# Teacher Spillover Effects Across Four Subjects in Middle Schools

**Kun Yuan**
**RAND Corporation**

## Abstract Body

### Background

Value-added modeling (VAM), one class of statistical models used to estimate 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 last decades (Sanders and Horn, 1994; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). Moreover, teacher value-added scores have been used for making important decisions such as teachers' compensation, bonus, and tenure (Harris, Sass, & Semykina, 2010; Winters, 2012).

Despite the increasing popularity of VAM, many researchers are concerned about the quality of value-added (VA) scores as a measure of teaching and call for more research on VAM. One area of VAM that needs more research is about teacher spillover effects (TSEs). Prior research has 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 two teachers (Jackson & Breugmann, 2009). The second type of TSEs, which is the focus of this study, refers to a teacher's influence on his/her students' achievement on another subject taught by another teacher (Koedel, 2009).

Situations where the second type of TSEs may happen are very common in middle or high schools. Ignoring TSEs when using VAM to assess teaching performance may lead to biased teacher value-added estimates as it ignores the contributions of cross-subject teachers who teach the same group of students with own-subject teachers, which threatens the fairness of any subsequent decision-makings based on these VA measures.

### Purpose of Study and Research Questions

In this study, I studied TSEs across four types of teachers, including math, ELA, science, and social studies, on student achievement in each of the four subjects. When estimating teachers' value-added scores, I used models similar to what is commonly used in the practice, which estimate teachers' annual effects and control for teacher or classroom aggregates of student variables (Bill & Melinda Gates Foundation, 2010). By using models that are commonly used in the practice, I expected to understand the magnitude of TSEs and the consequence of ignoring such effects for estimating own-subject teachers' value-added scores in the common practice. Specifically, I addressed the following two research questions:

1. Do TSEs exist across teachers of math, ELA, science, and social studies on any of the four subjects at the middle school level?
2. If so, how does controlling for TSEs affect the relative stance and variation of own-subject teachers' effects?

### Significance of the Study:

Four studies have examined TSEs in high schools. They focused on TSEs between math and English language arts (ELA) teachers on students' math and ELA achievement. Three studies found TSEs from math teachers to students' ELA achievement and two found TSEs from ELA teachers to students' math achievement.

This is the first study on TSEs at the middle school level. Moreover, I expanded the research on TSEs to more teacher and test subjects than previous studies. I investigated TSEs not only on teacher subjects studied by prior studies (i.e., math and ELA), but also on teacher subjects that are rarely studied by previous research, including science and social studies. I used student achievement on each of these four subjects as the outcome measures and examined TSEs for the other three groups of cross-subject teachers.

**Setting**

Data came from an urban school district in the Southern United States. This district serves a student population of about 70-80 thousand students annually, which is roughly 50 percent African American, 36 percent white, 11 percent Hispanic, and about three percent Asian and other ethnic groups. Ten percent of the students were English language learners (ELL). Over 60 percent of the student population were eligible for free and reduced-price lunch (FRPL). The district's performances on the statewide math and ELA tests were below the state average.

**Participants**

Data used for analysis followed students at grades 6-8 and their math, ELA, science, and social studies teachers from 2006-07 to 2008-09. In total, the analytical sample included 16,886 students at grades 6-8 and 899 teachers. Table 1 shows more details about sample size and demographic characteristics for each grade-year group (see Appendix B).

"(please insert Table 1 here)."

**Statistical Model**

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 nine grade-year groups.

I investigated the existence of TSEs using two types of models. First, I implemented a value-added model with fixed teacher effects (see Model 1) and used F-test to test the significance of teachers' joint contributions to student achievement.

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

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 tested 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. $\lambda$ and $\beta_p$ $(p = 1, 2, 3)$ are vectors of parameters to be estimated for student demographic variables and prior achievement. $D_{ijt}^{school}$, $D_{ijt}^{L(language)}$, $D_{ijt}^{M(math)}$, $D_{ijt}^{R(reading)}$, $D_{ijt}^{Q(science)}$, and $D_{ijt}^{U(social\ studies)}$ are indicator variables for the school and teachers of five subjects for student $i$ in school $j$ in year $t$. $\delta_J$ represents fixed school effect for school $j$. $\theta_L, \theta_M, \theta_R, \theta_Q,$ and $\theta_U$ are fixed teacher effects and parameters of interest to be estimated from the model. $\varepsilon_{ijt}$ is residual error assumed to be mean zero, independently and identically distributed across students.

I applied Model 1 to each of the nine grade-year groups, using students' rank-based z scores on each of the four subjects as the outcome variable. For each grade-year group, I first fit the model with all groups of teachers (referred to as full model). Then I excluded one group of teachers and reran the model (referred to as reduced model). Next I compared the results of the full model with those of the reduced model and used the F-test to examine the significance of effects for the group of teachers excluded from the reduced model. I repeated this process for all types of teachers on all four subjects for each grade-year group. Finally, I pooled the F-test results by grades and across years using Fisher's combined probability test (Fisher, 1925).

I also implemented a value-added model with random teacher effects (see Model 2) to further examine the existence of TSEs.

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

The notations for most model components remain the same as in Model 1. $C_{ij(t-1)}^{T}$ 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. $\gamma$ is a vector of parameters for the teacher-level aggregated achievement and socio-economic status variables. $\zeta_L, \zeta_M, \zeta_R, \zeta_Q,$ and $\zeta_U$ are random teacher effects for five groups of teachers, whose variance are estimated from the model.

I applied Model 2 to each of the nine grade-year groups and collected estimated variances for every type of teacher for each tested subject. I used the bootstrap method to obtain the 95 confidence intervals for the estimated variances of teacher effects. I pooled results by grade and across years and calculated the average variance and its 95 confidence interval for each type of teacher on each tested subject at each grade level. I also calculated the differences in the estimated variance of teacher effects between the own- and cross-subject teachers and their 95 confidence intervals to examine the relative contributions of teachers of different subjects.

To examine whether controlling for TSEs affects the relative stance of own-subject teachers' value-added scores, I implemented Model 2 with and without controlling for TSEs. I obtained estimated value-added scores for own-subject teachers from both runs of Model 2 and pooled them across years by grade level. I examined differences in the quartile rankings of the same teachers' value-added scores before and after controlling for spillover effects. To understand whether controlling for TSEs affects the variation of own-subject teachers' effects, I compared estimated variability across teachers in their true impacts for own-subject teachers' effects from Model 2 before and after controlling for TSEs.

**Findings**

Results from both types of value-added models showed consistent evidences of teacher spillover effects at all three grade levels in middle schools, especially from ELA teachers to students' achievement on the other three subjects.

Results from the fixed teacher effect model showed that although own-subject teachers contributed significantly to students' achievement for most subjects at three grade levels, cross-subject teachers, especially ELA or reading teachers, had significant TSEs on students' achievement on the other three subjects across three grade levels (see Table 2). Significant TSEs found included spillover from reading teachers to students' science and social studies achievement at grade six, from ELA teachers to students' social studies achievement at grade seven, and from ELA teachers to students' achievement in math, science, and social studies at grade eight. In addition, math teachers at grade seven had significant TSEs on students' ELA achievement. Social studies teachers showed significant TSEs on students' math achievement at grade six and science achievement at grade eight.

"(please insert Table 2 here)."

Results from the random teacher effect model showed that although own-subject teachers were the dominant contributors to student achievement for most tested subjects at three grade levels, TSEs existed and were not negligible (see Table 3). Specifically, estimated effect sizes of own-subject teachers were significantly greater than those of TSEs for all tested subjects at grade seven, social studies at grade six, and ELA at grade eight. TSEs that were not statistically different from those of own-subject teachers were found for most subjects at grade six and eight. At grade eight, ELA teachers had TSEs similar to those from own-subject teachers on math, science, and social studies. For ELA and science at grade six, TSEs from all three groups of cross-subject teachers were similar to those from own-subject teachers. For math at grade six, science teachers' spillover effects were similar to those of math teachers.

" please insert Table 3 here"

Results also showed that controlling for TSEs affected the relative stance of value-added scores for up to one-third of own-subject teachers and reduced the variation and increased the estimation errors of own-subject teachers' effects. Specifically, controlling for TSEs resulted in changes of the quartile rankings of value-added scores for 10-35 percent of own-subject teachers (see Table 4). In addition, controlling for TSEs reduced the variation of own-subject teachers' effects by 0.01-0.05 standard deviations, and the magnitude of change varied by grade.

" please insert Table 4 here"

**Conclusions**

Consistent with most prior studies on teacher spillover effects, this study found evidence of TSEs. Although these studies differed in many ways such as the specific magnitude of TSEs reported and the relative magnitude of effect sizes between own- and cross-subject teachers, common findings from these studies suggest that TSEs exist at the secondary school level.

Although results from this study and those from prior research showed that the impact of controlling for TSEs on the variation of own-subject teachers' effects is relatively small (Aaronson, Barrow, & Sander, 2007; Buddin & Zamarro, 2009), findings from this study that up to one-third of own-subject teachers were affected in the quartile rankings of value-added scores by controlling of TSEs indicate that cautions are warranted when quartile rankings of teacher value-added scores are used for high-stakes decision-making such as performance bonuses.

Results from this study suggest it is necessary to at least examine the existence and magnitude of TSEs when conducting teacher value-added analysis and using such results for decision-making. However, lack of teacher-student linkage data for subjects not tested by the state accountability exams might be a constraint for such analysis. Thus, efforts need to be made early to collect data required for such analysis. Moreover, results indicate group-based incentives may be reasonable for some groups of teachers at certain grade levels such as for math and ELA teachers at grade eight in order to improve students' math achievement.

This study contains two limitations. First, I had to apply a set of restrictions on students and teachers to identify the analytical sample, which might have led to more homogeneous samples than their respective populations and thus underestimated the variation of teacher effects in middle schools. However, these restrictions were necessary to ensure a stable student and teacher sample for analysis. Second, findings of this study about TSEs at the middle school level came from only one school district. Other factors such as the curriculum used in this particular district may also have certain effects on the TSEs observed in this study. More research on TSEs using data from other school districts will help us better understand the prevalence and magnitude of TSEs at the middle school level.

## Appendices

### Appendix A. 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.

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

Buddin, R. & Zamarro, G. (2009). *Teacher effectiveness in urban high schools (WR-693-IES)*. Santa Monica, CA: RAND Corporation.

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

Harris, D., Sass, T., & Semykina, A. (2010). Value-added models and the measurement of teacher productivity. Washington, D.C.: The Urban Institute.

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.

Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economics of Education Review*, *28,* 682-692.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, Laura S. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*, 67-101.

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.

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

**Appendix B. Tables and Figures**

Table 1. *Sample Size and Demographic Characteristics of Each Grade-Year Group*

| Grade | | School Year | | |
|---|---|---|---|---|
| | | 2006-07 | 2007-08 | 2008-09 |
| 6 | Number of students | 3060 | 2888 | 2833 |
| | Number of math teachers | 95 | 96 | 93 |
| | Number of language teachers | 94 | 106 | 108 |
| | Number of science teachers | 85 | 87 | 84 |
| | Number of social studies teachers | 131 | 127 | 127 |
| | Number of reading teachers | 153 | 134 | 121 |
| | Number of schools | 36 | 35 | 36 |
| | % of ELL students | 1% | 1% | 3% |
| | % of white students | 38% | 37% | 38% |
| | % of African American students | 49% | 47% | 44% |
| | % of Hispanic students | 10% | 12% | 14% |
| | % of students eligible for FRPL | 60% | 63% | 62% |
| | % of special education students | 5% | 6% | 5% |
| 7 | Number of students | 2822 | 2840 | 3080 |
| | Number of math teachers | 72 | 71 | 65 |
| | Number of language teachers | 67 | 64 | 82 |
| | Number of science teachers | 58 | 52 | 49 |
| | Number of social studies teachers | 65 | 63 | 63 |
| | Number of schools | 34 | 34 | 35 |
| | % of ELL students | 2% | 2% | 5% |
| | % of white students | 35% | 35% | 35% |
| | % of African American students | 50% | 48% | 46% |
| | % of Hispanic students | 11% | 13% | 16% |
| | % of students eligible for FRPL | 58% | 57% | 63% |
| | % of special education students | 4% | 5% | 6% |
| 8 | Number of students | 2842 | 2860 | 2966 |
| | Number of math teachers | 69 | 65 | 58 |
| | Number of language teachers | 64 | 59 | 78 |
| | Number of science teachers | 50 | 51 | 46 |
| | Number of social studies teachers | 69 | 62 | 55 |
| | Number of schools | 35 | 35 | 32 |
| | % of ELL students | 2% | 2% | 6% |
| | % of white students | 38% | 36% | 34% |
| | % of African American students | 51% | 48% | 47% |
| | % of Hispanic students | 8% | 12% | 15% |
| | % of students eligible for FRPL | 57% | 57% | 61% |
| | % of special education students | 6% | 6% | 6% |

Note: ELL = English language learner; FRPL = free and reduced-price lunch. The percentage of male students ranged from 48 to 50 percent for all grade-year groups.

Table 2. *Significance of Chi-Square Test for Pooled F-Test*

| Grade | Teacher Subject | Test Subject | | | |
|---|---|---|---|---|---|
| | | Math | ELA | Science | Social Studies |
| 6 | Math | | | | |
| 6 | ELA | | | | |
| 6 | Reading | | | | |
| 6 | Science | | | | |
| 6 | Social Studies | | | | |
| 7 | Math | | | | |
| 7 | ELA | | | | |
| 7 | Science | | | | |
| 7 | Social Studies | | | | |
| 8 | Math | | | | |
| 8 | ELA | | | | |
| 8 | Science | | | | |
| 8 | Social Studies | | | | |

Note: Grey cells indicate significant teacher effects for a particular type of teacher on a tested subject in a certain year after FDR adjustment. Dashed lines indicate a grade block. Grey cells on the main diagonal within each grade block indicate significant own-subject teacher effects, and grey cells off the main diagonal within each grade block indicated significant TSEs. For grade six, ELA and reading teachers' effects on students' ELA achievement are considered as own-subject teachers' effects.

Table 3. *Standard Deviations and 95 Confidence Intervals of Teacher Effects for Each Type of Teacher on Each Test Subject*

| Grade | Teacher Subject | Test Subject: Math | ELA | Science | Social Studies |
|---|---|---|---|---|---|
| 6 | Math | **0.14**$^{\#}$ | 0.06$^{n\#\#}$ | 0.1$^{n}$ | 0.06 |
| | | **[.11, .18]** | [.02, .08] | [.06, .13] | (0$^{\#\#\#}$, .09] |
| 6 | ELA | 0.04 | **0.07** | 0.04$^{n}$ | 0.05 |
| | | (0, .08] | **[.02, .1]** | (0, .08] | (0, .08] |
| 6 | Reading | 0.06 | **0.07** | 0.07$^{n}$ | 0.05 |
| | | [.04, .09] | **[.03, .09]** | [.03, .09] | (0, .07] |
| 6 | Science | 0.09$^{n}$ | 0.03$^{n}$ | **0.09** | 0.08 |
| | | [.06, .13] | (0, .06] | **[.05, .13]** | [.03, .11] |
| 6 | Social Studies | 0.06 | 0.04$^{n}$ | 0.08$^{n}$ | **0.13** |
| | | [.03, .08] | [.01, .07] | [.06, .11] | **[.1, .15]** |
| 7 | Math | **0.16** | 0.06 | 0.07 | 0.08 |
| | | **[.13, .2]** | [.02, .08] | [.03, .09] | [.05, .11] |
| 7 | ELA | 0.05 | **0.1** | 0.05 | 0.08 |
| | | [.01, .07] | **[.07, .13]** | [.02, .08] | [.04, .1] |
| 7 | Science | 0.03 | 0.04 | **0.19** | 0.05 |
| | | (0, .06] | (0, .07] | **[.14, .23]** | (0, .08] |
| 7 | Social Studies | 0.04 | 0.04 | 0.04 | **0.17** |
| | | [.01, .07] | (0, .07] | (0, .07] | **[.13, .2]** |
| 8 | Math | **0.16** | 0.06 | 0.04 | 0.06 |
| | | **[.13, .19]** | [.02, .08] | (0, .07] | [.02, .08] |
| 8 | ELA | 0.12$^{n}$ | **0.16** | 0.12$^{n}$ | 0.17$^{n}$ |
| | | [.09, .16] | **[.12, .2]** | [.08, .15] | [.13, .21] |
| 8 | Science | 0.03 | 0.03 | **0.13** | 0.05 |
| | | (0, .06] | (0, .06] | **[.08, .17]** | (0, .08] |
| 8 | Social Studies | 0.05 | 0.04 | 0.07 | **0.12** |
| | | [.02, .08] | (0, .06] | [.04, .1] | **[.09, .15]** |

Notes:

$^{\#}$ Bolded cells indicate effect sizes for own-subject teachers. Normal text cells represent spillover effect sizes.

$^{\#\#}$ Superscript n indicates the difference between own- and cross-subject teachers was non-significant. Cells without superscript *n* indicate significant differences between own- and cross-subject teachers' effects.

$^{\#\#\#}$ 0 indicates the estimated standard deviation of a certain types of teachers' effects was smaller than 0.01 after I rounded it up to two digits after the decimal.

Table 4

*Statistics for the Percentage of Teachers Who Changed Quartile Ranking of Their Value-Added Estimates after Controlling for Teacher Spillover Effects*

| Grade | Teacher | N | Percent of Teachers Who Changed Quartile Ranking | Mode of Changed Quartile Ranking | Max of Changed Quartile Ranking |
|---|---|---|---|---|---|
| 6 | Math | 284 | 33% | 1 | 1 |
| 6 | ELA | 308 | 10% | 1 | 1 |
| 6 | Reading | 408 | 35% | 1 | 2 |
| 6 | Science | 256 | 24% | 1 | 1 |
| 6 | Social Studies | 385 | 15% | 1 | 3 |
| 7 | Math | 208 | 11% | 1 | 1 |
| 7 | ELA | 213 | 18% | 1 | 3 |
| 7 | Science | 159 | 14% | 1 | 3 |
| 7 | Social Studies | 191 | 12% | 1 | 3 |
| 8 | Math | 192 | 18% | 1 | 2 |
| 8 | ELA | 201 | 15% | 1 | 3 |
| 8 | Science | 147 | 23% | 1 | 3 |
| 8 | Social Studies | 186 | 27% | 1 | 3 |

Note: N represents the total number of own-subject teachers for a particular subject across three years.