fisher/Methods/

Goldhaber, D., Goldschmidt, P., Sylling, P. & Tseng F. (2011). *Teacher value-added at the high school level: Different models, different answers?* Seattle, WA: Center for Education Data & Research. Retrieved from http://www.cedr.us/papers/working/CEDR%20WP%202011-4%20Value-added%20Assessment%20(10-19-2011).pdf

Harris, D. (2011). Value-added measures and the future of educational accountability. *Science, 333*, 826–827. http://dx.doi.org/10.1126/science.1193793

Harris, D., Sass, T., & Semykina, A. (2010). *Value-added models and the measurement of teacher productivity.* Washington, D.C.: The Urban Institute. Retrieved from http://www.urban.org/publications/1001508.html

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relation between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology, 79,* 192–227. http://dx.doi.org/10.1006/jecp.2000.2586

Hill, H. C. (2009). Evaluating value-added models: A validity argument approach. *Journal of Policy Analysis and Management, 28,* 700–709. http://dx.doi.org/10.1002/pam.20463

Jackson, C. K. (2012). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina* (Working Paper 18624). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18624

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

Jacob, B. A. & Rockoff, J. E. (2011). *Organizing schools to improve student achievement: Start times, grade configurations, and teacher assignments.* Washington, D.C.: The Hamilton Project. Retrieved from http://www0.gsb.columbia.edu/faculty/jrockoff/papers/092011_organize_jacob_rockoff_paper.pdf

Jordan, N., Kaplan, D., & Hanich, L. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology, 94*(3), 586–597. http://dx.doi.org/10.1037/0022-0663.94.3.586

Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economics of Education Review*, *28,* 682–692. http://dx.doi.org/10.1016/j.econedurev.2009.02.003

Koretz, D. (2002). Limitations in the use of achievement tests as measures of educators' productivity [Special issue]. *The Journal of Human Resources*, *37*(4), 752–777. http://dx.doi.org/10.2307/3069616

Lockwood, J. R., McCaffrey, D. F., Mariano, L. T., & Setodji, C. (2007). Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics, 32,* 125–150. http://dx.doi.org/10.3102/1076998606298039

McCaffrey, D. F., Han, B., & Lockwood, J. R. (2009a). Turning student test scores into teacher compensation systems. In M. Springer (Ed.), *Performance incentives: Their growing impact on American K-12 education* (pp. 113–147). Washington, D.C.: Brookings Institution Press.

McCaffrey, D. F., Koretz, D., Lockwood, J. R., & Hamilton, L. S. (2004a). *Evaluating value-added models for teacher accountability* (MG-158-EDU). Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG158.pdf

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. S. (2004b). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*, 67–101. http://dx.doi.org/10.3102/10769986029001067

McCaffrey, D. F., Lockwood, J. R., Miraly, K., & Sass, T. R. (2012). A review of Stata routines for fixed effects estimation in normal linear models. *The Stata Journal, 12*(3)*,* 406–432. Retrieved from http://www.stata-journal.com/article.html?article=st0267

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009b). The intertemporal variability of teacher effect estimates. *Education Finance and Policy, 4*(4), 572–606. http://dx.doi.org/10.1162/edfp.2009.4.4.572

National Governors Association Center for Best Practices, & Council of Chief State School Officers (2010). *Common Core State Standards for Mathematics.* Washington, D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers.

Newton, X. A., Darling-Hammond, L., Haertel, E. & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, *18*(23). Retrieved from http://epaa.asu.edu/ojs/article/view/810

O'Reilly, T., & McNamara, D. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal, 44*(1)*,* 161–196. http://dx.doi.org/10.3102/0002831206298171

Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value□added models for estimating school effects. *Education Finance and Policy*, *4*(4), 492–519. http://dx.doi.org/10.1162/edfp.2009.4.4.492

Sanders, W. L., & Horn, S. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, *8*(3), 299–311. http://dx.doi.org/10.1007/BF00973726

Sass, T., Semykina, A., & Harris, D. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review, 38,* 9–23. http://dx.doi.org/10.1016/j.econedurev.2013.10.003

Strauss, V. (2013, April 11). What teachers need and reformers ignore: Time to collaborate. *Washington Post* [online]. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/04/11/what-teachers-need-and-reformers-ignore-time-to-collaborate/

U.S. Department of Education (2014). *Setting the pace: Expanding opportunity for American's students under Race to the Top*. Washington, D.C.: U.S. Department of Education. Retrieved from http://www.whitehouse.gov/sites/default/files/docs/settingthepacerttreport_3-2414_b.pdf

Winters, M. A. (2012). *Transforming tenure: Using value-added modeling to identify ineffective teachers.* New York, NY: Manhattan Institute. Retrieved from http://www.manhattan-institute.org/html/cr_70.htm#.UQrDk47DPww

About the Author

Kun Yuan
RAND Corporation
kyuan@rand.org
Kun Yuan is a behavioral scientist at the RAND Corporation. Her research focuses on measurement of teaching and learning, evaluation of education reforms and programs, STEM education, and technology-based education. She can be reached at kyuan@rand.org

Acknowledgements

This work was supported by the National Center on Performance Incentives, which is funded by the United States Department of Education's Institute of Education Sciences (R305A06034). The author would like to thank the participating district for providing data for analysis and Dr. J. R. Lockwood at the Educational Testing Service for his invaluable input and comments on an earlier version of this paper. The content or opinions expressed do not necessarily reflect the views of the funders of this project.

education policy analysis archives

Volume 23 Number 38 March 30th, 2015 ISSN 1068-2341