ACT Research & Policy

Issue Brief

April 2015

The Effects of Model Choice on Estimates of Teacher Effectiveness

Edgar I. Sanchez and Yan Zhou

Introduction

Effective teachers and schools play important roles in improving student academic achievement.[1] Accordingly, US Department of Education grants, including Race to the Top, and state waivers under the Elementary and Secondary Education Act support the evaluation of teacher effectiveness.[2] Efficient teacher and school evaluation methods that recognize exemplary teachers and identify teachers or schools in need of additional support are vital.

In this issue brief we explore the practical effects of ignoring or including school effects when estimating teacher effectiveness. This is operationalized by using a two- or three-level hierarchical value-added model (VAM). We also offer points for consideration when deciding which mathematical model to use. This will allow us to see if the conclusions reached under two- and three-level VAMs are consistent.

Value-added models attempt to isolate the impact of teachers and schools on achievement by adjusting for the impact of confounding factors such as prior achievement and demographics on achievement scores.[3] Education researchers and practitioners have advocated using VAMs to evaluate teacher and school effects on student achievement in order to identify higher- and lower-performing teachers.[4]

Hierarchical models are one way in which VAMs can be operationalized. Hierarchical models are advantageous when the response data is nested, as education data is. In education, students in a classroom are taught by the same teacher, and teachers can work at the same school. When we analyze data focusing on students nested within (taught by) teachers, we are using a two-level model. When we analyze data with students nested in teachers and teachers nested within (employed by) schools, we are using a three-level model. This type of model is shown in figure 1.

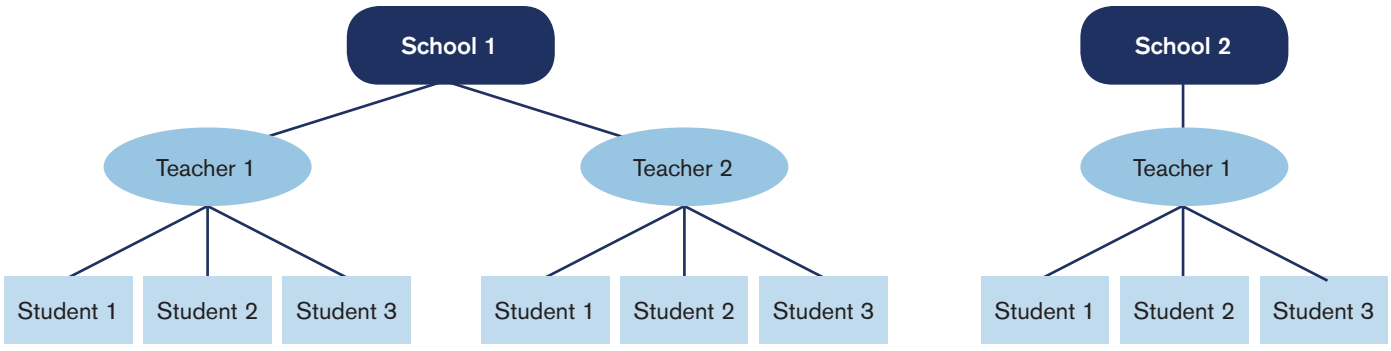

Figure 1. Example of Three-level Hierarchical VAM

research.policy@act.org for more information.

A two-level student and teacher hierarchical VAM attempts to account for the common environment of students who have the same teacher (e.g., students being exposed to the same instruction from the same teacher). A three-level student, teacher, and school hierarchical model also accounts for teachers being nested within school and for the similarity of teachers within schools (e.g., having similar resources and principal or leadership support). This is accomplished by partitioning the variance between teacher and school levels, thereby attempting to isolate the influence of each. This partitioning has an effect on the statistical tests that are conducted at each level and can therefore influence findings of statistical significance. In practical applications where these mathematical models are used to make decisions about teachers and schools, the partitioning of variance could affect conclusions about teachers' performance relative to their peers.

The two- and three-level VAMs provide alternative ways to conceptualize the instructional process. The two-level VAM ignores the impact of school-level effects such as differences in available resources. This model assumes such effects will not affect student achievement and do not need to be considered when evaluating teacher effects. The three-level model explicitly includes school effects. This model therefore could be used when there is a theoretical basis for concluding school effects play an important role in student achievement.

The question we wish to have answered can help determine which model to use. For example, would we rather know how a teacher compares to all teachers regardless of school or how a teacher compares to teachers at the same school? In a two-level model, teachers are compared to all other teachers regardless of their school assignment. In this way we can compare a science teacher from school 1 to a science teacher from school 2. A three-level model, however, compares teachers to other teachers at the same school.

To explore how the two VAMs affect evaluation of teacher effects, data from seven ACT QualityCore® courses between 2010 and 2013 were included and analyzed separately by course. ACT QualityCore is a research-based system of educator's resources, formative items, and end-of-course assessments designed to help schools better prepare all students for college and career. It accomplishes this goal by ensuring that high school core courses follow a rigorous curriculum focused on essential postsecondary skills and by enabling educators to monitor progress toward and attainment of those skills.

The ACT QualityCore courses; the total number of students, teachers, and schools; the minimum, median, and maximum number of students per teacher; and the number of teachers per school in this study are shown in table 1. Both the two- and three-level models used the following predictors: prior achievement test score in the subject area, gender, limited English proficiency status, special education status, poverty status, minority status, the length of time between the prior achievement test and the end-of-course exam, grade level, classroom average for the prior achievement test score, percentage of the classroom with limited English proficiency, percentage of the classroom with a poverty classification, and percentage of the classroom identified as minority.[5]

Table 1. Course-Specific Sample Sizes

				Students per teacher			Teachers per school		
Course (pretest)	**Schools**	**Teachers**	**Students**	**Min***	**Median**	**Max†**	**Min**	**Median**	**Max**
English 10 (English 9)	80	253	16,780	11	52	298	1	2	20
English 11 (English 10)	56	157	12,181	10	60	284	1	2	10
English 12 (English 11)	13	23	823	11	36	72	1	1	4
Algebra 2 (Geometry)	67	184	10,130	11	45	241	1	2	14
Geometry (Algebra 1)	75	201	11,843	11	40	363	1	2	11
Precalculus (Algebra 2)	56	103	5,968	11	48	208	1	1	6
Chemistry (Biology)	56	121	8,724	11	55	264	1	2	6

* A minimum number of ten students per teacher was required for inclusion.

† Data were collected across years and classes for each teacher.

Differences in Teacher Effect Estimates from Two- and Three-Level VAMs

Across the seven courses, the correlations of teacher effects between the two models were all strongly positive (see table 2).[6] These correlations ranged from a low of 0.81 for Algebra 2 to a high of 0.90 for both English 11 and Precalculus. This finding indicates there is strong agreement in the direction and magnitude of teacher effect estimates regardless of the inclusion of school-level effects.

Differences in Statistical Significance of Teacher Effect Estimates from Two- and Three-Level VAMs

Fewer teachers were found to have statistically different teacher effects from the average teacher under the three-level VAMs than under the two-level VAMs. As shown in table 3, approximately 20% to 48% of teachers were found to have statistically significant teacher effects under the two-level VAMs.[7] For the three-level VAMs, the percentages of teachers with statistically significant teacher effects range from approximately 0% to 14%. Recall from table 2 that the teacher effect estimates were similar for both the two- and three-level models. We see fewer teachers having a significantly higher or lower effect than their peers because of the comparison of the smaller teacher effect—due to schools' average effect of zero—to the smaller school variance. This additional nesting makes it more difficult to identify teachers as having a significantly higher or lower effect than their peers.

Determination of significance of teacher effects, regardless of directionality or magnitude, was fairly consistent across the two- and three-level VAMs for all courses examined. All courses had an agreement rate above 60%. Table 4 and table 2 together suggest that for many teachers the teacher

Table 2. Correlation of Teacher Effects under the Two- and Three-Level VAMs

Course (pretest)	Correlation of teacher effects
English 10 (English 9)	.82
English 11 (English 10)	.90
English 12 (English 11)	.84
Geometry (Algebra 1)	.87
Algebra 2 (Geometry)	.81
Precalculus (Algebra 2)	.90
Chemistry (Biology)	.83

Table 3. Percentage of Significant Teacher/School Effects under the Two- and Three-Level VAMs

	Two-level VAM	Three-level VAM	
Course (pretest)	Significant teachers (%)	Significant teachers (%)	Significant schools (%)
English 10 (English 9)	28	10	15
English 11 (English 10)	39	14	7
English 12 (English 11)	20	0	7
Geometry (Algebra 1)	31	8	9
Algebra 2 (Geometry)	42	12	16
Precalculus (Algebra 2)	43	5	4
Chemistry (Biology)	48	13	16

Table 4. Percentage of Agreement for Teacher Effect Significance under the Two- and Three-level VAMs

Course (pretest)	Significant under two-level VAM only (%)	Significant under three-level VAM only (%)	Significant under both VAMs (%)	Not significant under both VAMs (%)	Overall agreement (%)*
English 10 (English 9)	21	1	8	70	78
English 11 (English 10)	27	2	11	60	71
English 12 (English 11)	13	0	0	87	87
Geometry (Algebra 1)	22	0	7	71	78
Algebra 2 (Geometry)	28	1	10	61	71
Precalculus (Algebra 2)	38	0	5	57	62
Chemistry (Biology)	36	2	11	50	61

* Overall agreement rate included teacher effect estimates that were either significant or not significant in both the two- and three-level VAMs.

effect estimates and their significance will be similar regardless of the inclusion of school-level effects. However, the inclusion of school-level effects did have an impact on the significance of estimates of many teachers. We can see that between 13% and 38% of teachers that were significant under the two-level VAM were not significant under the three-level VAM. The ability to account for school-level effects has a clear impact on identifying teachers with significantly higher or lower effects.[8]

How Variance Estimates from Two- and Three-Level VAMs Differ

Student-level residual variances from the two- and three-level VAMs should be virtually identical because of how variance is partitioned by the models. In the three-level VAMs, the teacher-level variance in the two-level VAMs is broken up into the teacher- and school-level variance. The school-level variance tends to be greater than the teacher-level variance for most courses. This suggests that school-level effects such as available resources might be factors worth considering when estimating student achievement scores. This is a finding that would be overlooked without the inclusion of school-level effects in a three-level VAM. While student-level characteristics account for a large portion of the variance in student achievement (see table 5), it is important to target the teacher and school levels for interventions. These may include professional development for teachers and principals.

Summary of Analysis

In this brief we compared value-added models for estimating teacher effects that include and exclude school-level effects. Our analysis showed several inherent consequences that should be considered when deciding how to model student achievement using hierarchical VAMs:

1. Teacher effect estimates from VAMs which did and did not account for school-level effects were highly correlated.
2. When the school-level effects were accounted for, the effects for fewer teachers were found to be significantly different from those of the average teacher.
3. There was a high agreement rate on the identification of teachers who were not significantly different from the average teacher regardless of whether school-level effects were taken into account.
4. School-level variance tended to be larger than teacher-level variance. That said, student-level characteristics accounted for considerably more of the variance in student achievement scores.

Implications for Practice

Based upon the conclusions presented in this brief, we can infer several implications for practice:

1. When selecting a VAM, it is important to take into account the types of comparisons one wishes to make. If there is a desire to compare teachers across a state, it would be appropriate to implement a two-level VAM. This type of model would also be appropriate for comparing teachers in the same subject within a single school district.
2. While both a two- and three-level VAM may result in high percentages of teachers identified as having an average impact on student achievement, a two-level model, which compares all teachers, will identify a greater number of higher- and lower-performing teachers.
3. If there is a strong theoretical basis for believing that school differences have an important influence on student achievement, it would be appropriate to use a three-level VAM. This type of model will estimate teacher impacts on student achievement while accounting for important school-level factors such as resource allocation. ■

Table 5. Variance Estimates under the Two- and Three-Level VAMs

	Two-level VAM		Three-level VAM		
Course (pretest)	**Student residual variance**	**Teacher residual variance**	**Student residual variance**	**Teacher residual variance**	**School residual variance**
English 10 (English 9)	15.3	1.9	15.3	0.9	1.1
English 11 (English 10)	20.5	3.0	20.5	1.8	1.6
English 12 (English 11)	17.9	4.2	17.9	1.0	4.0
Geometry (Algebra 1)	9.8	1.3	9.8	0.6	0.8
Algebra 2 (Geometry)	9.3	2.3	9.3	1.0	2.0
Precalculus (Algebra 2)	11.7	3.4	11.7	2.0	1.5
Chemistry (Biology)	14.7	4.9	14.7	2.4	2.9

Notes

1 William L. Sanders, "Value-Added Assessment from Student Achievement Data: Opportunities and Hurdles," *Journal of Personnel Evaluation in Education* 14, no. 4 (2000): 329–339; Robert Gordon, Thomas J. Kane, and Douglas O. Staiger, "Identifying Effective Teachers Using Performance on the Job," in *Path to Prosperity: Hamilton Project Ideas on Income Security, Education, and Taxes*, ed. Jason Furman and Jason Bordoff (Washington, DC: Brookings Institution Press, 2008); Robert H. Meyer, "Value-Added Indicators of School Performance: A Primer," *Economics of Education Review* 16, no. 3 (1997): 283–301; Jennifer King Rice, *Teacher Quality: Understanding the Effectiveness of Teacher Attributes* (Washington, DC: Economic Policy Institute, 2003); Susan J. Rosenholtz, "Effective Schools: Interpreting the Evidence," *American Journal of Education* 93, no. 3 (1985): 352–388; William L. Sanders and June C. Rivers, *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement* (Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center, 1996), http://mccluelearning.com/wp-content/uploads/2011/09/Cumulative-and-Residual-Effects-of-Teachers.pdf.

2 Glenda L. Partee, *Using Multiple Evaluation Measures to Improve Teacher Effectiveness: State Strategies from Round 2 of No Child Left Behind Act Waivers* (Washington, DC: Center for American Progress, 2012).

3 Daniel F. McCaffrey, J.R. Lockwood, Daniel M. Koretz, and Laura S. Hamilton, *Evaluating Value-Added Models for Teacher Accountability* (Santa Monica, CA: The RAND Corporation, 2003), http://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG158.pdf.

4 See John P. Papay, "Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates Across Outcome Measures," *American Educational Research Journal* 48, no. 1 (2011): 163–193; and Stephen W. Raudenbush, "What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice?," *Journal of Educational and Behavioral Statistics* 29, no. 1 (2004): 121–129.

5 Student prior-achievement test, or pretest, scores were based on the ACT QualityCore End-of-Course (EOC) Assessment for the course preceding the course of interest. For example, for English 10 the pretest would be the EOC exam score for English 9. Data for this analysis came from high school information collected as part of ACT QualityCore program. Due to sample size restrictions, not all ACT QualityCore subject courses were used in this analysis.

6 Recall from table 2 that many schools only have a single teacher for each subject. The three-level model, which compares teachers within schools, produces an estimate near zero for an individual teacher. In a two-level VAM where that same teacher is compared to all other teachers, most teachers will likely receive an estimate near zero, the average teacher effect. That may be contributing to the high correlation.

7 Statistical significance was evaluated at the 0.05 level.

8 The school effect may be confounded by the teacher effect at schools with small numbers of teachers in each subject.