Measuring Teacher Effectiveness in Memphis

April 30, 2008

Eric Isenberg

Submitted to:
New Leaders for New Schools
30 West 26th Street
New York, NY 10010

Project Officer: Dianne Houghton

Submitted by:
Mathematica Policy Research, Inc.
600 Maryland Ave., SW, Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Duncan Chaplin
THE MATHEMATICA POLICY RESEARCH
TEACHER-LEVEL VALUE-ADDED MODEL

INTRODUCTION

Mathematica Policy Research, Inc. (MPR) has adapted its school-level value-added model to teacher-level data to help New Leaders for New Schools identify high-performing teachers in award-winning schools within Memphis City Schools (MCS). An earlier memo, “Measuring School Effectiveness in Memphis” (Booker and Isenberg 2008), illustrates how MPR identifies high-performing schools and includes a nontechnical description of the MPR value-added model for schools (VAM-S). This memo describes the technical details of the value-added model for teachers (VAM-T). Readers are advised to consult the earlier memo before reading this piece.

DATA

Data are provided by MCS and originate from three sources: student-level test score and demographic data used in the VAM-S and two files of enrollment data that pair students with teachers. Booker and Isenberg (2008) describe the construction of the test scores. Only students included in the estimation sample for the VAM-S are included in the VAM-T.

For each student-teacher pair, the enrollment data give dates that the student entered and exited the teacher’s classroom. One of these files—“homeroom data”—links elementary school students to teachers. The other file—“subject data”—links middle and high school students to teachers by math and reading courses, and indicates the course names. Test score data for high school students come from Gateway exams, which cover Algebra I or English II; the high school sample is limited to students who took the Gateway exam for the first time in spring 2007, were enrolled in one or both of these courses in 2006–2007, and did not previously fail this course.

Similar to the VAM-S model, dosage variables are created for each student corresponding to the proportion of the year the student spent with each teacher. In the homeroom and subject data sets, some students are assigned simultaneously to more than
one teacher. MPR uses a procedure described in the appendix to assign approximate teacher dosage values to these students. Because a teacher is unlikely to have an appreciable educational impact on a student who spends a very short time in that teacher’s classroom, the dosage variable is set to zero for students who spent fewer than two weeks in a teacher’s classroom. The dosage variable is set to one for students who spent all but two weeks or fewer in a given classroom. Once teacher dosage values have been assigned for each student, test score and demographic data are merged with the dosage data.

The homeroom and subject data sets were merged and subsequently split into elementary school, middle school, and high school data sets. (The school levels are defined in Booker and Isenberg [2008].) Students were dropped if they did not attend any schools that were included in the VAM-S model. The high school data set includes only subject data. The elementary and middle school data sets include both homeroom and subject data because some students split time between both types of schools. For example, a sixth grader who attended an elementary school with grades K–6 for part of the year and a middle school with grades 6–8 for the remainder will appear in both data sets. Dosage data are combined for students who appear in both data sets. Only one teacher overlaps the homeroom and subject data; for this teacher, student dosage data are combined into a single variable.

Teachers with very few students are not ranked. These are primarily teachers who did not teach at schools that are in the data but who taught some mobile students for part of the school year. Each teacher dosage variable is summed across all students. If a teacher has five or fewer dosage-weighted students, then the dosage variable for this teacher is set to zero. The model implicitly estimates the effect of all omitted teachers (and any teachers employed outside of MCS who taught mobile students for part of the year) as a single “other teacher.”

The number of students, student-test-years, and teachers in each data set are summarized in Table 1. Student-test years are calculated by summing the teacher dosage variables. For example, a math teacher who taught 10 students for the whole year and one student for half of the year would have 10.5 student-test-years. If she also taught reading to the same students, she would have 21.0 student-test-years.

<table>
<thead>
<tr>
<th></th>
<th>Students</th>
<th>Student-Test-Years</th>
<th>Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary</td>
<td>25,115</td>
<td>97,799</td>
<td>1,285</td>
</tr>
<tr>
<td>Middle</td>
<td>19,755</td>
<td>36,490</td>
<td>549</td>
</tr>
<tr>
<td>High</td>
<td>9,733</td>
<td>11,482</td>
<td>229</td>
</tr>
</tbody>
</table>
**MODEL**

The specification of the VAM-T is similar to the VAM-S described in Booker and Isenberg (2008):

\[ Y_{i,j,t} = \beta_1 \hat{Y}_{i,j,t-1} + \beta_2 \cdot X_{i,t} + \beta_3 \cdot T_{i,t} + e_{i,j,t} \]

where \( Y_{i,j,t} \) is the 2006–2007 test score for student \( i \) in subject \( j \), \( \hat{Y}_{i,j,t-1} \) is the predicted prior test score for student \( i \) in subject \( j \), \( X_{i,t} \) is a vector of control variables for individual student characteristics (described below), \( T_{i,t} \) is a vector of teacher dosage variables, and \( e_{i,j,t} \) is the error term. As in the VAM-S model, to correct for measurement error in the pretest, the model uses two-stage least squares with the average of the student’s prior test scores in other subjects as an instrumental variable for the student’s prior test score. The value of \( \hat{Y}_{i,j,t-1} \) is assumed to capture all previous inputs into student achievement. The vector \( T_{i,t} \) includes one variable for each teacher in the model. Each variable equals the percentage of the year student \( i \) spent with that teacher. The value of any element of \( T_{i,t} \) is zero if student \( i \) was not enrolled in that teacher’s classroom. The teacher performance measures are the coefficients on \( T_{i,t} \), which are the elements of the vector \( \beta \).

The elementary school model, based primarily on the homeroom data, uses four observations for most students, one for each subject (math, reading, science, and social studies.) Thus, each student can contribute to a teacher performance measure four times, once for each test. The middle school data are derived primarily from the subject data, which only include math and reading.1 Each student contributes one observation for single-subject math or reading teachers, up to two observations for teachers who teach all subjects when data are derived from the subject data, and up to four observations when the homeroom data are used. In the high school model, all data are derived from the subject data, so there are at most two observations per student. The standard errors of the teacher performance measures are adjusted for the clustering of observations by student.

The VAM-T model and the VAM-S model include almost identical control variables for exogenous student characteristics. In addition to the student’s lagged test score, the teacher dosage variables, and a constant term, the VAM-T regressions include the following variables:

- Gender indicator
- Race/ethnicity indicators (white, African American, Hispanic, Asian, Native American)
- Free or reduced-price lunch indicator

---

1 The VAM-S used science and social studies tests for middle school students. Those data could not be used in the VAM-T because the subject data do not match students and teachers for those subjects.
• Limited English proficiency indicator
• Special education status indicator
• An indicator for skipping a grade
• An indicator for being held back a grade
• Subject-by-grade indicators
• For the high school model, an indicator for being enrolled in a review course for the Gateway exam

The last variable is not included in the VAM-S model. Two variables are included in the VAM-S model but are excluded from the VAM-T model: an indicator for switching schools within the 2006–2007 school year and the combined percentage of the school year that the student spent outside schools in the estimation sample. These variables are excluded from the VAM-T because they are collinear with the teacher dosage variables.

**HOW TO INTERPRET THE RESULTS**

The VAM-T produces an estimated teacher effect across all subjects for which there are data on the teacher’s students. Within each school level (elementary, middle, or high school), the average teacher score is set to zero so that teachers with a positive score are higher than average and teachers with a negative score are lower than average. Based on these scores, MPR also reports the percentile to which each teacher would raise or lower the test score of the median student. For example, consider a teacher with a percentile value of 65. A student who would have scored at the 50th percentile of the student distribution with an average teacher would be estimated to score at the 65th percentile with this teacher.

The standard error is used to calculate a 90 percent confidence interval for each teacher. This confidence interval is used to report a high and low rank for each teacher, which corresponds to the ranks the teacher would have received if his or her performance measure was at the high or low end of the 90 percent confidence interval.

Although teachers are ranked against all teachers within their school level, the most meaningful comparisons can be made between teachers in the same school. Because the rankings measure the net influence of the classroom plus school characteristics (like the effect of the principal or school culture on student achievement), it is not possible to disentangle the effect of the teacher from that of the school. Thus, one cannot infer that if teachers transferred across schools, they would be as effective at raising student test scores in their new school as they were in their old one. Aaronson, Barrow, and Sander (2007) arrive at a similar conclusion using data on Chicago high school teachers.
When comparing teachers within a school, one should take note of the confidence intervals around the estimates of teacher performance.\textsuperscript{2} Figures 1 through 3 show the teacher scores (marked with a diamond) along with upper and lower bounds of a 90 percent confidence interval for an award-winning elementary school, middle school, and high school. As Figure 1 shows, the lower bound of the top-ranked teacher is above the estimated teacher score of the second-best teacher. The second-best teacher’s lower bound is higher than the fourth-best teacher’s score, and the third-best teacher’s lower bound is above the score of the fifth best teacher. The bottom 10 teachers, however, are difficult to distinguish from each other because of the large overlaps in their confidence intervals. In Figure 2, the top-ranked middle school teacher is clearly ranked better than the second-best teacher, but the second to sixth best teachers are harder to distinguish from each other. The top-ranked high school teacher in Figure 3 is not easily distinguished from the next six teachers.

\textsuperscript{2} A related concern is that measures of teacher effectiveness may vary from year to year. Lockwood et al. (2008) have found moderate correlations—ranging from 0.3 to 0.5—between value-added measures of teacher performance across years.
Figure 1: Teacher Scores at an Award-Winning Elementary School
Figure 2: Teacher Scores at an Award-Winning Middle School
Figure 3: Teacher Scores at an Award-Winning High School
REFERENCES


APPENDIX A

MCS provided MPR with classroom entry and exit dates for each student. According to the homeroom data, some students are enrolled simultaneously in more than one teacher’s classroom within the same school. If a student has exactly two teachers, the exit date for the first classroom is assigned as the day before the entry date for the second classroom. This is not done in the subject data. Apart from this correction, the same procedures are followed for the homeroom and subject data. For the subject data, math and reading teachers for each student are treated separately.

Total days enrolled for each student with each teacher are approximated based on the classroom entry and exit dates. Total days in the school year are based on an estimate of total days a student could be enrolled in MCS in 2006–2007. An initial estimate of teacher dosage was derived by dividing days enrolled with a particular teacher by total days in the school year. With this procedure, some students, especially those who changed schools, had an initial total dosage (the sum of the student-teacher dosages) in excess of 100 percent. For these students, updated teacher dosages were assigned by dividing the days enrolled with each teacher by the total days enrolled for that student in all teachers’ classrooms. As a result, no student has a total dosage of more than 100 percent.

Finally, the sum of a student’s dosage values for each school attended are adjusted to align with the dosage data from the VAM-S model. The school-level data used in the VAM-S report actual days enrolled for each student in each school; therefore, these data are assumed to be more accurate than the teacher-level data, which approximate days enrolled based on entry and exit dates. To make this final adjustment, the dosage for each student/teacher pair is multiplied by the VAM-S dosage estimate for the student at that school and then divided by the sum of the teacher dosages for all of the student’s teachers at that school.

Students are flagged when their dosage measures are approximated due to reported simultaneous enrollments. This accounts for between 19 and 25 percent of the sample of each subject in each data set except for middle school reading courses, in which 60 percent of the students are reported to have simultaneous enrollments. The rate of simultaneous
enrollment is higher among middle school students because many of them are enrolled in a “reading” course and a “language arts” course concurrently. Model results excluding these students are qualitatively similar to results that include them.