A Comparison of Alternative Models for Estimating School Performance in Mathematics and Reading/Language Arts in a State Accountability System, Part A: School Performance Estimates

NCAASE Technical Report

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

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

Address all correspondence to Joseph J. Stevens, University of Oregon, Department of Educational Methodology, Policy, and Leadership, College of Education, 5267 University of Oregon, Eugene, OR 97403-5267; 541-346-2445, stevensj@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.

A Comparison of Alternative Models for Estimating School Performance in Mathematics and Reading/Language Arts in a State Accountability System, Part A: School Performance Estimates

The purpose of this technical report is to describe the results of the first part (A) of a large study examining alternative models for estimating school performance. Despite the central importance of analytic models used in evaluating teacher and school effects in modern accountability systems, there are relatively few studies that examine the reliability and validity of these high-stakes systems (see, for example, Goldschmidt, Choi, & Beaudoin, 2012). The results reported here examine several alternative models using Oregon operational state accountability data in mathematics and reading. We sought to address four questions surrounding the use of analytic models for the evaluation of school performance:

(a) Are estimates of school performance stable across successive cohorts of students?

(b) How well do estimates of school performance correlate among models?

(c) How do estimates of school performance correlate with variables describing the student composition of the school?

(d) Do estimates of school performance vary from one model to another based on the school composition of SWD?

Because of the volume of results being reported, this technical report (Part A) describes correlations among study cohorts, models, and variables describing the student composition of schools. A second portion of the technical report (Part B) reports additional results based on school rankings derived from the models reported here.

Method

Sample

The initial sample consisted of all Oregon students that took the Oregon Assessment of Knowledge and Skills (OAKS) mathematics (N = 483,502) or reading/language arts (N = 474,080) general assessment in any one school year from 2007-08 through 2011-12, and whose records in each year were included in the state calculation of Adequate Yearly Progress (AYP). 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 were retained or accelerated from the previous year.

The initial 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 cohorts (a) 2007/08 through 2009/2010; (b) 2008/09 through 2010/11; and (c) 2009/10 through 2011/12; see Table 1). The initial elementary school sample for the mathematics test was 137,744 students. The initial middle school sample for the mathematics test was 136,535 students. The initial elementary school sample for the reading test was 137,535 students. The initial middle school sample for the reading test was 137,343 students. To create an analytic sample that was appropriate for our research questions, we systematically retained specific students and schools from the initial sample. We included students with valid test scores in all three years, schools that served all three grades (Grades 3 through 5 or 6 through 8) for a cohort, and schools with $N \geq 10$

students in each of the three cohorts in the final reference year (i.e., Grade 5 for elementary school and Grade 8 for middle school). As is the case in most operational and research applications of these models, there was 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. These inclusion rules were applied to ensure that there were no differences in the analytic sample for different school models and comparisons of school models were therefore 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 90,679 students. The final middle school analytic sample for the mathematics test was 75,318 students. The final elementary school analytic sample for the reading test was 89,627 students. The final middle school analytic sample for the reading test was 75,193 students.

Table 1
Research Design Indicating Academic Years and Longitudinal Cohorts Studied.

	Academic Year				
Grade	2007-08	2008-09	2009-10	2010-11	2011-12
3	E1	E2	E3		
4		E1	E2	E3	
5			E1	E2	E3
6	M1	M2	M3		
7		M1	M2	M3	
8			M1	M2	M3

Note. E denotes an elementary school cohort, M denotes a middle school cohort.

Table 2 provides summary statistics describing the school-level analytical samples of Oregon fifth and eighth grade students in the three cohorts for mathematics and reading. Although there is variation from cohort to cohort in sample demographic characteristics, generally the composition of the samples is quite similar across the three cohorts and for mathematics and reading/language arts at each grade level band. From elementary to middle school cohorts, there are small but consistent decreases in the proportion of English learners (EL), free or reduced lunch (FRL), 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, almost 10% of the students were EL, almost 50% of the students were female, over 50% were FRL, approximately 33% were racial/ethnic minority students, and about 13% were SWD. It is also noteworthy that there is much greater school level variation–as indicated by the values of the standard deviations in parentheses–in FRL and racial/ethnic minority student school composition than other student characteristics.

Table 2.

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

	Cohort		
Math Elementary	1	2	3
ELL	.102 (.136)	.097 (.129)	.094 (.120)
Female	.495 (.093)	.496 (.086)	.496 (.078)
FRL	.514 (.262)	.523 (.251)	.543 (.247)
Minority	.323 (.217)	.332 (.218)	.343 (.216)
SWD	.127 (.065)	.131 (.068)	.127 (.063)
Math Middle			
ELL	.056 (.082)	.047 (.066)	.033 (.050)
Female	.495 (.087)	.500 (.092)	.494 (.090)
FRL	.482 (.238)	.491 (.228)	.510 (.227)
Minority	.307 (.229)	.320 (.220)	.330 (.218)
SWD	.116 (.065)	.111 (.060)	.114 (.066)
Reading Elementary			
ELL	.093 (.121)	.097 (.129)	.093 (.119)
Female	.497 (.093)	.497 (.087)	.498 (.078)
FRL	.511 (.262)	.523 (.251)	.543 (.247)
Minority	.316 (.211)	.333 (.218)	.344 (.216)
SWD	.123 (.064)	.127 (.067)	.123 (.062)
Reading Middle			
ELL	.056 (.082)	.047 (.066)	.033 (.050)
Female	.495 (.087)	.500 (.092)	.494 (.090)
FRL	.482 (.238)	.491 (.228)	.510 (.227)
Minority	.307 (.229)	.320 (.220)	.330 (.218)
SWD	.116 (.065)	.111 (.060)	.114 (.066)

Instrument

The outcome measures for all analyses were the standardized Oregon Assessment of Knowledge and Skills (OAKS; Oregon Department of Education [ODE], 2012a) mathematics and reading/language arts tests. The OAKS is a summative, computer-adaptive assessment based on the Oregon content standards (ODE, 2008). OAKS test specifications varied by grade and content area and were intended to measure the core content standards in the state curriculum (ODE, 2012a). The tests were administered under standardized conditions (ODE, 2012b). OAKS raw scores were converted to scale scores based on the number of items answered correctly while taking item difficulty into account using one parameter item response theory (IRT) methods and a vertical linking design over grades to create a student developmental scale score (ODE, 2009).

School Performance Models

For all models, we estimated school performance in the focal year (Grade 5 or 8), using prior years achievement scores as dictated by the particular model. We applied seven alternative analytic models of school performance to the Oregon mathematics and reading achievement data in elementary and middle school. The seven models were: Percent Proficient (PP), gain score (Gain), transition matrix (TM), student growth percentile (SGP), value-added model (VAM), and Multilevel Linear Model initial status (MLM0) and growth rate (MLM Growth).

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 (see ODE, 2008).

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 five OR state proficiency categories in the prior year crossed with the five state proficiency categories in the focal year (Grade 5 or Grade 8) which 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, 2016). The weighted values were averaged across all cells to create an overall school TM index.

Student Growth Percentile (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 was discussed extensively by Ballou, Sanders, and Wright (2004) and is known generally as the "layered model" because layers of equations are added with each year of schooling. For example, the model for a student 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 represents an assessment for student *i* in grade *g*, and the student attended school *j*. The fixed mean for all students in the combination of grades and schools was μ_{ijg}, while e_{ijg} was the random deviation for student *n* from the mean, μ_{ijg}. The layered model we used was limited to a maximum of three years and was applied separately to the two content areas.

Multilevel Linear Growth Model Initial Status and Growth (MLM0 and MLM Growth Rate). We modeled student growth over the three elementary or three middle school grades with multilevel longitudinal analyses (Raudenbush & Bryk, 2002) using MLM 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 first grade (3 or 6). We used a quadratic model based on previous findings in the literature 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) in order to obtain a model solution. The resulting 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}). We used two parameters from the MLM models as estimates of school performance, the empirical Bayes (EB) estimates of school intercepts (labeled "MLM0" below) and the EB estimates of school average growth rate (labeled "MLM" below). Growth rate was calculated as: $\pi_{1ij} + (2)(\pi_{2ij})(t)$.

Comparison of Models

In order to evaluate the comparability and stability of school estimates across school performance models and across cohorts we used several comparison criteria. In Part A of this technical report we describe comparisons based on the correlation of the school estimates just

described. In Part B of the report we describe comparisons based on school ranks arising from these school performance estimates. In the present evaluation of school performance estimates, we examined: (a) the correlations of each model across the three cohorts, (b) the correlations among school estimates from one model to another, (c) the correlations among the school estimates and school composition variables, and (d) the correlations of each model with the percentage of students with disabilities in the school.

Results and Discussion

Cohort Stability. We first consider the stability of model estimates across the three successive cohorts of students. We computed the correlations among cohorts of the school estimates from each school performance model to determine the stability of estimates across successive groups 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 3 shows the correlation of model estimates across cohorts for mathematics and reading/language arts in the elementary school and middle school samples. As can be seen in Table 3, correlations generally range from small to moderate indicating some substantial instability in school performance estimates across cohorts. Correlations between adjacent years in the first two columns (cohort 1 to 2 or 2 to 3) are generally somewhat higher than the comparisons across two years (cohort 1 to 3). Although there is also some variation from elementary to middle school or from

Table 3. *Correlations of School Performance Model Estimates across Cohorts*.

Elementary School Mathematics

	Cohort		
Model	**1 to 2**	**2 to 3**	**1 to 3**
PP	0.593	0.630	0.529
MLM0	0.796	0.808	0.745
Gain	0.369	0.305	0.174
TM	0.319	0.267	0.139
SGP	0.416	0.432	0.237
VAM	0.443	0.468	0.275
MLM	0.325	0.241	0.145

Elementary School Reading/Language Arts

	Cohort		
Model	**1 to 2**	**2 to 3**	**1 to 3**
PP	0.655	0.667	0.593
MLM0	0.818	0.840	0.796
Gain	0.281	0.261	0.084
TM	0.208	0.093	0.115
SGP	0.364	0.279	0.204
VAM	0.396	0.312	0.215
MLM	0.211	0.123	0.113

Middle School Mathematics

	Cohort		
Model	**1 to 2**	**2 to 3**	**1 to 3**
PP	0.713	0.670	0.586
MLM0	0.815	0.824	0.790
Gain	0.300	0.299	0.212
TM	0.321	0.259	0.205
SGP	0.399	0.429	0.221
VAM	0.414	0.491	0.269
MLM	0.233	0.228	0.211

Middle School Reading/Language Arts

	Cohort		
Model	**1 to 2**	**2 to 3**	**1 to 3**
PP	0.674	0.668	0.672
MLM0	0.846	0.839	0.826
Gain	0.294	0.255	0.246
TM	0.321	0.259	0.205
SGP	0.338	0.176	0.153
VAM	0.308	0.314	0.171

MLM	0.282	0.376	0.368

mathematics to reading, trends in cohort stability were fairly similar across content and grade level. To facilitate interpretation of the cohort results, we also averaged correlations across the two content areas and grade levels (see Table 4). It can be seen that the correlations across cohorts were largest for the two status based school performance measures (PP and MLM0) and noticeably lower for all other models that used two or three years of data to estimate school performance. The two rightmost columns of Table 4 show the overall means and standard deviations 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 MLM0 estimates (MLM EB intercepts) closely followed by the PP model estimates. All of the remaining multi-year performance models had much greater instability. The standard deviation of correlations across cohort comparisons are shown in the rightmost column of Table 4 and show the least variability over cohorts for the status models and the greatest variability across cohort correlations for the SGP model followed by the VAM model.

Table 4. *Average Correlations across Content and Grade Level and Overall Mean and Standard Deviation (SD) Across the Three Cohort Comparisons*.

	Cohort				
Model	**1 to 2**	**2 to 3**	**1 to 3**	**Mean**	**SD**
PP	0.659	0.659	0.595	0.638	0.040
MLM0	0.819	0.828	0.789	0.812	0.021
Gain	0.311	0.280	0.179	0.257	0.071
TM	0.292	0.220	0.166	0.226	0.067
SGP	0.379	0.329	0.204	0.304	0.100
VAM	0.390	0.396	0.233	0.340	0.097
MLM	0.263	0.242	0.209	0.238	0.052

Comparison of Models. We next computed the correlations of school performance models 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 A1 to A4. Table 5 shows model correlations for mathematics and reading/language arts in the elementary school and middle school samples.

Table 5. *Correlations of School Performance Estimates across Models*.

Elementary School Mathematics

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.617	0.261	0.293	0.495	0.516	0.169
MLM0		–	-0.103	-0.086	0.026	0.043	-0.048
Gain			–	0.684	0.849	0.872	0.960
TM				–	0.664	0.664	0.621
SGP					–	0.957	0.717
VAM						–	0.735
MLM							–

Elementary School Reading/Language Arts

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.739	-0.015	0.260	0.448	0.519	-0.013
MLM0		–	-0.321	-0.004	0.134	0.173	-0.215
Gain			–	0.423	0.481	0.499	0.442
TM				–	0.709	0.716	0.672
SGP					–	0.913	0.535
VAM						–	0.576
MLM							–

Middle School Mathematics

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.586	0.456	0.572	0.513	0.508	0.431
MLM0		–	0.093	0.134	0.021	0.031	0.222
Gain			–	0.887	0.88	0.909	0.951
TM				–	0.811	0.809	0.836
SGP					–	0.949	0.762
VAM						–	0.803
MLM							–

Middle School Reading/Language Arts

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.755	-0.139	0.279	0.395	0.399	-0.307
MLM0		–	-0.489	0.100	0.033	0.019	-0.542
Gain			–	0.260	0.661	0.675	0.916
TM				–	0.366	0.365	0.177
SGP					–	0.897	0.438
VAM						–	0.499
MLM							–

As can be seen in Table 5, substantial variability is present in the degree to which school performance estimates for one model are related to other models. On average, across content and grade level, the lowest correlations were between the PP and MLM growth models (.07) and the MLM0 model and the TM (.04), SGP (.05), and VAM (.07) models. The MLM0 model was also negatively correlated with the Gain score model (-.21) and the MLM growth model (-.15). The highest correlations were among the PP and MLM0 models (.67), and the Gain score model with the TM model (.56), SGP model (.72), the VAM model (.74), and the MLM growth model (.82). The average correlations among the TM model and the SGP model (.64), VAM model (.64), and MLM growth models (.58) were also high as was the correlation between the VAM and MLM growth models (.65). The highest average correlation among models was between the SGP and VAM models (.93). It is also worth noting, however, that the correlations among models varied a good deal depending on content and grade level. For example, the correlation between the PP model and the Gain score model ranged from -.14 to +.46 and between MLM0

and MLM growth rate ranged from -.54 to +.22. The least variation in model correlations across content and grade level was for the SGP and VAM models from +.90 to +.96.

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 model 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 B1-B4.

Table 6. *Correlations of Elementary School Mathematics Model Estimates with School Composition Variables*.

Elementary School Mathematics

	School Composition Variable						
Model	FRL	LEP	SWD	Female	Minority	School Size	Mean
PP	-0.500	-0.329	-0.194	0.0123	-0.319	0.147	-0.197
MLM0	-0.658	-0.426	-0.188	-0.032	-0.377	0.126	-0.259
Gain	-0.015	0.052	0.017	-0.003	0.021	0.099	0.028
TM	-0.058	0.032	-0.016	-0.006	-0.011	0.097	0.006
SGP	-0.158	-0.016	-0.036	-0.009	-0.028	0.171	-0.013
VAM	-0.180	-0.024	-0.052	0.006	-0.035	0.187	-0.016
MLM	-0.013	0.032	0.029	-0.007	-0.003	0.062	0.017

Elementary School Reading/Language Arts

	School Composition Variable						
Model	FRL	LEP	SWD	Female	Minority	School Size	Mean
PP	-0.600	-0.526	-0.204	0.002	-0.492	0.064	-0.293
MLM0	-0.719	-0.569	-0.186	0.002	-0.495	0.090	-0.313
Gain	0.158	0.184	0.076	-0.024	0.117	0.067	0.096
TM	-0.072	0.051	0.035	0.003	0.008	0.084	0.018
SGP	-0.210	-0.061	-0.041	-0.003	-0.105	0.110	-0.052
VAM	-0.258	-0.110	-0.059	0.008	-0.141	0.125	-0.072
MLM	0.142	0.130	0.054	0.023	0.090	0.005	0.074

Middle School Mathematics

Model	School Composition Variable						
	FRL	LEP	SWD	Female	Minority	School Size	Mean
PP	-0.330	-0.229	-0.267	0.051	-0.266	0.169	-0.145
MLM0	-0.573	-0.320	-0.183	-0.007	-0.264	0.156	-0.199
Gain	-0.009	0.017	-0.070	0.030	0.056	0.140	0.027
TM	-0.047	-0.001	-0.118	0.037	0.005	0.166	0.007
SGP	0.005	0.045	-0.049	0.026	0.059	0.172	0.043
VAM	0.006	0.045	-0.048	0.020	0.062	0.181	0.044
MLM	-0.090	-0.045	-0.074	0.011	-0.001	0.138	-0.010

Middle School Reading/Language Arts

Model	School Composition Variable						
	FRL	LEP	SWD	Female	Minority	School Size	Mean
PP	-0.525	-0.444	-0.263	0.135	-0.481	0.088	-0.248
MLM0	-0.704	-0.513	-0.217	0.112	-0.478	0.066	-0.289
Gain	0.326	0.238	0.092	-0.026	0.226	0.031	0.148
TM	-0.049	0.001	-0.112	0.047	0.006	0.167	0.010
SGP	-0.063	0.016	-0.035	0.056	-0.047	0.128	0.009
VAM	-0.058	0.002	-0.038	0.040	-0.058	0.121	0.002
MLM	0.365	0.255	0.126	-0.054	0.235	0.007	0.156

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 and MLM0, are noticeably higher than the correlations of the other school performance models with school composition variables. On average across content and grade level, the correlation of the school composition variables was -0.221 for the PP model and -0.265 for the MLM0 model. In contrast, the average correlations of the school composition variables with the remaining models were quite low ranging from -0.003 to +0.075.

Relation of Model Estimates to SWD School Composition. Because of the NCAASE focus on the performance and academic growth of SWD, we also specifically examined the relations between the percentage of SWD students served by a school and the school performance model estimates. Table 7 shows the correlation of model estimates with the percentage of SWD in each school for mathematics and reading/language arts in the elementary school and middle school

samples. Correlations of model estimates with SWD school composition within each individual cohort are presented in Appendix C1-C4. As can be seen in the bottom row of Table 7, on

Table 7. *Average School Performance Model Estimates as a Function of the Percentage of SWD in the School by Content and Grade Level*.

Content and Grade Level	School Performance Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM
Math. Elementary	-0.194	-0.188	0.017	-0.016	-0.036	-0.052	0.029
Math. Middle	-0.267	-0.183	-0.070	-0.118	-0.049	-0.048	-0.074
Reading Elementary	-0.204	-0.186	0.076	0.035	-0.041	-0.059	0.054
Reading Middle	-0.263	-0.217	0.092	-0.112	-0.035	-0.038	0.126
Mean	-0.221	-0.197	0.062	-0.031	-0.037	-0.050	0.070

average, school performance estimates based on the single-year, status models (PP and MLM0) had substantially higher correlations with school SWD composition than the other school performance models. With the PP and MLM0 models, school performance estimates were higher the lower the percentage of SWD students in the school.

Summary. We evaluated seven alternative models for estimating school academic performance in mathematics and reading/language arts using operational Oregon state accountability data. We observed substantial variability in model estimates across three successive student cohorts in mathematics and reading/language arts in both elementary and middle school grades. There was somewhat less variability across cohorts for status models (PP and MLM0) than for the models that used more than one year of data. We also compared the estimates of school performance from one model to another and found substantial disagreement across models. Generally, the status models based on a single year of data were more similar to each other and differed from the remaining models that examined more than one year of data. Comparison of model estimates to school composition variables showed that status models (PP and MLM0) had substantially higher correlations than the remaining school performance models. Finally, we correlated school performance estimates with the percentage of SWD in each school. Ideally one would hope that an estimate of school performance would not be related to the student composition of the school, but as with the other school composition variables, we found that the status models were more

highly correlated with SWD school composition but there was little relation of the other model estimates with the percentage of SWD students in the school.

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/)

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.

Oregon Department of Education (ODE). (2008). *Technical report: Oregon's statewide assessment system test development. Standard Setting, Volume 3.* Salem, OR.

Oregon Department of Education. (2009, January). *2007-2008 technical report: Score interpretation guide.* Retrieved from ODE website: http://www.ode.state.or.us/teachlearn/testing/manuals/2008/asmttechmanualvol6_interpguide.pdf

Oregon Department of Education (ODE). (2012a). *Technical report: Oregon's statewide assessment system test development, Volumes 1-10.* Salem, OR.

Oregon Department of Education (ODE). (2012b). *Technical report: Oregon's statewide assessment system test development. Test Administration, Volume 5.* Salem, OR.

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. (2016). *Documenting Student Growth and Estimating School Effects with a State Testing Program using Transition Matrices.* Manuscript submitted for publication.

Appendices

Appendix A1. *Mathematics Elementary School Model Correlations by Cohort*.

Appendix A2. *Mathematics Middle School Model Correlations by Cohort*.

Appendix A3. *Reading/Language Arts Elementary School Model Correlations by Cohort*.

Appendix A4. *Reading/Language Arts Middle School Model Correlations by Cohort*.

Appendix B1. *Mathematics Elementary School Model Correlations with School Composition Variables by Cohort.*

Appendix B2. *Mathematics Middle School Model Correlations with School Composition Variables by Cohort.*

Appendix B3. *Reading/Language Arts Elementary School Model Correlations with School Composition Variables by Cohort*.

Appendix B4. *Reading/Language Arts Middle School Model Correlations with School Composition Variables by Cohort*.

Appendix C1. *Mathematics Elementary School Model Correlations with School Percentage SWD by Cohort*.

Appendix C2. *Mathematics Middle School Model Correlations with School Percentage SWD by Cohort*.

Appendix C3. *Reading/Language Arts Elementary School Model Correlations with School Percentage SWD by Cohort*.

Appendix C4. *Reading/Language Arts Middle School Model Correlations with School Percentage SWD by Cohort*.

Appendix A1. *Mathematics Elementary School Model Correlations by Cohort*.

Cohort 1

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.640	0.150	0.295	0.465	0.473	0.049
MLM0		–	-0.212	-0.091	0.023	0.040	-0.165
Gain			–	0.887	0.816	0.832	0.952
TM				–	0.792	0.796	0.827
SGP					–	0.956	0.659
VAM						–	0.665
MLM							–

Cohort 2

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.579	0.364	0.375	0.527	0.542	0.296
MLM0		–	-0.099	-0.126	-0.029	-0.017	-0.031
Gain			–	0.901	0.892	0.916	0.970
TM				–	0.826	0.826	0.865
SGP					–	0.961	0.800
VAM						–	0.822
MLM							–

Cohort 3

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.632	0.268	0.208	0.494	0.534	0.161
MLM0		–	0.002	-0.042	0.084	0.105	0.051
Gain			–	0.265	0.838	0.868	0.959
TM				–	0.374	0.369	0.172
SGP					–	0.955	0.692
VAM						–	0.719

MLM							–

Appendix A2. *Mathematics Middle School Model Correlations by Cohort.*

Cohort 1

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.666	0.360	0.530	0.435	0.421	0.306
MLM0		–	0.138	0.216	0.069	0.104	0.241
Gain			–	0.845	0.839	0.878	0.932
TM				–	0.719	0.705	0.781
SGP					–	0.931	0.671
VAM						–	0.723
MLM							–

Cohort 2

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.586	0.425	0.523	0.535	0.512	0.393
MLM0		–	-0.014	0.033	-0.012	-0.020	0.103
Gain			–	0.889	0.897	0.913	0.957
TM				–	0.844	0.844	0.84
SGP					–	0.953	0.783
VAM						–	0.818
MLM							–

Cohort 3

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.506	0.583	0.662	0.568	0.591	0.594
MLM0		–	0.154	0.154	0.005	0.010	0.321
Gain			–	0.927	0.904	0.936	0.965
TM				–	0.870	0.877	0.887

SGP					–	0.962	0.832
VAM						–	0.868
MLM							–

Appendix A3. *Reading/Language Arts Elementary School Model Correlations by Cohort*.

Cohort 1

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.718	-0.012	0.276	0.466	0.547	-0.152
MLM0		–	-0.361	0.026	0.162	0.205	-0.336
Gain			–	0.718	0.651	0.67	0.907
TM				–	0.722	0.721	0.593
SGP					–	0.917	0.409
VAM						–	0.435
MLM							–

Cohort 2

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.735	0.071	0.200	0.432	0.497	-0.008
MLM0		–	-0.252	-0.084	0.118	0.156	-0.213
Gain			–	0.769	0.727	0.766	0.931
TM				–	0.673	0.684	0.685
SGP					–	0.914	0.571
VAM						–	0.605
MLM							–

Cohort 3

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.763	-0.104	0.303	0.445	0.512	0.122

MLM0		–	-0.350	0.046	0.121	0.158	-0.095
Gain			–	-0.219	0.065	0.060	-0.512
TM				–	0.732	0.743	0.738
SGP					–	0.909	0.624
VAM						–	0.689
MLM							–

Appendix A4. *Reading/Language Arts Middle School Model Correlations by Cohort*.

Cohort 1

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.751	-0.197	0.258	0.390	0.347	-0.357
MLM0		–	-0.531	0.116	0.022	-0.052	-0.588
Gain			–	0.191	0.646	0.690	0.916
TM				–	0.367	0.326	0.111
SGP					–	0.897	0.417
VAM						–	0.509
MLM							–

Cohort 2

Model	School Effects Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.741	-0.106	0.251	0.401	0.392	-0.303
MLM0		–	-0.476	0.029	0.006	0.034	-0.500
Gain			–	0.317	0.647	0.649	0.891
TM				–	0.364	0.332	0.221
SGP					–	0.880	0.368
VAM						–	0.439
MLM							–

Cohort 3

	School Effects Model

Model							
	PP	MLM0	Gain	TM	SGP	VAM	MLM Growth
PP	–	0.774	-0.113	0.329	0.393	0.457	-0.262
MLM0		–	-0.459	0.156	0.071	0.076	-0.538
Gain			–	0.273	0.689	0.686	0.942
TM				–	0.368	0.436	0.200
SGP					–	0.913	0.528
VAM						–	0.549
MLM							–

Appendix B1. *Mathematics Elementary School Model Correlations with School Composition Variables by Cohort.*

Cohort 1

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.457	-0.333	-0.201	-0.017	-0.315	0.148
MLM0	-0.622	-0.442	-0.185	-0.051	-0.415	0.096
Gain	0.102	0.156	0.081	-0.022	0.099	0.164
TM	0	0.113	0.006	-0.015	0.043	0.174
SGP	-0.132	0.028	-0.033	-0.062	0.005	0.248
VAM	-0.156	0.020	-0.029	-0.033	-0.001	0.270
MLM	0.117	0.151	0.099	-0.009	0.076	0.110

Cohort 2

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.507	-0.319	-0.209	0.075	-0.291	0.181
MLM0	-0.681	-0.457	-0.229	-0.015	-0.387	0.134
Gain	-0.039	0.08	0.019	0.023	0.027	0.093
TM	-0.001	0.108	-0.012	0.046	0.042	0.093
SGP	-0.122	0.026	-0.014	0.031	-0.001	0.142
VAM	-0.146	0.029	-0.041	0.045	0	0.170
MLM	-0.052	0.051	0.019	0.017	-0.005	0.063

Cohort 3

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.536	-0.334	-0.172	-0.021	-0.351	0.111
MLM0	-0.672	-0.379	-0.151	-0.029	-0.329	0.147
Gain	-0.109	-0.08	-0.049	-0.009	-0.064	0.039
TM	-0.173	-0.126	-0.041	-0.049	-0.119	0.025
SGP	-0.219	-0.102	-0.06	0.003	-0.088	0.122

VAM	-0.239	-0.120	-0.086	0.005	-0.104	0.120
MLM	-0.104	-0.105	-0.031	-0.028	-0.080	0.013

Appendix B2. *Mathematics Middle School Model Correlations with School Composition Variables by Cohort.*

Cohort 1

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.362	-0.198	-0.248	-0.013	-0.309	0.166
MLM0	-0.546	-0.285	-0.228	-0.080	-0.269	0.148
Gain	-0.050	0.116	-0.016	-0.021	0.112	0.156
TM	-0.076	0.117	-0.132	0.011	0.024	0.227
SGP	-0.005	0.130	0.039	-0.009	0.103	0.165
VAM	-0.031	0.118	0.024	-0.038	0.089	0.188
MLM	-0.109	0.055	-0.038	-0.054	0.068	0.135

Cohort 2

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.331	-0.234	-0.275	0.084	-0.313	0.137
MLM0	-0.568	-0.349	-0.184	0.098	-0.265	0.137
Gain	0.039	-0.018	-0.058	0.032	-0.040	0.065
TM	-0.015	-0.055	-0.099	0.035	-0.084	0.090
SGP	0.011	0.008	-0.084	0.015	-0.046	0.128
VAM	0.024	0.023	-0.050	0.026	-0.006	0.143
MLM	-0.025	-0.082	-0.059	0.029	-0.081	0.049

Cohort 3

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.296	-0.254	-0.278	0.082	-0.175	0.205
MLM0	-0.605	-0.326	-0.137	-0.040	-0.259	0.184

Gain	-0.016	-0.047	-0.137	0.080	0.097	0.198
TM	-0.049	-0.064	-0.124	0.065	0.076	0.181
SGP	0.008	-0.003	-0.102	0.073	0.121	0.223
VAM	0.024	-0.005	-0.118	0.073	0.104	0.211
MLM	-0.137	-0.107	-0.126	0.057	0.009	0.229

Appendix B3. *Reading/Language Arts Elementary School Model Correlations with School Composition Variables by Cohort*.

Cohort 1

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.539	-0.530	-0.174	-0.035	-0.488	0.073
MLM0	-0.654	-0.551	-0.180	-0.023	-0.476	0.106
Gain	0.212	0.201	0.130	0.001	0.122	0.086
TM	-0.080	0.049	0.081	-0.033	-0.012	0.156
SGP	-0.187	-0.055	-0.021	0.001	-0.111	0.174
VAM	-0.265	-0.124	-0.034	-0.007	-0.176	0.187
MLM	0.251	0.198	0.136	0.023	0.123	0.061

Cohort 2

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.610	-0.506	-0.255	0.013	-0.472	0.078
MLM0	-0.743	-0.587	-0.225	0.013	-0.503	0.084
Gain	0.100	0.175	0.019	-0.002	0.110	0.054
TM	-0.045	0.101	0.025	-0.005	0.046	0.084
SGP	-0.227	-0.051	-0.083	-0.008	-0.097	0.127
VAM	-0.271	-0.077	-0.106	0.005	-0.116	0.151
MLM	0.119	0.152	0.035	0	0.101	0.004

Cohort 3

	School Composition Variable

Model	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.651	-0.542	-0.184	0.029	-0.516	0.040
MLM0	-0.761	-0.568	-0.152	0.017	-0.507	0.081
Gain	0.163	0.175	0.079	-0.071	0.120	0.060
TM	-0.091	0.004	0	0.046	-0.010	0.011
SGP	-0.215	-0.078	-0.019	-0.002	-0.108	0.029
VAM	-0.239	-0.129	-0.037	0.027	-0.130	0.038
MLM	0.057	0.041	-0.009	0.046	0.047	-0.050

Appendix B4. *Reading/Language Arts Middle School Model Correlations with School Composition Variables by Cohort*.

Cohort 1

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.505	-0.412	-0.342	0.197	-0.496	0.104
MLM0	-0.665	-0.495	-0.241	0.114	-0.459	0.059
Gain	0.312	0.259	0.127	-0.006	0.197	0.017
TM	-0.081	0.111	-0.129	0.044	0.024	0.227
SGP	-0.054	0.054	-0.041	0.065	-0.097	0.129
VAM	-0.017	0.055	-0.042	0.013	-0.058	0.138
MLM	0.369	0.29	0.163	-0.061	0.236	-0.005

Cohort 2

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.501	-0.451	-0.197	0.135	-0.510	0.031
MLM0	-0.712	-0.548	-0.203	0.159	-0.508	0.045
Gain	0.359	0.280	0.083	-0.053	0.187	-0.023
TM	-0.016	-0.048	-0.083	0.030	-0.077	0.092
SGP	-0.007	0.033	0.033	0.008	-0.036	0.093
VAM	-0.033	-0.002	-0.023	0.055	-0.087	0.064

MLM	0.369	0.257	0.102	-0.057	0.178	-0.058

Cohort 3

Model	School Composition Variable					
	FRL	LEP	SWD	Sex	Minority	School Size
PP	-0.569	-0.468	-0.251	0.072	-0.438	0.129
MLM0	-0.735	-0.497	-0.207	0.063	-0.467	0.094
Gain	0.306	0.175	0.065	-0.018	0.294	0.099
TM	-0.049	-0.059	-0.124	0.067	0.072	0.182
SGP	-0.128	-0.038	-0.097	0.096	-0.009	0.163
VAM	-0.124	-0.048	-0.048	0.053	-0.028	0.162
MLM	0.357	0.219	0.113	-0.044	0.292	0.085

Appendix C1. *Mathematics Elementary School Model Correlations with School Percentage SWD by Cohort*.

Cohort	School Performance Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM
1	-.201	-.185	.081	.006	-.033	-.029	.099
2	-.209	-.229	.019	-.012	-.014	-.041	.019
3	-.172	-.151	-.049	-.041	-.060	-.086	-.031

Appendix C2. *Mathematics Middle School Model Correlations with School Percentage SWD by Cohort*.

Cohort	School Performance Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM
1	-0.248	-0.228	-0.016	-0.132	0.039	0.024	-0.038
2	-0.275	-0.184	-0.058	-0.099	-0.084	-0.05	-0.059
3	-0.278	-0.137	-0.137	-0.124	-0.102	-0.118	-0.126

Appendix C3. *Reading/Language Arts Elementary School Model Correlations with School Percentage SWD by Cohort*.

Cohort	School Performance Model						
	PP	MLM0	Gain	TM	SGP	VAM	MLM
1	-.174	-.180	.130	.081	-.021	-.034	.136
2	-.255	-.225	.019	.025	-.083	-.106	.035
3	-.184	-.152	.079	.000	-.019	-.037	-.009

Appendix C4. *Reading/Language Arts Middle School Model Correlations with School Percentage SWD by Cohort*.

	School Performance Model

Cohort	PP	MLM0	Gain	TM	SGP	VAM	MLM
1	-0.342	-0.241	0.127	-0.129	-0.041	-0.042	0.163
2	-0.197	-0.203	0.083	-0.083	0.033	-0.023	0.102
3	-0.251	-0.207	0.065	-0.124	-0.097	-0.048	0.113