A Comparison of Alternative Models for Estimating School Performance in Mathematics and Reading/Language Arts in Four State Accountability Systems:

Pennsylvania Results

NCAASE Technical Report

March, 2018

Joseph F. T. Nese[a]
Joseph J. Stevens[a]
Ann C. Schulte[b]
Gerald Tindal[a]
Nedim Yel[c]
Daniel Anderson[a]
Tyler Matta[a]
Stephen N. Elliott[b]

[a] University of Oregon
[b] Arizona State University
[c] Indiana University

Address all correspondence to Joseph F. T. Nese, University of Oregon, Behavioral Research and Teaching, College of Education, 5262 University of Oregon, Eugene, OR 97403-5262; 541-346-2174, jnese@uoregon.edu

This research was funded in part by a Cooperative Service Agreement from the Institute of Education Sciences (IES) establishing the National Center on Assessment and Accountability for Special Education – NCAASE (PR/Award Number R324C110004); the findings and conclusions expressed do not necessarily represent the views or opinions of the U.S. Department of Education.

Table of Contents

Background and Introduction..... 1
General Method Description..... 1
Sample..... 1
Instruments..... 1
School Performance Models..... 2
Percent proficient..... 2
Average gain score..... 2
Transition matrix..... 3
Student growth percentiles..... 3
Value-added models..... 3
Multilevel linear growth model initial status, focal year growth, and average growth (MLM0, MLM growth rate and MLM average growth rate)..... 3
Comparison of Model Estimates..... 4
Comparison of School Ranks Based on Model Estimates..... 5
Summary..... 5
Pennsylvania Study..... 6
Method..... 6
Sample..... 6
Instrument..... 8
Results and Discussion..... 9
Section A: School Performance Estimates..... 9
Cohort stability..... 9
Comparison of models..... 10
Relation with school composition variables..... 12
Relation of model estimates to SWD school composition..... 13
Summary of Section A..... 14
Section B: School Ranks Based on School Performance Estimates..... 14
Comparison of cohorts..... 14
Comparison of models..... 20
Relation with school composition variables..... 24
Relation of school ranks with SWD school composition..... 25
Summary of Section B..... 26
Conclusion..... 26
References..... 28
Appendix A. Correlations among School Performance Model Estimates for Each Individual Cohort by Content Area and Grade Level Band..... 30

Appendix B. Correlations of School Performance Model Estimates with School Composition Variables for each Individual Cohort by Content Area and Grade Level Band..... 33
Appendix C. Correlations of School Performance Model Estimates with School Percentage SWD for each Individual Cohort by Content Area and Grade Level Band.. 36
Appendix D. Proportion of Elementary or Middle Schools within 5, 10, or 20 Ranks of Each Other for Each Pair of School Performance Models in Mathematics and English Language Arts by Cohort..... 37
Appendix E. RMSD in School Ranks for Pairs of School Performance Models for each Individual Cohort by Content Area and Grade Level Band..... 40
Appendix F. Correlations of School Ranks with School Composition Variables by Content Area and Grade Level Band for each Individual Cohort..... 43
Appendix G. Correlations of School Ranks with School Percentage SWD for each Individual Cohort by Content Area and Grade Level Band..... 47

List of Tables

1. Proportion and Standard Deviation (in parentheses) of Student Subgroups for the Pennsylvania Analytical Samples by Content Area and Grade Level Band..... 7
2. Correlations of School Performance Model Estimates Across Cohorts by Content Area and Grade Level Band..... 9
3. Average Correlations across Content Area and Grade Level Band and Overall Mean and Standard Deviation (SD) Across the Three Cohort Comparisons.....10
4. Correlations of School Performance Estimates across Models by Content Area and Grade Level Band.....10
5. Correlations of School Performance Model Estimates between Mathematics and English Language Arts by Grade Level Band in each Cohort and Averaged over Cohorts..... 11
6. Correlations of Model Estimates with School Composition Variables by Content Area and Grade Level Band.....12
7. Average School Performance Model Estimates as a Function of the Percentage of SWD in the School by Content and Grade Level Band..... 13
8. Spearman's Correlations of Model School Ranks for Each Pair of Cohorts by Content Area and Grade Level Band..... 15
9. Spearman's Correlations of Model School Ranks Averaged Across Content Area and Grade Level Band and Overall Mean and Standard Deviation (SD) Across the Three Cohort Comparisons..... 16
10. Proportion of Elementary or Middle Schools Within 5, 10, or 20 Ranks of Each Other for Each School Performance Model for Each Pair of Cohorts in Mathematics and English Language Arts.....16
11. RMSD in School Ranks for Each Student Cohort for Each School Performance Model by Content Area and Grade Level Band.....19
12. Proportion of Elementary or Middle Schools within 5, 10, or 20 Ranks of Each Other for Each Pair of School Performance Models in Mathematics and English Language Arts Averaged over Cohorts.....20
13. Average Across Cohorts of RMSD in School Ranks Between School Performance Models by Content Area and Grade Level Band.....22
14. Spearman's Correlations of School Performance Model Estimates Across Mathematics and English Language Arts Comprehension by Cohort.....23
15. Proportion of Elementary or Middle Schools within 5, 10, or 20 Ranks of Each Other in Mathematics versus English Language Arts for Each School Performance Model Averaged Over Cohorts.....23
16. RMSD in School Ranks for Mathematics and English Language Arts by Cohort and Grade Level Band and Overall Means and Standard Deviations (SD)..... 24
17. Spearman's Correlations of School Ranks With School Composition Variables by Content and Grade Level Band..... 24
18. Average School Rank as a Function of the Percentage of SWD in the School by Model, Content Area, and Grade Level Band.....26

A Comparison of Alternative Models for Estimating School Performance in Mathematics and Reading/Language Arts in Four State Accountability Systems: Pennsylvania Results

Background and Introduction

This technical report is one of a series of four technical reports that describe the results of a study comparing eight alternative models for estimating school academic performance using data from Arizona, North Carolina, Oregon, and Pennsylvania accountability systems. Our purpose was not to evaluate or examine the accountability systems in use by these states, but to evaluate a broader range of models commonly used for estimating school performance that are applied in many states and frequently reported in the school effectiveness research literature. This introduction briefly describes the study background and details the methods and procedures we used to estimate the eight school performance models and compare model. The individual state technical reports including details on each state's accountability data, assessment instruments, and results are provided at: http://www.ncaase.com/publications/tech-reports.

Despite the central importance of analytic models used in evaluating teacher and school effects in modern accountability systems, there are relatively few studies of the reliability and validity of these high-stakes systems (see, for example, Goldschmidt, Choi, & Beaudoin, 2012). The results reported here examine eight models using operational state accountability data in mathematics and reading/language arts from the four participating states. We addressed four questions surrounding the use of analytic models for the evaluation of school performance:

1. Are estimates of school performance stable across successive cohorts of students?
2. How well do estimates of school performance correlate among models?
3. How do estimates of school performance correlate with variables describing the student composition of the school?
4. Do estimates of school performance vary from one model to another based on the school composition of students with disabilities (SWD)?

General Method Description

Sample

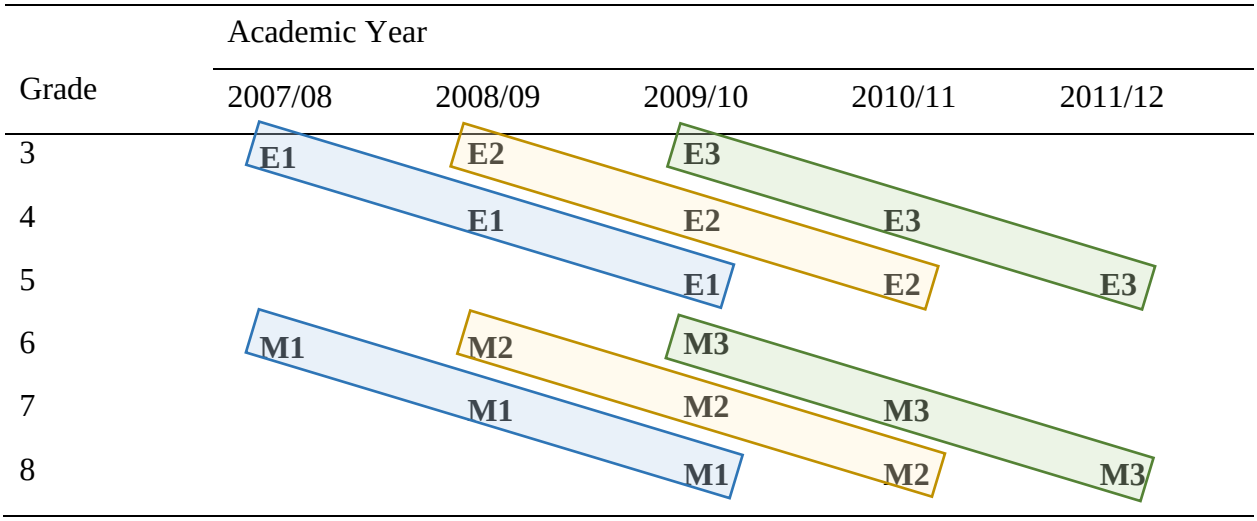
The sample from each state is described in each individual state technical report. In three of the four states, the sample consisted of all students who took the state's mathematics or reading/language arts general assessment in any one school year from 2007-08 through 2011-12, and whose records in each year were included in the state's calculation of Adequate Yearly Progress (AYP). Samples were separated into two grade level bands: a longitudinal elementary school sample (Grades 3 through 5) and a longitudinal middle school sample (Grades 6 through 8), each consisting of three cohorts (a) 2007/08 through 2009/2010; (b) 2008/09 through 2010/11; and (c) 2009/10 through 2011/12 (see research design schematic below). In Arizona, only one elementary and middle school cohort was used (2006/07 through 2008/09) due to changes in the Arizona testing program in 2010.

Instruments

The outcome measures for all analyses were the standardized mathematics and reading/language arts tests used for accountability in each state. In three of the states, the instruments used vertically linked developmental scales created using item response theory (IRT) methods. In Pennsylvania, the test was not vertically linked over grades preventing the

estimation of certain school performance models described in the next section. More detail about the Pennsylvania test is contained in the next section.

Research design indicating academic years and longitudinal cohorts studied:



Note. E denotes an elementary school cohort, M denotes a middle school cohort; only one elementary and one middle school cohort were available in the Arizona data.

School Performance Models

For all models, we estimated school performance in the last focal year (Grade 5 or 8) of the two grade level bands, adding prior years of achievement data as dictated by the particular model. We applied eight alternative analytic models of school performance to the mathematics and reading/language arts achievement data in elementary and middle school for each state. The eight school performance models were: Percent Proficient (PP), gain score (Gain), transition matrix (TM), student growth percentile (SGP), value-added model (VAM), and three Multilevel Linear Model (MLM) estimates: focal year intercept or status (MLM0), focal year growth rate (Grate), and average MLM growth rate across the three years (AvGrate). Because the Pennsylvania test was not vertically linked over grades, we could not apply models that required a vertical scale that were applied in the other states (AZ, NC, and OR), namely: the gain score model (Gain; focal year minus previous year); and three Multilevel Linear Models (MLM), focal year (Grade 5 or 8) intercept or status (MLM0), focal year growth rate (Grate), and average MLM growth rate across the three years (AvGrate). Although we did not apply all performance models to the Pennsylvania data, for completeness we include a brief description of all eight models here.

Percent Proficient (PP). PP was the NCLB required metric used by the state that calculated the percentage of students in each school that met or exceeded state benchmarks for proficiency in either mathematics or reading/language arts in each grade.

Average Gain Score. Gain scores were calculated as the prior academic year (Grade 4 or Grade 7) scale score in mathematics or reading/language arts subtracted from the focal year scale score (Grade 5 or Grade 8):

$$\text{Gain}_i = \Delta_i = Y_{it} - Y_{i(t-1)} \quad (1)$$

where Y_{it} was the assessment outcome for student *i* at time *t*. Student gain scores were averaged for each school (labeled "Gain" below).

Transition Matrix (TM). School performance estimates were computed from a table of the state's proficiency categories in the prior year crossed with the proficiency categories in the focal year (Grade 5 or Grade 8) which, in the case of five proficiency categories, created a transition matrix table of 25 cells. The percentage of students occurring in each of the cells was entered and then a weighting scheme was applied to each cell and the products were summed to create a TM school performance index. The weighting scheme awarded one of three scores: (a) -1 was recorded if the student moved down one or more categories from the previous year, (b) 0 was recorded if the student stayed in the same category, and (c) +1 was recorded if the student moved up one or more categories from the previous year (see Tindal, Nese, & Stevens, 2017). The weighted values were averaged across all cells to create an overall school TM index.

Student Growth Percentiles (SGP). Student growth percentiles were computed at the student level using the approach described by Betebenner (2009). A student's SGP was calculated by taking the current year test score and regressing it on the two prior years of test scores. Betebenner's (2009) approach uses ordinal methods (quantile regression) as well as B-spline, cubic polynomial smoothing of the resulting normative distribution of conditional regression estimates. The analysis results in a relative rank for each student in a conditional distribution of those who had similar scores in previous years. We used the R package *SGP* (Betebenner, & Iwaarden, 2011) to compute student estimates based on the regression of the two prior years of test scores on the current year's test score and then we aggregated student SGP for each school to create a median SGP as each school's SGP performance estimate.

Value-added Models (VAM). This mixed effects approach examined performance gains over years and included indicators for student membership in a particular school. This model is known generally as the "layered model" because layers of equations are added with each year of schooling (Ballou, Sanders, and Wright, 2004). For example, the model for our case with students with three years of data would be specified as follows:

$$Y_{0ij} = b_0 + u_0 + e_0 \quad (2a)$$
$$Y_{1ij} = b_1 + u_0 + u_1 + e_1 \quad (2b)$$
$$Y_{2ij} = b_2 + u_0 + u_1 + u_2 + e_2\ , \quad (2c)$$

where Y_{tij} represents an assessment for student *i* at time *t* (grade) attending school *j*. The fixed mean for all students in the combination of grades and schools was μ_{tij}, while e_{tij} was the random deviation for student *n* from the mean, μ_{tij}. The layered model we used was limited to a maximum of three years and was applied separately to mathematics and reading/language arts.

Multilevel Linear Growth Model Initial Status, Focal Year Growth, and Average Growth (MLM0, MLM Growth Rate and MLM Average Growth Rate). We modeled student growth over the three elementary or three middle school grades with multilevel longitudinal analyses (Raudenbush & Bryk, 2002) using HLM 7.1 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011) and full maximum likelihood estimation. The conditional models included a

level-1 model that specified student mathematics or reading/language arts scores predicted by a quadratic function of time of measurement, a level-2 model composed of the prediction of level-1 model parameters as a function of student mean values, and a level-3 model composed of the prediction of level-2 parameters as a function of school mean parameter values. Time was centered on the focal year (Grades 5 or 8) for computation of MLM0 and MLM growth rate but was centered on the middle year (Grades 4 or 7) for computation of MLM average growth rate. We used a quadratic model based on previous findings (Bloom, Hill, Black, & Lipsey, 2008) as well as inspection of the data and statistical testing of alternative growth functions. Because only three time points were present, the model intercept and linear slope were random parameters but the variance of the quadratic parameter was fixed (note the omission of a residual term in equation 4c below) to obtain a model solution. We used two different centering definitions to take into account the curvilinear nature of growth. Although centering in the last, focal year is most consistent with the definition of other models, it likely underestimates the amount of growth that occurs over the three year period because of deceleration. We therefore also centered on the middle grade in the three year span to produce an average growth rate over the three years. The resulting MLM model equations were:

Level 1 (Time):

$$(Y_{tij}) = \pi_{0ij} + \pi_{1ij}(\text{time}_{tij}) + \pi_{2ij}(\text{time squared}_{tij}) + e_{tij} \quad (3)$$

Level 2 (Students):

$$\pi_{0ij} = \beta_{00j} + r_{0ij} \quad (4a)$$
$$\pi_{1ij} = \beta_{10j} + r_{1ij} \quad (4b)$$
$$\pi_{2ij} = \beta_{20j} \quad (4c)$$

Level 3 (Schools):

$$\beta_{00j} = \gamma_{000} + u_{00j} \quad (5a)$$
$$\beta_{10j} = \gamma_{100} + u_{10j} \quad (5b)$$
$$\beta_{20j} = \gamma_{200} + u_{20j} \quad (5c)$$

where Y_{tij} was the mathematics or reading/language arts scale score for student i at time t in school j, π_{0ij} was the initial status or intercept for student i at time 0 in school j, π_{1ij} was the linear rate of change, π_{2ij} was the quadratic curvature representing the acceleration or deceleration in each student's growth trajectory and e_{tij} was the residual for each student. At level-2, the level-1 parameters were modeled using mean parameter values across students (β_{k0j}) and at level-3, the level-2 parameters were modeled using mean parameter values across schools (γ_{k0j}).

Comparison of Model Estimates

We used several comparison criteria to evaluate the comparability and stability of school estimates across school performance models and across cohorts. In each state technical report we describe the results of our evaluation of school performance estimates. We examined: (a) correlations of model estimates for each school across the three cohorts, (b) correlations among school estimates from one model to another, (c) correlations among the school estimates and school composition variables (e.g., percent free/reduced lunch in the school, percent minority students in the school), and (d) correlations of each model with the percentage of students with disabilities in the school.

Comparison of School Ranks Based on Model Estimates

Many states and districts create school ranks based on their accountability system results. To compare the alternative school performance models using this metric, we created school percentile ranks (from 1 to 99, with 99 being the highest performance) based on each of the school performance models described above. In one of the only studies evaluating school performance models, Goldschmidt, Choi, and Beaudoin (2012) compared models using quintiles. They examined the percentage of times schools remained in the same quintile band based on one school performance model versus another. Similarly, Castellano and Ho (2013) compared SGP and conditional regression models by examining the percentage of times schools remained within 1, 5 or 10 percentile ranks for each model. To maintain some comparability with each of these studies, we used three levels of similarity in school ranks, computing the percentage of schools within 5, 10, or 20 ranks of each other. We also computed the Spearman's correlation of school ranks from one cohort to another or from one school performance model to another. As a final comparison metric, we computed the root mean squared difference (RMSD) between school ranks based on each pair of cohorts or each pair of school performance models (see Castellano & Ho, 2013):

$$RMSD_{c,c} = \sqrt{\frac{\sum_{j=1}^{j}(Rank_{jc} - Rank_{jc})^2}{n}} \quad (6)$$

In equation 6, for a particular school performance model, the RMSD computes the difference ($Rank_{it}$) between each school's rank in one cohort (*jt*) versus the school's rank in a second cohort (*ju*), squaring the difference, summing across all schools, dividing by the number of schools, *n*, and taking the square root of the result.

$$RMSD_{mn} = \sqrt{\frac{\sum(Rank_{jm} - Rank_{jn})^2}{n}} \quad (7)$$

Similarly, in equation 7, the school ranks arising from alternative school performance models are compared in which $Rank_{jm}$ and $Rank_{jn}$ represent the rank of school *j* using school performance model m compared to that school's rank using school performance model *n*. As in equation 6, differences in ranks are then summed, squared, divided by the number of schools and taken to the ½ power. The RMSD was a measure of similarity in school performance models where a lower value indicates a pair of models that rank schools most similarly.

Summary

We evaluated eight models for estimating school academic performance in mathematics and reading/language arts using operational state accountability data; in PA, however, we evaluated four models. In NC, OR, and PA, we examined stability in model estimates across three successive student cohorts in mathematics and reading/language arts in both elementary and middle school grades. In all four states, we also compared the estimates of school performance from one model to another to determine whether the models provided similar or different depictions of school performance, although several models could not be estimated in Pennsylvania because their test did not have a vertically linked score scale. We then compared

the degree to which model estimates correlated with variables describing the student composition of the school, a likely indication of construct irrelevant variance. Ideally estimates of school performance should not be related to the student composition of the school. Last, we evaluated the school performance models in terms of the way they ranked schools, the stability of school ranks across cohorts, and the degree of agreement in school rankings from one school performance model to another. Detailed results of these analyses and comparisons follow for the state of Pennsylvania.

Pennsylvania Study

Method

Sample

The Pennsylvania sample was separated into an elementary school sample (Grades 3 through 5) and a middle school sample (Grades 6 through 8), each consisting of three successive cohorts of students enrolled in school years: (a) 2007/08 through 2009/2010; (b) 2008/09 through 2010/11; and (c) 2009/10 through 2011/12. The initial sample included students across the three cohorts whose Grade 5 (elementary school sample) or Grade 8 (middle school sample) Pennsylvania System of School Assessment (PSSA) English language arts or mathematics scores on the general or alternate assessment were included in the state calculation of Adequate Yearly Progress (AYP). There was a small number of cases where a unique student identifier appeared to have been associated with more than one student in a year. When conflicting reading or mathematics scores were associated with a student identifier, all records for that student identifier in that year were removed. The initial elementary school sample for the mathematics test was 393,065 students. The initial middle school sample for the mathematics test was 399,933 students. The initial elementary school sample for the reading/language arts test was 392,180 students. The initial middle school sample for the reading/language arts test was 398,951 students.

To create an analytic sample that was appropriate for our research questions, we only included students with valid reading or mathematics general assessment scores in all three grades (Grades 3 through 5, or Grades 6 through 8). Students who did not follow the typical grade level sequence due to grade retention, acceleration, or dubious progressions were excluded from the sample; this included the transition from 2006/07 to 2007/08, so that no students present in 2007/08 had been retained or accelerated from the previous year. We included only schools that served the grade spans 3 to 5 or 6 to 8, and schools with $N \geq 10$ students in each of the three cohorts in the final reference year of the three-year grade level band (i.e., Grade 5 for elementary Grades 3 to 5 and Grade 8 for middle Grades 6 to 8). Students and schools that did not meet these criteria were excluded from analyses. As is the case in most operational and research applications of these models, we made no attempt to account for student mobility in years prior to the focal year or to make any attributions of "school effects" based on how many years the student had been in the focal year school. Our strategy in creating the analytic sample was to maximize the interpretation of comparisons of the models rather than to ensure complete representativeness of the samples. These inclusion rules were applied to ensure that there were no differences in the analytic samples for different school models so that comparisons of school models were a function only of differences in the models and not the composition of the sample analyzed. The final elementary school analytic sample for the mathematics test was 257,811 students (65.6% of the initial sample). The final middle school analytic sample for the

mathematics test was 213,873 students (53.5%). The final elementary school analytic sample for the reading/language arts test was 252,035 students (64.3%). The final middle school analytic sample for the reading/language arts test was 209,923 students (52.6%).

Table 1 provides summary statistics describing the school-level analytical samples of Pennsylvania elementary and middle school students in the three cohorts for mathematics and English language arts. Although variation existed from cohort to cohort in sample demographic characteristics, generally the composition of the samples was quite similar across the three cohorts and for mathematics and English language arts at each grade level band. From elementary to middle school cohorts, there were small but consistent decreases in the proportion of English learners (EL), economically disadvantaged students (EDS), racial/ethnic minority students (i.e., American Indian/Alaskan Native, Asian/Pacific Islander, Black/African American, Hispanic, Multi-Ethnic, and Declined to report), and students with disabilities (SWD). At the elementary school level, about 9% (English Language Arts) and 13% (Mathematics) of the students were EL, almost 50% of the students were female, about 46% were EDS, approximately 30% were racial/ethnic minority students, and about 18% were SWD. At the middle school level, about 13% (English Language Arts) and 17% (Mathematics) of the students were EL, 50% of the students were female, about 50% were EDS, approximately 40% were racial/ethnic minority students, and about 15% to 18% were SWD. It is also noteworthy that there was much greater school level variation–as indicated by the values of the standard deviations in parentheses–in EDS and racial/ethnic minority student school composition (and also EL at the middle school level) than other student characteristics. It should also be noted that when we refer to "school" composition, it references variables representing a particular cohort in each school in our analytic samples. Because we excluded students and schools to create our analytic samples, "total school" characteristics may differ slightly from the variables reported here.

Table 1

Proportion and Standard Deviation (in parentheses) of Student Subgroups for the Pennsylvania Analytical Samples by Content Area and Grade Level Band

		Cohort		
		1	2	3
Mathematics Elementary	EL	0.133	0.148	0.153
		(0.266)	(0.286)	(0.285)
	Female	0.492	0.491	0.490
		(0.074)	(0.074)	(0.074)
	EDS	0.461	0.471	0.477
		(0.299)	(0.300)	(0.294)
	Ethnic Minority	0.306	0.309	0.316
		(0.348)	(0.347)	(0.346)
	SWD	0.181	0.158	0.157
		(0.077)	(0.069)	(0.068)
English Language Arts Elementary	EL	0.089	0.066	0.151
		(0.215)	(0.164)	(0.283)

	Female	0.493	0.491	0.490
		(0.075)	(0.075)	(0.074)
	EDS	0.456	0.466	0.476
		(0.299)	(0.300)	(0.295)
	Ethnic Minority	0.298	0.299	0.316
		(0.346)	(0.344)	(0.346)
	SWD	0.181	0.157	0.157
		(0.078)	(0.070)	(0.068)
Mathematics Middle	EL	0.167	0.179	0.187
		(0.339)	(0.327)	(0.359)
	Female	0.495	0.505	0.496
		(0.069)	(0.068)	(0.076)
	EDS	0.501	0.515	0.524
		(0.314)	(0.313)	(0.312)
	Ethnic Minority	0.412	0.419	0.422
		(0.389)	(0.390)	(0.388)
	SWD	0.180	0.146	0.149
		(0.081)	(0.065)	(0.070)
English Language Arts Middle	EL	0.129	0.121	0.177
		(0.309)	(0.255)	(0.333)
	Female	0.495	0.506	0.497
		(0.069)	(0.069)	(0.077)
	EDS	0.496	0.510	0.522
		(0.314)	(0.313)	(0.311)
	Ethnic Minority	0.405	0.411	0.419
		(0.390)	(0.390)	(0.387)
	SWD	0.180	0.145	0.149
		(0.081)	(0.067)	(0.070)

Instrument

The outcome measures for all analyses were the standardized Pennsylvania System of School Assessment (PSSA; Pennsylvania Department of Education [PDE], 2008, 2009, 2010, 2011, 2012) mathematics and English language arts tests. The PSSA is a summative, standards-based, criterion-referenced paper-pencil assessment aligned with PA Academic Standards and designed to assess knowledge and skills described in the PA Assessment Anchor Content Standards (PDE, 2008, 2009, 2010, 2011, 2012) which vary by grade and content area. The PSSA mathematics and English language arts employs multiple-choice and open-ended item-types, and were administered under standardized conditions (PDE, 2008, 2009, 2010, 2011, 2012). PSSA raw scores were converted to scale scores based on the total test score while taking item difficulty into account using one parameter item response theory (IRT) methods. Each grade and content area has its own unique PSSA scaled score and a chained linking design (within-year linking) was used to place the item parameters and student ability estimates on the same scale

across forms (within grade and content area). The PSSA was not designed to have a developmental scale score that could be applied across grades.

Results and Discussion

This technical report is organized in three sections: Section A describes school performance model estimates, Section B describes school ranks, and several Appendices provide additional detailed results.

Section A: School Performance Estimates

Cohort stability. We first considered the stability of model estimates by computing the correlations among estimates across the three successive cohorts of students. It should be noted that cohort comparisons are both an indication of changes in the composition of students in the school from one academic year to another as well as any other temporal changes that occur from one year to another including changes in policy, practice, instruction, or other factors that impact student test scores. Table 2 shows the correlation of model estimates across cohorts for mathematics and English language arts in the elementary school and middle school samples. As can be seen in Table 2, correlations generally ranged from very low (.003 for TM 1 with 3) to large (.857 for PP 1 with 2) for the model estimates indicating some stability in school performance estimates across cohorts for the PP estimates, but little stability for the other models. Correlations between adjacent years in the first two columns (cohort 1 with 2 or 2 with 3) are generally larger than the comparisons across two years (cohort 1 with 3). Although there is also some variation from elementary to middle school or from mathematics to English language arts, trends in cohort stability were fairly similar across content area and grade level band.

Table 2

Correlations of School Performance Model Estimates across Cohorts by Content Area and Grade Level Band

Elementary Schools

	Mathematics			English Language Arts		
Model	1 with 2	2 with 3	1 with 3	1 with 2	2 with 3	1 with 3
PP	0.806	0.774	0.768	0.798	0.783	0.782
TM	0.411	0.334	0.258	0.292	0.136	0.031
SGP	0.525	0.448	0.344	0.401	0.373	0.207
VAM	0.568	0.466	0.362	0.463	0.434	0.240

Middle Schools

	Mathematics			English Language Arts		
Model	1 with 2	2 with 3	1 with 3	1 with 2	2 with 3	1 with 3
PP	0.857	0.852	0.818	0.833	0.854	0.821
TM	0.276	0.316	0.126	0.185	0.204	0.003
SGP	0.532	0.469	0.346	0.414	0.426	0.186
VAM	0.551	0.515	0.367	0.484	0.561	0.299

To facilitate interpretation of the cohort results, we also averaged correlations across the two content areas and grade levels (see Table 3). It can be seen that the correlations across cohorts were greatest for the status based school performance measure (PP) and noticeably lesser for all other models, particularly for TM model estimates. The two rightmost columns of Table 3 show the overall mean and standard deviation across the cohort comparisons for each school performance model. It can be seen that the greatest agreement over cohorts, content, and grade level was for the PP model estimates. All remaining multi-year performance models had greater instability. The standard deviations of correlations across cohort comparisons shown in the rightmost column of Table 3 also show the least variability over cohorts for the status model and the greatest variability across cohort correlations for the VAM model.

Table 3

Average Correlations across Content Area and Grade Level Band and Overall Mean and Standard Deviation (SD) Across the Three Cohort Comparisons

Model	1 with 2	2 with 3	1 with 3	Mean	SD
PP	0.824	0.816	0.797	0.812	0.017
TM	0.291	0.248	0.104	0.214	0.105
SGP	0.468	0.430	0.270	0.389	0.107
VAM	0.516	0.494	0.317	0.442	0.114
Mean	0.525	0.497	0.372	--	--

Comparison of models. We next computed the correlations of school performance estimates from one model to another within each of the three cohorts and then took the mean correlation across cohorts. Correlations of model estimates within each individual cohort are presented in Appendix A. Table 4 shows model correlations for mathematics and English language arts in the elementary school and middle school samples averaged over the three cohorts.

Table 4

Correlations of School Performance Estimates across Models by Content Area and Grade Level Band

Elementary School Mathematics

Model	TM	SGP	VAM
PP	0.441	0.539	0.573
TM		0.869	0.875
SGP			0.964

Elementary School English Language Arts

Model	TM	SGP	VAM
PP	0.190	0.582	0.658
TM		0.735	0.712
SGP			0.943

Middle School Mathematics

Model	TM	SGP	VAM
PP	0.326	0.489	0.55
TM		0.830	0.820
SGP			0.965

Middle School English Language Arts

Model	TM	SGP	VAM
PP	0.272	0.509	0.593
TM		0.688	0.669
SGP			0.934

Average over Content Area and Grade Level Band

Model	TM	SGP	VAM
PP	0.307	0.530	0.594
TM		0.781	0.769
SGP			0.952

As evident in Table 4, the degree of model agreement depended on which models were being compared. As shown by the averages in the last panel of Table 4, across content area and grade level band, the largest correlations were among the SGP and VAM models (+.952), the TM and SGP models (+.781), and the TM and VAM models (+.769). The smallest correlation was between the TM and PP models (+.307). The average correlation of the PP and SGP models (+.530) and the PP and VAM models (+.594) were moderate in magnitude, but it should be noted that even moderate correlations may indicate substantial disagreement between model estimates. For example, the PP-VAM correlation of +.594 indicates that about 35% of variance in school estimates is shared across the two models but 65% of the variance of estimates across the two models is not in agreement.

We also examined the degree to which school performance model estimates were consistent from one content area to the other. Table 5 shows model estimate agreement across content areas in each cohort as well as the average across the three cohorts. As can be seen in Table 5, correlations were generally larger between content areas in elementary than middle school. On average, the correlations for the status models (PP) were larger than +.849 and were also larger than the average correlations for the other models that ranged from +.303 to +.724.

Table 5

Correlations of School Performance Model Estimates between Mathematics and English Language Arts by Grade Level Band in each Cohort and Averaged over Cohorts

	Elementary Schools				Middle Schools			
	Cohort				Cohort			
Model	1	2	3	Mean	1	2	3	Mean

PP	0.860	0.860	0.888	0.869	0.849	0.868	0.877	0.865
TM	0.505	0.510	0.618	0.544	0.388	0.303	0.438	0.376
SGP	0.565	0.578	0.652	0.598	0.431	0.422	0.473	0.442
VAM	0.626	0.623	0.724	0.658	0.497	0.476	0.567	0.513

Relation with school composition variables. We computed the correlation of model estimates with school composition variables to determine whether estimates were related to the aggregated student characteristics in each school. Table 6 shows the correlations of model estimates with school composition variables for mathematics and reading/language arts in the elementary school and middle school samples. Correlations of model estimates with school composition variables within each individual cohort are presented in Appendix B.

The rightmost column of Table 6 shows the average correlation of each school performance model with the school composition variables. As can be seen, correlations of the status models, PP, were negative and noticeably larger than the correlations of the other school performance models with school composition variables. On average across content and grade level band, the correlation of the school composition variables was -0.247 for the PP model. In contrast, the average correlations of the school composition variables with the remaining models were noticeably smaller, ranging from -0.087 (VAM) to -0.018 (TM). Thus there was relatively little relation of the multiyear models with school composition, but for the status model, school performance estimates were higher the fewer the number of students from protected groups present in the school and lower as the number of students from protected groups increased. No clear pattern was present for the relation between school size and model estimates.

Table 6

Correlations of Model Estimates with School Composition Variables by Content Area and Grade Level Band

Elementary School Mathematics

Models	EDS	EL	SWD	Female	Minority	School Size	Mean
PP	-0.702	-0.184	-0.150	-0.033	-0.632	0.182	-0.253
TM	-0.183	-0.026	0.012	-0.002	-0.103	0.037	-0.044
SGP	-0.253	-0.013	-0.006	-0.022	-0.157	0.052	-0.067
VAM	-0.295	-0.025	-0.009	-0.024	-0.186	0.060	-0.080

Elementary School English Language Arts

Models	EDS	EL	SWD	Female	Minority	School Size	Mean
PP	-0.759	-0.136	-0.163	0.002	-0.647	0.240	-0.244

TM	0.024	-0.010	0.036	-0.020	0.044	-0.035	0.006
SGP	-0.297	-0.014	-0.031	-0.004	-0.202	0.084	-0.077
VAM	-0.368	-0.026	-0.045	-0.005	-0.272	0.103	-0.102

Middle School Mathematics

Models	EDS	EL	SWD	Female	Minority	School Size	Mean
PP	-0.706	-0.241	-0.286	-0.018	-0.674	0.403	-0.254
TM	-0.087	0.011	-0.027	-0.012	-0.105	0.046	-0.029
SGP	-0.210	-0.032	-0.061	-0.015	-0.200	0.106	-0.069
VAM	-0.268	-0.047	-0.076	-0.012	-0.252	0.140	-0.086

Middle School English Language Arts

Models	EDS	EL	SWD	Female	Minority	School Size	Mean
PP	-0.723	-0.200	-0.345	0.058	-0.636	0.417	-0.238
TM	0.014	0.007	-0.062	0.015	-0.002	-0.010	-0.006
SGP	-0.223	-0.013	-0.091	0.037	-0.122	0.143	-0.045
VAM	-0.331	-0.043	-0.129	0.036	-0.219	0.215	-0.079

Relation of model estimates to SWD school composition. Because of the NCAASE emphasis on the performance and academic growth of SWD, we also focused more specifically on the relations between the percentage of SWD students served by a school and the school performance model estimates. Correlations of model estimates with SWD school composition within each individual cohort are presented in Appendix C. Table 7 shows the correlation of model estimates with the percentage of SWD in each school for mathematics and English language arts in the elementary school and middle school samples averaged over cohorts. As can be seen in the bottom row of Table 7, average school performance estimates based on the status model (PP) had substantially larger negative correlations with school SWD composition than the other school performance models. With the PP model, school performance estimates were higher the smaller the percentage of SWD students in the school and smaller to the extent that the school served larger proportions of SWD.

Table 7

Average School Performance Model Estimates as a Function of the Percentage of SWD in the School by Content and Grade Level Band

Content Area and	PP	TM	SGP	VAM

Grade Level Band				
Math Elementary	-0.150	0.012	-0.006	-0.009
Math Middle	-0.286	-0.027	-0.061	-0.076
English Language Arts Elementary	-0.163	0.036	-0.031	-0.045
English Language Arts Middle	-0.345	-0.062	-0.091	-0.129
Mean	-0.236	-0.010	-0.047	-0.065

Summary of Section A. We evaluated four alternative models for estimating school academic performance in mathematics and English language arts using operational Pennsylvania state accountability data. We observed limited stability in model estimates across three successive student cohorts in mathematics and English language arts in both elementary and middle school grades. We also compared the estimates of school performance from one model to another and found substantial disagreement across models. Generally, the status model (PP) based on a single year of data differed from the remaining models that examined more than one year of data. There was greater agreement among the models that used multiple years of data.

We also compared school performance estimates in mathematics with those in English language arts. Again, agreement was greater across content areas for the status models than for the multiple year models. Comparison of model estimates with school composition variables showed that, compared to the remaining school performance models, the status model (PP) had substantially larger correlations with the student makeup of the school; lower PP estimates were related to larger proportions of protected student subgroups in the school. Finally, we correlated school performance estimates with the percentage of SWD in each school. Ideally, estimates of school performance should be unrelated to the student composition of the school, but as with the other school composition variables, we found that the status model (PP) was more highly correlated with SWD school composition than the multiyear model estimates.

Section B: School Ranks Based on School Performance Estimates

In this section, we focus on the examination of school ranks based on the school performance estimates reported in the previous section. It is common practice for states and other jurisdictions to rank schools as a method for evaluating and reporting academic performance. Therefore, using the estimates of school performance generated by the four models described previously, we computed percentile ranks for each school (from 1, lowest to 99, highest). We then compared school ranks within each school performance model across the three cohorts used in the study. Next, we compared the school ranks for each model to the ranks obtained from each of the other models. Finally, we examined the relation between school ranks from each model with variables describing the student composition of each school. Three criteria were used to evaluate the comparisons of school ranks: (a) the Spearman's correlation between school ranks, (b) the proximity of absolute school ranks, and (c) the root mean square difference (RMSD) in school ranks.

Comparison of cohorts. We first consider the stability of school ranks within each school performance model across the three successive cohorts of students in mathematics and reading/language arts in the elementary and middle school grades. We computed the Spearman's correlation of the school ranks from one cohort to the school ranks from each of the other two

cohorts within each of the four school performance models to determine the stability of school ranks. As mentioned in Section A, cohort comparisons are both an indication of changes in the composition of students in the school from one academic year to another as well as any other temporal changes that occur from one year to another including changes in policy, practice, instruction, or other factors that impact student test scores. Table 8 shows the correlation of school ranks across cohorts for mathematics and English language arts in the elementary school and middle school samples. As can be seen in Table 8, the correlations ranged from small to large, indicating substantial variability in school ranks from one cohort to another. As would be expected, correlations between adjacent years in the first two columns (cohort 1 with 2 or 2 with 3) were generally somewhat larger than the comparison across two years (cohort 1 with 3). Although there was some variation, results were generally similar from elementary to middle school or from mathematics to English language arts.

Table 8

Spearman's Correlations of Model School Ranks for Each Pair of Cohorts by Content Area and Grade Level Band

Elementary Schools

	Mathematics			English Language Arts		
Model	1 with 2	2 with 3	1 with 3	1 with 2	2 with 3	1 with 3
PP	0.789	0.750	0.756	0.775	0.768	0.766
TM	0.416	0.351	0.273	0.285	0.161	0.044
SGP	0.503	0.451	0.347	0.383	0.401	0.220
VAM	0.539	0.482	0.379	0.433	0.480	0.277

Middle Schools

	Mathematics			English Language Arts		
Model	1 with 2	2 with 3	1 with 3	1 with 2	2 with 3	1 with 3
PP	0.839	0.844	0.801	0.811	0.844	0.787
TM	0.320	0.357	0.236	0.230	0.204	0.112
SGP	0.520	0.470	0.356	0.428	0.446	0.242
VAM	0.562	0.519	0.395	0.501	0.572	0.334

To facilitate further interpretation, we averaged the results shown in Table 8 across content area and grade level band. As can be seen in Table 9, on average the greatest stability was for the status model (PP). Noticeably smaller correlations occurred for the remaining school

performance models, all of which were based on more than one year of data, with the TM model showing the least stability.

Table 9

Spearman's Correlations of Model School Ranks Averaged across Content Area and Grade Level Band and Overall Mean and Standard Deviation (SD) Across the Three Cohort Comparisons

Model	1 with 2	2 with 3	1 with 3	Mean	SD
PP	0.804	0.802	0.778	0.795	0.019
TM	0.313	0.268	0.166	0.249	0.079
SGP	0.458	0.442	0.291	0.397	0.094
VAM	0.509	0.513	0.346	0.456	0.099

Our second criterion for comparing school ranks was to determine how much a school's rank changed from one cohort to another. Table 10 shows the proportion of schools that were within 5, 10, or 20 ranks in one cohort versus another for each school performance model in mathematics and English language arts at each grade level band. The last table entry for each school performance model shows the average differences in school ranks averaged over content area and grade level band. It can be seen that on average for the PP model, about one third of the schools differed by only 5 percentile ranks or less, over 50% of schools differed by 10 ranks or less, and more than 75% differed by 20 ranks or less. However, the level of agreement in school ranks across cohorts was noticeably lower for all of the remaining models that were based on two or more years of achievement data. For example, school ranks based on the remaining models differed by more than 20 ranks for about 50% or more of the schools.

Table 10

Proportion of Elementary or Middle Schools Within 5, 10, or 20 Ranks of Each Other for Each School Performance Model for Each Pair of Cohorts in Mathematics and English Language Arts

PP

	Cohort	r = 5	r = 10	r = 20
Mathematics Elementary	1 vs. 2	0.308	0.520	0.760
	2 vs. 3	0.273	0.477	0.735
	1 vs. 3	0.293	0.468	0.729
English Language Arts Elementary	1 vs. 2	0.312	0.506	0.751
	2 vs. 3	0.276	0.485	0.745
	1 vs. 3	0.281	0.488	0.750
Mathematics Middle	1 vs. 2	0.372	0.595	0.825
	2 vs. 3	0.349	0.574	0.840

	1 vs. 3	0.339	0.547	0.784
English Language Arts Middle	1 vs. 2	0.351	0.533	0.798
	2 vs. 3	0.367	0.585	0.826
	1 vs. 3	0.380	0.561	0.798
Mean	1 vs. 2	0.336	0.538	0.784
	2 vs. 3	0.316	0.530	0.786
	1 vs. 3	0.323	0.516	0.765

TM

	Cohort	r = 5	r = 10	r = 20
Mathematics Elementary	1 vs. 2	0.194	0.314	0.550
	2 vs. 3	0.164	0.305	0.514
	1 vs. 3	0.150	0.278	0.477
English Language Arts Elementary	1 vs. 2	0.150	0.270	0.487
	2 vs. 3	0.125	0.237	0.439
	1 vs. 3	0.130	0.224	0.404
Mathematics Middle	1 vs. 2	0.189	0.319	0.523
	2 vs. 3	0.188	0.318	0.537
	1 vs. 3	0.175	0.330	0.518
English Language Arts Middle	1 vs. 2	0.149	0.243	0.475
	2 vs. 3	0.169	0.286	0.453
	1 vs. 3	0.134	0.236	0.431
Mean	1 vs. 2	0.170	0.286	0.509
	2 vs. 3	0.162	0.286	0.486
	1 vs. 3	0.147	0.267	0.458

SGP

	Cohort	r = 5	r = 10	r = 20
Mathematics Elementary	1 vs. 2	0.212	0.347	0.570
	2 vs. 3	0.187	0.322	0.541

	1 vs. 3	0.176	0.290	0.514
English Language Arts Elementary	1 vs. 2	0.174	0.307	0.512
	2 vs. 3	0.173	0.320	0.527
	1 vs. 3	0.150	0.261	0.471
Mathematics Middle	1 vs. 2	0.218	0.377	0.586
	2 vs. 3	0.221	0.358	0.579
	1 vs. 3	0.207	0.335	0.542
English Language Arts Middle	1 vs. 2	0.211	0.346	0.569
	2 vs. 3	0.186	0.329	0.576
	1 vs. 3	0.174	0.304	0.497
Mean	1 vs. 2	0.204	0.344	0.559
	2 vs. 3	0.192	0.332	0.556
	1 vs. 3	0.177	0.298	0.506

VAM

	Cohort	r = 5	r = 10	r = 20
Mathematics Elementary	1 vs. 2	0.208	0.357	0.585
	2 vs. 3	0.190	0.327	0.576
	1 vs. 3	0.159	0.302	0.505
English Language Arts Elementary	1 vs. 2	0.180	0.326	0.530
	2 vs. 3	0.192	0.350	0.563
	1 vs. 3	0.176	0.289	0.488
Mathematics Middle	1 vs. 2	0.218	0.384	0.626
	2 vs. 3	0.209	0.351	0.604
	1 vs. 3	0.214	0.353	0.565
English Language Arts Middle	1 vs. 2	0.186	0.330	0.583
	2 vs. 3	0.202	0.383	0.594
	1 vs. 3	0.167	0.309	0.506
Mean	1 vs. 2	0.198	0.349	0.581

2 vs. 3	0.198	0.353	0.584
1 vs. 3	0.179	0.313	0.516

Our third criterion for comparing school ranks was to calculate the root mean square difference (RMSD) between cohorts or models as defined in the report introduction and general methods. Table 11 shows the RMSD across pairs of cohorts by content area and grade level band for each of the four school performance models and in the last two columns the mean and standard deviation (SD) across cohort comparisons. As can be seen in the table, the smallest differences in rank were for the PP model, about 17 to 20 ranks on average. Average differences in school rank across cohorts for the remaining models ranged from about 29 to 37.

Table 11

RMSD in School Ranks for each Student Cohort for each School Performance Model by Content Area and Grade Level Band

Elementary School Mathematics

Model	1 with 2	2 with 3	1 with 3	Mean	SD
PP	18.554	20.205	19.949	19.569	0.889
TM	30.862	32.523	34.432	32.606	1.786
SGP	28.471	29.931	32.630	30.344	2.110
VAM	27.433	29.066	31.819	29.439	2.217
Mean	26.330	27.931	29.708	--	--

Elementary School English Language Arts

Model	1 with 2	2 with 3	1 with 3	Mean	SD
PP	19.142	19.444	19.547	19.378	0.210
TM	34.157	36.996	39.477	36.877	2.662
SGP	31.734	31.256	35.671	32.887	2.423
VAM	30.408	29.118	34.336	31.287	2.718
Mean	28.860	29.203	32.258	--	--

Middle School Mathematics

Model	1 with 2	2 with 3	1 with 3	Mean	SD
PP	16.171	15.943	18.008	16.707	1.132
TM	33.271	32.359	35.273	33.634	1.491
SGP	27.954	29.373	32.368	29.898	2.253
VAM	26.689	27.985	31.376	28.683	2.420

Mean	26.021	26.415	29.256	--	--

Middle School English Language Arts

Model	1 with 2	2 with 3	1 with 3	Mean	SD
PP	17.543	15.925	18.601	17.356	1.348
TM	35.388	36.001	38.024	36.471	1.379
SGP	30.517	30.057	35.132	31.902	2.807
VAM	28.495	26.407	32.932	29.278	3.332
Mean	27.986	27.098	31.172	--	--

Comparison of models. We next compared school ranks from one model to another within each of the three cohorts. Comparisons of school ranks within each individual cohort were computed and are presented in Appendix D. We averaged those results by taking the median absolute difference in school ranks over the three cohorts in mathematics and reading/language arts in the elementary and middle school grades. For each pair of school performance models, Table 12 shows the average percentage of schools that were within 5, 10, or 20 percentile ranks in one model versus the other. As can be seen in the table, the SGP and VAM models ranked schools most similarly, over 75% of schools were within 10 ranks and over 95% were within 20 ranks for these two models. The level of agreement in school ranks was lower when comparing either the SGP or the VAM models with the TM model. The PP (status) model school rankings agreed with the multiyear models within 20 ranks in about 43% to 66% of schools.

The lowest agreement in ranks occurred between the PP and TM rankings, ranging from about 43% to about 54% of school within 20 ranks of each other.

Table 12

Proportion of Elementary or Middle Schools within 5, 10, or 20 Ranks of Each Other for Each Pair of School Performance Models in Mathematics and English Language Arts Averaged over Cohorts

Model Comparison:	r = 5	r = 10	r = 20
PP vs. TM			
Math Elementary	0.156	0.304	0.539
English Language Arts Elementary	0.135	0.239	0.432
Math Middle	0.156	0.274	0.482
English Language Arts Middle	0.144	0.262	0.441
Mean	0.148	0.270	0.474
PP vs. SGP			
Math Elementary	0.183	0.331	0.576
English Language Arts Elementary	0.199	0.359	0.602

Math Middle	0.182	0.315	0.544
English Language Arts Middle	0.197	0.336	0.561
Mean	0.190	0.335	0.571

PP vs. VAM

Math Elementary	0.209	0.367	0.612
English Language Arts Elementary	0.241	0.410	0.656
Math Middle	0.209	0.346	0.582
English Language Arts Middle	0.213	0.374	0.607
Mean	0.218	0.374	0.614

TM vs. SGP

Math Elementary	0.394	0.601	0.846
English Language Arts Elementary	0.264	0.444	0.687
Math Middle	0.373	0.589	0.798
English Language Arts Middle	0.249	0.426	0.643
Mean	0.320	0.515	0.744

TM vs. VAM

Math Elementary	0.382	0.601	0.844
English Language Arts Elementary	0.244	0.419	0.667
Math Middle	0.325	0.530	0.780
English Language Arts Middle	0.233	0.383	0.617
Mean	0.296	0.483	0.727

SGP vs. VAM

Math Elementary	0.626	0.858	0.981
English Language Arts Elementary	0.512	0.758	0.955
Math Middle	0.654	0.870	0.985
English Language Arts Middle	0.500	0.765	0.953
Mean	0.573	0.813	0.968

Our last criterion for comparing school ranks across cohorts was the RMSD between pairs of school performance model rankings. Appendix E shows the RMSD between pairs of school performance model rankings for each individual cohort. Table 13 shows the RMSD averaged over the three cohorts by content area and grade level band. The RMSD values reflect the same patterns of results for models as described previously. The greatest agreement in average ranks was between the SGP and VAM models for which schools differed by about 10 ranks or less on average. Much larger differences (about 23 ranks or more on average) occurred between the PP and the other school performance models. Agreement in school ranks between the remaining models was generally in the range of 14 to 25 ranks on average.

Table 13

Average across Cohorts of RMSD in School Ranks between School Performance Models by Content Area and Grade Level Band

Elementary School Mathematics

Model	TM	SGP	VAM
PP	29.284	27.322	25.55
TM		15.004	14.434
SGP			7.506

Elementary School English Language Arts

Model	TM	SGP	VAM
PP	36.044	26.107	23.379
TM		21.618	22.477
SGP			9.697

Middle School Mathematics

Model	TM	SGP	VAM
PP	32.676	29.375	27.372
TM		16.988	17.471
SGP			7.232

Middle School English Language Arts

Model	TM	SGP	VAM
PP	36.556	28.527	25.668
TM		24.354	24.763
SGP			9.747

We also evaluated the extent to which school ranks agreed from one content area to the other. Table 14 shows the Spearman's correlation of school ranks in mathematics with school ranks in English language arts by cohort and grade level band. The table also shows the mean correlation across cohorts at the two grade level bands. As can be seen in Table 14, on average correlations of school ranks across mathematics and English language arts in elementary schools ranged from +.500 to +.846 for the different school performance models. For middle schools, the average correlations ranged from +.307 to +.838. Correlations were larger for the status

models and smaller for the multiyear models at both grade level bands. Average correlations at the middle school level were also consistently smaller than for elementary schools for all models.

Table 14

Spearman's Correlations of School Performance Model Estimates across Mathematics and English Language Arts by Cohort

	Elementary Schools				Middle Schools			
Model	Cohort 1	Cohort 2	Cohort 3	Mean	Cohort 1	Cohort 2	Cohort 3	Mean
PP	0.839	0.834	0.865	0.846	0.816	0.848	0.849	0.838
TM	0.474	0.477	0.548	0.500	0.330	0.247	0.345	0.307
SGP	0.537	0.553	0.627	0.572	0.393	0.389	0.445	0.409
VAM	0.591	0.603	0.699	0.631	0.469	0.449	0.530	0.483

Table 15 shows the proportion of schools that shared similar ranks in mathematics as in reading/language arts for each school performance model by school level and averaged over grade level band. Similar to results previously described, Table 15 shows greater agreement for the PP model than the other school performance models with over 80% of the schools having ranks within 20 places across grade level bands. In contrast, there was substantially less agreement across the two content areas for the remaining, multiyear models with only approximately 50% to 64% of schools agreeing within 20 ranks for most models in either grade level band.

Table 15

Proportion of Elementary or Middle Schools within 5, 10, or 20 Ranks of Each Other in Mathematics versus English Language Arts for Each School Performance Model Averaged Over Cohorts

Model Comparison		r = 5	r = 10	r = 20
PP				
	Elementary	0.367	0.567	0.823
	Middle	0.364	0.575	0.809
	Mean	0.366	0.571	0.816
TM				
	Elementary	0.203	0.337	0.559
	Middle	0.172	0.287	0.491
	Mean	0.188	0.312	0.525
SGP				
	Elementary	0.224	0.379	0.600
	Middle	0.176	0.306	0.522

Mean	0.200	0.342	0.561
VAM			
Elementary	0.233	0.396	0.635
Middle	0.180	0.324	0.547
Mean	0.207	0.360	0.591

Calculation of the RMSD in school ranks for mathematics versus reading/language arts by cohort and grade level band and averaged over cohorts showed similar results (see Table 16). The difference in school ranks averaged over cohorts for the PP model was about 16. Average differences in rank across the two content areas were substantially greater for the remaining models ranging from 22 to about 35 depending on model and grade level band.

Table 16

RMSD in School Ranks for Mathematics and English Language Arts by Cohort and Grade Level Band and Overall Means

	Elementary Schools				Middle Schools			
Model	Cohort 1	Cohort 2	Cohort 3	Mean	Cohort 1	Cohort 2	Cohort 3	Mean
PP	16.229	16.443	14.816	15.829	17.214	15.636	15.641	16.164
TM	29.297	29.211	27.152	28.553	32.893	34.917	32.543	33.451
SGP	27.495	27.001	24.673	26.390	31.327	31.460	29.932	30.906
VAM	25.829	25.446	22.168	24.481	29.313	29.868	27.586	28.922

Relation with school composition variables. We computed the correlation of school ranks based on each school performance model with school composition variables to determine whether estimates were related to the aggregated student characteristics in each school. Table 17 shows these correlations for mathematics and English language arts in the elementary school and middle school samples. Correlations of model estimates with school composition variables within each individual cohort are presented in Appendix F. The rightmost column of Table 17 shows the correlation of each school performance model averaged over all of the school composition variables. As can be seen, correlations of the status model (PP) ranged from -.179 to -.216 depending on content and grade level band, and were noticeably larger than the correlations of the other school performance models with school composition variables, which ranged from -.077 to +.006 depending on content and grade level band.

Table 17

Spearman's Correlations of School Ranks With School Composition Variables by Content Area and Grade Level Band

Elementary School Mathematics

	Ethnic	School

Model	EDS	EL	SWD	Female	Minority	Size	Mean
PP	-0.689	-0.075	-0.149	-0.036	-0.446	0.224	-0.195
TM	-0.194	0.012	0.004	-0.002	-0.080	0.078	-0.030
SGP	-0.260	0.033	-0.007	-0.022	-0.115	0.101	-0.045
VAM	-0.315	0.033	-0.016	-0.026	-0.141	0.119	-0.058

Elementary School English Language Arts

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.750	-0.029	-0.148	-0.008	-0.428	0.290	-0.179
TM	0.021	-0.005	0.040	-0.021	0.026	-0.023	0.006
SGP	-0.310	0.021	-0.026	-0.015	-0.140	0.125	-0.058
VAM	-0.384	0.016	-0.045	-0.013	-0.190	0.154	-0.077

Middle School Mathematics

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.717	-0.107	-0.300	-0.024	-0.588	0.442	-0.216
TM	-0.087	0.036	-0.020	0.000	-0.104	0.055	-0.020
SGP	-0.207	0.003	-0.045	-0.013	-0.193	0.127	-0.055
VAM	-0.265	0.002	-0.065	-0.013	-0.236	0.170	-0.068

Middle School English Language Arts

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.745	-0.070	-0.360	0.030	-0.507	0.456	-0.199
TM	0.031	0.026	-0.043	0.008	0.002	0.002	0.004
SGP	-0.244	0.032	-0.087	0.032	-0.087	0.139	-0.036
VAM	-0.340	0.028	-0.128	0.035	-0.156	0.205	-0.059

Relation of school ranks with SWD school composition. We also specifically examined the relations between the percentage of SWD students served by a school and the school ranks based on the school performance model. Table 18 shows these correlations for mathematics and reading/language arts in the elementary school and middle school samples averaged over cohorts. Correlations of model estimates with SWD school composition within each individual cohort are presented in Appendix G. As can be seen in the bottom row of Table 18, on average there was a substantially larger negative correlation of the PP status model with school SWD composition (-0.239) than the other school performance models. With the PP

model, school ranks were higher with smaller percentages of SWD students in the school and school ranks were lower as schools served larger proportions of SWD. Little relation was present between school ranks based on the other models and SWD school composition.

Table 18

Average School Rank as a Function of the Percentage of SWD in the School by Model, Content Area, and Grade Level Band

Content Area and Grade Level Band	PP	TM	SGP	VAM
Math Elementary	-0.149	0.004	-0.007	-0.016
Math Middle	-0.300	-0.020	-0.045	-0.065
English Language Arts Elementary	-0.148	0.040	-0.026	-0.045
English Language Arts Middle	-0.360	-0.043	-0.087	-0.128
Mean	-0.239	-0.005	-0.041	-0.064

Summary of Section B. We evaluated the school ranks arising from four alternative models for estimating school academic performance in mathematics and English language arts across three sequential cohorts of students. As with the school performance estimates described in Section A, substantial variability in school ranks was present across the three student cohorts regardless of content area or grade level band. Using any of our comparison criteria (Spearman's correlations, absolute difference in ranks, RMSD), there was somewhat less variability across cohorts for the status model (PP) than for the models that used more than one year of data. When we compared school ranks arising from one model to school ranks from other models, we found disagreement across models. Generally, the PP status model differed from the remaining models that examined more than one year of data. Comparison of model estimates to school composition variables showed that the PP status model had substantially larger negative correlations than the remaining school performance models. Finally, we correlated school ranks arising from the four performance models with the percentage of SWD in each school. As with the school performance model estimates, we found that the status model was more strongly correlated with SWD school composition but there was little relation of the other model estimates with the percentage of SWD students in the school.

Conclusion

This report described the Pennsylvania results of a large study examining four alternative methods of estimating school performance across four states. In addition to this Pennsylvania report, there are reports describing results for the three other states (AZ, OR, NC) included in the study. The four alternative school performance models were representative of types of models often used in state accountability systems, although none were the actual model used in Pennsylvania at the time. We represented school performance in two ways, the actual model estimates and school ranks based on model estimates. Our primary interest in these comparisons

was estimating the impact of cohort and student composition (including the percent of SWD) on school performance estimates, as well as examining the extent to which different estimates of school performance correlated with each other.

A number of general conclusions can be drawn from the results of the Pennsylvania analyses. First, model representations of school performance over successive cohorts of students were somewhat unstable, irrespective of whether representations were based on school performance model estimates or on school ranks. There was somewhat greater cohort stability for status models (PP) than for the multiyear models. Nonetheless, even with the most stable PP model, Spearman's correlations showed that less than two-thirds of the variance was common across cohorts, and over all the models, there was substantial instability over cohorts. These results were also reflected in the examination of differences in absolute or average (RMSD) differences in ranks over cohorts.

Our examination of the relations of the school performance models with each other produced similar results. Generally, the status model estimates (PP) that were based on a single year of data did not agree with the remaining multiyear models. However, there was some substantial agreement of the SGP and VAM multiyear models with somewhat lower agreement of those models with the TM model.

We also examined the relation of school performance model estimates with variables describing the student composition of the schools. These results showed a pattern of results that differed between the status and the multiyear models. The status model had substantially larger negative correlations with school composition variables than the multiyear models. This was also true in terms of the percentage of SWD students served by a school. The greater the percentage of SWD in the school, the lesser the status model estimates of school performance.

Thus, the Pennsylvania results showed consistent patterns of instability of estimates of school performance over successive cohorts of students, estimates of school performance arising from the alternative school performance models – especially for status versus multiyear models – and stronger relations of the status model with the student composition of the school than multiyear models. Taken together, these results suggest the need for substantial caution in the way that school performance models are used and interpreted. Cohort instability suggests that rolling averages or some other mechanism is needed to provide more dependable depictions of school performance that are more stable over time. The substantial disagreement among the school performance models suggests that the choice of model matters a great deal. This choice should be made very carefully. A single model estimate of school performance may not be trustworthy and may need to be augmented by the results from additional models or metrics of school performance.

References

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics,* 29(1), 37-65.

Betebenner, D. (2009). Norm-and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42-51.

Betebenner, D. W., & Iwaarden, A. V. (2011). *SGP: An textupR package for the calculation and visualization of student growth percentiles* [Computer software manual]. (R package version 0.4-0.0 available at http://cran.r-project.org/web/packages/SGP/)

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289‑328.

Castellano, K. E., & Ho, A. D. (2013). Contrasting OLS and quantile regression approaches to student "growth" percentiles. *Journal of Educational and Behavioral Statistics, 38*, 190–215.

Goldschmidt, P., Choi, K., & Beaudoin, J. P. (2012, February). *Growth model comparison study: Practical implications of alternative models for evaluating school performance.* Council of Chief State School Officers: Washington, DC.

Pennsylvania Department of Education (2008). *Technical report for the Pennsylvania System of School Assessment 2008 Reading and Mathematics grades 3, 4, 5, 6, 7, 8, and 11.* Harrisburg, PA: PDE. Retrieved from http://www.pde.state.pa.

Pennsylvania Department of Education (2009). *Technical report for the 2009 Pennsylvania System of School Assessment.* Harrisburg, PA: PDE. Retrieved from http://www.pde.state.pa.

Pennsylvania Department of Education (2010). *Technical report for the 2010 Pennsylvania System of School Assessment.* Harrisburg, PA: PDE. from http://www.pde.state.pa.

Pennsylvania Department of Education (2011). *Technical report for the 2011 Pennsylvania System of School Assessment.* Harrisburg, PA: PDE. Retrieved from http://www.pde.state.pa.

Pennsylvania Department of Education (2012). *Technical Report for the 2012 Pennsylvania System of School Assessment.* Harrisburg, PA: PDE. Retrieved October 27, 2017 from http://www.pde.state.pa

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *MLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Tindal, G., Nese, J. F. T., and Stevens, J. J. (2017). Estimating school effects with a state testing program using transition matrices. *Educational Assessment, 22,* 189-204.

Appendix A

Correlations among School Performance Model Estimates for Each Individual Cohort by Content Area and Grade Level Band

Mathematics Elementary Schools

Cohort 1

Model	TM	SGP	VAM
PP	0.398	0.498	0.543
TM		0.86	0.863
SGP			0.963

Cohort 2

Model	TM	SGP	VAM
PP	0.322	0.475	0.500
TM		0.853	0.852
SGP			0.959

Cohort 3

Model	TM	SGP	VAM
PP	0.602	0.643	0.675
TM		0.894	0.909
SGP			0.97

Mathematics Middle Schools

Cohort 1

Model	TM	SGP	VAM
PP	0.370	0.492	0.548
TM		0.846	0.822
SGP			0.972

Cohort 2

Model	TM	SGP	VAM
PP	0.191	0.427	0.472
TM		0.806	0.798
SGP			0.961

Cohort 3

Model	TM	SGP	VAM
PP	0.417	0.548	0.63
TM		0.839	0.839
SGP			0.964

English Language Arts Elementary Schools

Cohort 1

Model	TM	SGP	VAM
PP	0.077	0.493	0.555
TM		0.729	0.713
SGP			0.941

Cohort 2

Model	TM	SGP	VAM
PP	0.053	0.568	0.653
TM		0.684	0.643
SGP			0.938

Cohort 3

Model	TM	SGP	VAM
PP	0.441	0.684	0.766
TM		0.794	0.779
SGP			0.951

English Language Arts Middle Schools

Cohort 1

Model	TM	SGP	VAM
PP	0.089	0.451	0.528
TM		0.684	0.669
SGP			0.945

Cohort 2

Model	TM	SGP	VAM
PP	0.390	0.597	0.674
TM		0.672	0.632
SGP			0.925

Cohort 3

Model	TM	SGP	VAM
PP	0.338	0.478	0.578
TM		0.709	0.704
SGP			0.931

Appendix B

Correlations of School Performance Model Estimates with School Composition Variables for Each Individual Cohort by Content Area and Grade Level Band

Mathematics Elementary Schools

Cohort 1

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.672	-0.186	-0.103	-0.003	-0.610	0.200
TM	-0.062	0.020	0.038	0.000	-0.007	0.023
SGP	-0.144	0.038	0.021	0.012	-0.050	0.042
VAM	-0.185	0.032	0.026	0.005	-0.088	0.054

Cohort 2

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.670	-0.171	-0.137	-0.069	-0.590	0.184
TM	-0.058	0.005	0.060	0.004	0.048	0.015
SGP	-0.160	0.024	0.033	-0.045	-0.051	0.048
VAM	-0.191	0.014	0.027	-0.046	-0.066	0.049

Cohort 3

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.763	-0.195	-0.209	-0.027	-0.698	0.162
TM	-0.427	-0.103	-0.061	-0.009	-0.351	0.073
SGP	-0.456	-0.102	-0.073	-0.033	-0.369	0.066
VAM	-0.508	-0.121	-0.081	-0.030	-0.404	0.077

Mathematics Middle Schools

Cohort 1

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.681	-0.238	-0.102	-0.127	-0.646	0.418
TM	-0.042	0.012	-0.043	-0.076	-0.070	0.022
SGP	-0.143	-0.028	-0.024	-0.097	-0.158	0.085

VAM	-0.195	-0.048	-0.031	-0.081	-0.196	0.110

Cohort 2

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.715	-0.271	-0.370	0.012	-0.684	0.401
TM	-0.013	0.059	0.025	-0.014	-0.026	0.001
SGP	-0.185	0.005	-0.051	-0.026	-0.166	0.088
VAM	-0.226	0.012	-0.061	-0.032	-0.206	0.120

Cohort 3

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.723	-0.215	-0.386	0.060	-0.693	0.390
TM	-0.206	-0.037	-0.064	0.053	-0.219	0.114
SGP	-0.303	-0.072	-0.108	0.078	-0.277	0.143
VAM	-0.383	-0.103	-0.136	0.076	-0.354	0.190

English Language Arts Elementary Schools

Cohort 1

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.710	-0.109	-0.131	0.039	-0.634	0.244
TM	0.242	0.093	0.039	-0.018	0.195	-0.087
SGP	-0.095	0.087	-0.025	0.028	-0.067	0.050
VAM	-0.143	0.079	-0.037	0.022	-0.117	0.068

Cohort 2

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.731	-0.069	-0.140	-0.027	-0.580	0.274
TM	0.185	-0.005	0.074	-0.034	0.245	-0.048
SGP	-0.246	0.006	-0.010	-0.022	-0.095	0.095
VAM	-0.332	0.001	-0.024	-0.025	-0.176	0.129

Cohort 3

	EDS	EL	SWD	Female	Minority	School Size

Model						
PP	-0.835	-0.230	-0.217	-0.005	-0.728	0.202
TM	-0.354	-0.118	-0.005	-0.008	-0.308	0.029
SGP	-0.551	-0.134	-0.060	-0.017	-0.445	0.108
VAM	-0.630	-0.158	-0.075	-0.012	-0.523	0.111

English Language Arts Middle Schools

Cohort 1

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.680	-0.143	-0.168	0.033	-0.592	0.418
TM	0.276	0.086	-0.088	0.032	0.252	-0.157
SGP	-0.047	0.068	-0.014	0.010	0.029	0.027
VAM	-0.127	0.058	-0.036	0.025	-0.044	0.095

Cohort 2

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.737	-0.199	-0.417	0.057	-0.654	0.430
TM	-0.132	-0.048	-0.031	0.001	-0.162	0.058
SGP	-0.350	-0.048	-0.140	0.027	-0.259	0.236
VAM	-0.475	-0.083	-0.188	0.034	-0.367	0.304

Cohort 3

Model	EDS	EL	SWD	Female	Minority	School Size
PP	-0.753	-0.257	-0.450	0.084	-0.662	0.403
TM	-0.101	-0.018	-0.066	0.011	-0.096	0.070
SGP	-0.272	-0.058	-0.119	0.074	-0.136	0.165
VAM	-0.392	-0.103	-0.165	0.049	-0.246	0.245

Appendix C

Correlations of School Performance Model Estimates with School Percentage SWD for Each Individual Cohort by Content Area and Grade Level Band

Mathematics Elementary Schools

Cohort	PP	TM	SGP	VAM
1	-0.103	0.038	0.021	0.026
2	-0.137	0.060	0.033	0.027
3	-0.209	-0.061	-0.073	-0.081

Mathematics Middle Schools

Cohort	PP	TM	SGP	VAM
1	-0.102	-0.043	-0.024	-0.031
2	-0.370	0.025	-0.051	-0.061
3	-0.386	-0.064	-0.108	-0.136

English Language Arts Elementary Schools

Cohort	PP	TM	SGP	VAM
1	-0.131	0.039	-0.025	-0.037
2	-0.140	0.074	-0.010	-0.024
3	-0.217	-0.005	-0.060	-0.075

English Language Arts Middle Schools

Cohort	PP	TM	SGP	VAM
1	-0.168	-0.088	-0.014	-0.036
2	-0.417	-0.031	-0.140	-0.188
3	-0.450	-0.066	-0.119	-0.165

Appendix D

Proportion of Elementary or Middle Schools within 5, 10, or 20 Ranks of Each Other for Each Pair of School Performance Models in Mathematics and English Language Arts by Cohort

Model Comparison	Cohort 1			Cohort 2			Cohort 3		
	r = 5	r = 10	r = 20	r = 5	r = 10	r = 20	r = 5	r = 10	r = 20
PP vs. TM									
Math Elementary	0.155	0.289	0.503	0.125	0.265	0.479	0.187	0.359	0.637
English Language Arts Elementary	0.103	0.188	0.376	0.127	0.225	0.389	0.174	0.304	0.531
Math Middle	0.140	0.267	0.486	0.132	0.221	0.430	0.196	0.333	0.530
English Language Arts Middle	0.123	0.227	0.373	0.167	0.281	0.469	0.141	0.278	0.480
Mean	0.130	0.243	0.434	0.138	0.248	0.442	0.174	0.318	0.544
PP vs. SGP									
Math Elementary	0.178	0.305	0.543	0.165	0.305	0.548	0.208	0.383	0.637
English Language Arts Elementary	0.162	0.317	0.532	0.207	0.351	0.600	0.229	0.410	0.673
Math Middle	0.174	0.316	0.561	0.168	0.281	0.489	0.204	0.347	0.582
English Language Arts Middle	0.174	0.286	0.501	0.211	0.380	0.617	0.207	0.341	0.566
Mean	0.172	0.306	0.534	0.188	0.329	0.564	0.212	0.370	0.614
PP vs. VAM									
Math Elementary	0.200	0.344	0.572	0.194	0.340	0.582	0.232	0.418	0.683

English Language Arts Elementary	0.194	0.334	0.572	0.249	0.424	0.673	0.282	0.472	0.723
Math Middle	0.225	0.367	0.568	0.181	0.293	0.533	0.221	0.379	0.644
English Language Arts Middle	0.179	0.316	0.533	0.255	0.434	0.666	0.206	0.373	0.624
Mean	0.200	0.340	0.561	0.220	0.373	0.614	0.235	0.410	0.668

TM vs. SGP

Math Elementary	0.369	0.569	0.845	0.404	0.579	0.807	0.433	0.654	0.888
English Language Arts Elementary	0.251	0.434	0.679	0.250	0.417	0.651	0.302	0.480	0.745
Math Middle	0.372	0.558	0.804	0.349	0.551	0.798	0.400	0.591	0.789
English Language Arts Middle	0.297	0.453	0.677	0.239	0.397	0.601	0.232	0.432	0.666
Mean	0.322	0.504	0.751	0.310	0.486	0.714	0.342	0.539	0.772

TM vs. VAM

Math Elementary	0.364	0.572	0.853	0.396	0.571	0.807	0.421	0.659	0.876
English Language Arts Elementary	0.250	0.430	0.683	0.236	0.412	0.636	0.305	0.489	0.742
Math Middle	0.384	0.602	0.819	0.319	0.558	0.786	0.414	0.607	0.788
English Language Arts Middle	0.278	0.434	0.656	0.230	0.399	0.599	0.239	0.445	0.675
Mean	0.319	0.509	0.753	0.295	0.485	0.707	0.345	0.550	0.770

SGP vs. VAM

Math Elementary	0.615	0.847	0.983	0.598	0.833	0.972	0.665	0.894	0.990
English Language Arts Elementary	0.488	0.742	0.940	0.504	0.745	0.959	0.542	0.786	0.966
Math Middle	0.681	0.872	0.993	0.623	0.858	0.977	0.658	0.879	0.984

English Language Arts Middle	0.510	0.764	0.965	0.487	0.757	0.944	0.504	0.773	0.951
Mean	0.574	0.806	0.970	0.553	0.798	0.963	0.592	0.833	0.973

Appendix E

RMSD in School Ranks for Pairs of School Performance Models for Each Individual Cohort by Content Area and Grade Level Band

Elementary School Mathematics: Cohort 1

Model	TM	SGP	VAM
PP	31.204	28.672	27.007
TM		15.522	15.089
SGP			7.716

Elementary School Mathematics: Cohort 2

Model	TM	SGP	VAM
PP	32.242	29.277	27.748
TM		16.263	15.678
SGP			8.108

Elementary School Mathematics: Cohort 3

Model	TM	SGP	VAM
PP	24.405	24.017	21.894
TM		13.227	12.534
SGP			6.694

Elementary School English Language Arts: Cohort 1

Model	TM	SGP	VAM
PP	39.187	29.163	27.421
TM		21.926	22.778
SGP			10.069

Elementary School English Language Arts: Cohort 2

Model	TM	SGP	VAM
PP	38.787	26.14	23.068
TM		24.021	25.299
SGP			9.977

Elementary School English Language Arts: Cohort 3

Model	TM	SGP	VAM
PP	30.159	23.019	19.648
TM		18.906	19.352
SGP			9.044

Middle School Mathematics: Cohort 1

Model	TM	SGP	VAM
PP	32.05	29.482	27.815
TM		15.928	16.689
SGP			6.732

Middle School Mathematics: Cohort 2

Model	TM	SGP	VAM
PP	35.659	31.195	29.759
TM		18.16	18.571
SGP			7.816

Middle School Mathematics: Cohort 3

Model	TM	SGP	VAM
PP	30.318	27.447	24.542
TM		16.876	17.155
SGP			7.149

Middle School English Language Arts: Cohort 1

Model	TM	SGP	VAM
PP	41.326	31.471	29.357
TM		24.711	25.287
SGP			9.510

Middle School English Language Arts: Cohort 2

Model	TM	SGP	VAM
PP	34.173	25.507	22.276
TM		25.653	25.885
SGP			10.045

Middle School English Language Arts: Cohort 3

Model	TM	SGP	VAM
PP	34.168	28.604	25.372
TM		22.697	23.118
SGP			9.687

Appendix F

Correlations of School Ranks with School Composition Variables by Content Area and Grade Level Band for Each Individual Cohort

Elementary School Mathematics: Cohort 1

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.662	-0.081	-0.121	-0.020	-0.422	0.241	-0.178
TM	-0.063	0.047	0.022	0.010	0.001	0.035	0.009
SGP	-0.158	0.078	0.013	0.002	-0.020	0.083	0.000
VAM	-0.208	0.079	0.016	-0.004	-0.052	0.095	-0.012

Elementary School Mathematics: Cohort 2

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.662	-0.069	-0.136	-0.068	-0.415	0.212	-0.190
TM	-0.085	0.007	0.043	-0.014	0.036	0.060	0.008
SGP	-0.172	0.039	0.024	-0.052	-0.038	0.089	-0.018
VAM	-0.217	0.030	0.010	-0.055	-0.055	0.102	-0.031

Elementary School Mathematics: Cohort 3

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.743	-0.076	-0.189	-0.020	-0.500	0.218	-0.218
TM	-0.433	-0.016	-0.053	-0.001	-0.276	0.139	-0.107
SGP	-0.450	-0.016	-0.057	-0.018	-0.288	0.131	-0.116
VAM	-0.521	-0.011	-0.074	-0.019	-0.316	0.159	-0.130

Elementary School English Language Arts: Cohort 1

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.698	-0.020	-0.142	0.015	-0.414	0.287	-0.162
TM	0.239	0.036	0.034	-0.012	0.150	-0.094	0.059
SGP	-0.111	0.067	-0.035	0.020	-0.023	0.068	-0.002
VAM	-0.161	0.067	-0.053	0.022	-0.066	0.087	-0.017

Elementary School English Language Arts: Cohort 2

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.730	0.000	-0.113	-0.031	-0.370	0.320	-0.154
TM	0.160	-0.005	0.081	-0.040	0.176	-0.038	0.056
SGP	-0.279	0.022	0.000	-0.036	-0.058	0.137	-0.036
VAM	-0.370	0.018	-0.013	-0.034	-0.113	0.187	-0.054

Elementary School English Language Arts: Cohort 3

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.822	-0.066	-0.190	-0.008	-0.501	0.264	-0.220
TM	-0.335	-0.047	0.004	-0.012	-0.250	0.064	-0.096
SGP	-0.538	-0.028	-0.044	-0.027	-0.340	0.170	-0.134
VAM	-0.622	-0.038	-0.070	-0.027	-0.392	0.190	-0.160

Middle School Mathematics: Cohort 1

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.695	-0.055	-0.155	-0.094	-0.555	0.462	-0.182
TM	-0.036	0.013	-0.003	-0.042	-0.079	0.029	-0.020
SGP	-0.134	0.008	-0.016	-0.075	-0.167	0.114	-0.045

VAM	-0.185	0.009	-0.026	-0.072	-0.194	0.144	-0.054

Middle School Mathematics: Cohort 2

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.724	-0.102	-0.377	0.005	-0.598	0.440	-0.226
TM	-0.030	0.106	-0.015	-0.012	-0.032	0.000	0.003
SGP	-0.191	0.075	-0.046	-0.046	-0.154	0.090	-0.045
VAM	-0.231	0.079	-0.063	-0.045	-0.186	0.125	-0.054

Middle School Mathematics: Cohort 3

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.731	-0.164	-0.369	0.018	-0.612	0.423	-0.239
TM	-0.194	-0.010	-0.043	0.056	-0.200	0.136	-0.042
SGP	-0.295	-0.075	-0.072	0.082	-0.257	0.178	-0.073
VAM	-0.377	-0.082	-0.105	0.077	-0.328	0.239	-0.096

Middle School English Language Arts: Cohort 1

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.708	-0.011	-0.236	0.043	-0.465	0.458	-0.153
TM	0.304	0.065	-0.060	0.020	0.212	-0.147	0.066
SGP	-0.062	0.086	-0.031	0.008	0.030	0.027	0.010
VAM	-0.134	0.087	-0.062	0.023	-0.014	0.091	-0.002

Middle School English Language Arts: Cohort 2

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean

PP	-0.754	-0.087	-0.404	0.017	-0.529	0.468	-0.215
TM	-0.129	-0.037	-0.006	-0.023	-0.163	0.088	-0.045
SGP	-0.382	-0.005	-0.109	0.022	-0.213	0.239	-0.075
VAM	-0.483	-0.009	-0.166	0.033	-0.293	0.295	-0.104

Middle School English Language Arts: Cohort 3

Model	EDS	EL	SWD	Female	Ethnic Minority	School Size	Mean
PP	-0.773	-0.112	-0.439	0.029	-0.528	0.442	-0.230
TM	-0.082	0.050	-0.064	0.028	-0.043	0.065	-0.008
SGP	-0.288	0.016	-0.121	0.066	-0.078	0.150	-0.042
VAM	-0.403	0.006	-0.155	0.049	-0.162	0.230	-0.073

Appendix G

Correlations of School Ranks with School Percentage SWD for Each Individual Cohort by Content Area and Grade Level Band

Elementary School Mathematics

Cohort	PP	TM	SGP	VAM
1	-0.121	0.022	0.013	0.016
2	-0.136	0.043	0.024	0.010
3	-0.189	-0.053	-0.057	-0.074

Elementary School English Language Arts

Cohort	PP	TM	SGP	VAM
1	-0.142	0.034	-0.035	-0.053
2	-0.113	0.081	0.000	-0.013
3	-0.190	0.004	-0.044	-0.070

Middle School Mathematics

Cohort	PP	TM	SGP	VAM
1	-0.155	-0.003	-0.016	-0.026
2	-0.377	-0.015	-0.046	-0.063
3	-0.369	-0.043	-0.072	-0.105

Middle School English Language Arts

Cohort	PP	TM	SGP	VAM
1	-0.236	-0.060	-0.031	-0.062
2	-0.404	-0.006	-0.109	-0.166
3	-0.439	-0.064	-0.121	-0.155