Spring 2014

A Comparison of a Value-Added Status Model Versus a Value-Added Growth Model for Identifying High Performing Maine Middle Schools

Ida A. Batista
University of Southern Maine, Muskie School of Public Service

A Comparison of a Value-Added Status Model Versus a Value-Added Growth Model for Identifying High Performing Maine Middle Schools

Ida A. Batista

Muskie Public Policy & Management Capstone

Spring 2014

Capstone Advisor: Josephine LaPlante

Email: ia_bat@yahoo.com

A Comparison of a Value-Added Status Model versus a Value-Added Growth Model for Identifying High Performing Maine Middle Schools

Maine policymakers and citizens continue to demand that a high-quality education be provided for all Maine children and provided in a fiscally sound and sustainable manner. High-quality being defined as some measure or measures of school academic performance. These measures are used to identify and classify schools as "High Performing" for those that have met the definition of high-quality and if a school is at the opposite end of the spectrum, "Low Performing". These classifications are not benign. The purposes for identifying schools may be low stakes such as to find attributes of success to share with other schools or to inform the community about their schools but they can also be high stakes as well, such as to determine funding or met federal, No Child Left Behind (NCLB) and state accountability laws. Currently Maine uses a value-added status model criteria system to identify higher performing schools to serve as model schools for determining adequate school funding allocations with the assumption being that higher performing schools are using their state funding in a manner that gets the desired results. Maine has separately funded in-depth studies of these identified High Performing schools to identify attributes of success and common attributes (Silvernail, 2007).

In the current federal school accountability system, NCLB, established in 2001 requires states to develop and evaluate schools' Adequate Yearly Progress (AYP) in getting their students to 100% proficiency by 2014. The charge to states leaves several decisions to the state, such as defining an outcome measure or measures for a school, whether to aggregate data across grades or years, whether to include proficiency across all subject matter areas or to limit to just the federally mandated subject matter areas of mathematics and reading, to name a few of the decisions that a state must make in developing a school accountability system. Since it was foreseen leading up to the 2014 target date that states were not going to met the goal of 100% student proficiency by 2014, the federal government allowed waivers for states which Maine applied for and was granted. Maine has been in the development of a complex method to identify High Performing schools for accountability purposes to meet the requirements of the waiver.

At present there are at least three separate methods in Maine for identifying High Performing schools. One method, a value-added status model is used for funding purposes and studying attributes of High Performing also known as successful schools. The second method, which uses a complex combination of models, status, student growth, & progress to target goal is used for federal and Maine state accountability (Maine Department of Education Accountability 2014). In addition to a third method used by the Maine Department of Education which informally identifies High Performing schools by school results as reported on the school's state report card which uses a status, growth and growth of bottom 25% model to provide snapshots of school performance to schools and communities (Maine Department of Education, Report Cards, 2014). Three pertain questions come immediately to mind:

1. What variables and data are used in these various methods?
2. Are the methods measuring the same thing? And should they be?
3. Is there overlap by method of the schools identified as High Performing?

This study focuses on the model developed for funding purposes since it is a high stake identification and has been the consistent method used for funding purposes within the state for at least the last ten years. The purpose of this study is to explore identification of higher performing Maine middle schools (grade 6 – grade 8) using student longitudinal data in a growth model evaluating growth versus the current value-added model of identification. Research questions include: Are there differences in who is classified as higher performing? What are the differences between the classification methods? Is one method better than the other? What are the implications to using one model versus the other in classification of higher performing schools? Can both models be used?

To be identified as higher performing for purposes of the funding model in Maine an elementary school has to meet four conditions, to be substantially above average in (1) average Maine Comprehensive Assessment System (MeCAS) assessment Score, and above average in (2) percent of students at least partially meeting the MeCAS assessment proficiency standard, (3) percent of students meeting the MeCAS assessment proficiency standard, and (4) standardized residual MeCAS assessment score in a regression analysis (value-added component). Elementary school performance at present is determined by performance in grade 4, or grade 8 for select school configurations.

The first three criteria account for aggregated school status results on the MeCAS assessment. Generally what is publicized is a school’s MeCAS assessment score with schools in wealthier communities receiving praise for higher scores and poorer community schools receiving the subtle shake of the head for not reaching the mark. The expectation and desire is that all schools receive high scores. What is not acknowledged is that it may be easier for schools in wealthier communities to get their students to proficiency and therefore receive high scores with little influence from the school. By adding criteria for the percent meeting proficiency and partially meeting proficiency formally documents schools' efforts in getting all their students to proficiency or at minimum showing a positive trend to proficiency with the attempt to raise school performance to at least proficiency (Figlio & Loeb, 2011). The value-added model criterion has been used to account for demographic and geographic differences among the variety of schools within the state and to utilize limited assessment data available to its full potential. Some limitations within this method have been its reliance on aggregated school assessment data, assessments limited to only certain grades across time, and defining school performance in a manner that is inclusive of only the majority school configurations.

Literature Review

Maine Educational Funding

The Essential Programs and Services report discusses the findings and recommendations for developing and implementing an adequacy based funding model. The Maine State Legislature requested in 1997 of the State Board of Education, establishment of a committee to study educational funding for essential programs and services needed for all Maine students to have the opportunity to achieve the Maine Learning Standards, Essential Programs & Services (EPS) (LD1137). Specifically, the task of the committee encompassed: identification of monetary and nonmonetary essential resources needed by students to reach the Learning Results; estimate costs of the essential resources, and establishing a school achievement accountability system (Maine State Board of Education, 2007).

There are several components of the EPS model that are determined using a successful schools approach such as funding for adequate staffing ratios, which naturally requires the identification of successful schools. The current middle school identification model requires the school to meet all seven criteria to be considered a successful school. The seven criteria are: 1. Scale scores above state average; 2. Scale scores above predicted scores; 3. Percent meets proficiency above state average; 4. Percent partially meets proficiency above state average.

School Accountability

The state has used the mandate from the federal government No Child Left Behind (NCLB) as the basis for structuring a school accountability system at this time. NCLB requires states to develop a school accountability system. The minimum requirement being that states use state assessment data in math and reading as the main source of evidence of academic achievement and either attendance or graduation data. However, the vague mandate leaves a lot of important decisions and their implications for the states to decide. Depending on the outputs and inputs placed in the model and how the data is aggregated and analyzed schools in states can shift wildly from passing to failing from one year to the next and what it means to be a passing or failing school is inconsistent from one state to the next. (NCLB, 2002)

Successful Schools

When studies are done to identify and explore the characteristics of high performing successful schools these schools first need to be identified by some matrix. Several studies of high performing successful schools in several states, including Delaware, Maine, New York City, and Tennessee, showed similar characteristics, such as high expectations, rigorous study, networks of support, monitor student performance, and provide additional attention and time, that are attributed to the success of the school. However, none of the methods for categorizing the schools as high performing and successful were similar. Tennessee used value-added year to year gains; Maine used value-added status; Delaware used a cluster analysis approach across years with performance being at least three points above average in several subject areas; and New York City used a value-added student nested within schools fixed effects model. (Buttram, Cooksy, & Rubright, 2009; Siegel, Amor, Zaltsman, and Fruchter 2005; Silvernail & Stump, 2012; Stone, Bruce, Hursh 2007) This begs the exploration of whether there is an absolute method for identifying high performing successful schools or whether the different methods are measuring similar things.

Measuring School Performance

Status methods, the simplest method and readily attainable through publicly available data provided on state Department of Education websites, represent current year measures, such as percent proficient, and are problematic (Betebenner, 2009a; Linn,2003; Linn, Baker, Betebenner 2002). The choice of the proficiency cut point may have a huge effect on the proficiency rate and may not be comparable across time, especially if there is no vertical alignment to scores, or subject area (Ho, 2008), Depending on how the school proficiency rate is defined it may lack meaning because of aggregation, how thinks are aggregated, and the inability to separate effects of other factors such as student characteristics that are captured in the measure. . In addition, there is no way to ascertain from the measure school effectiveness, what the school adds to the measure (Betebenner, 2009a). A school may be doing a poor job educating students and still have a high level of proficiency because the students are academically capable or a school may be doing a good educating and students may have low proficiency even if the school has helped them make gains. The value-added methods attempt to determine the justifiable contribution of schools to student performance using statistical techniques which leave the school accountable (Betebenner, 2009a; Ballou, Sanders, & Wright, 2004; Raudenbush, 2004). Growth models measure student growth, student achievement across time, using longitudinal data, prior achievement, and sophisticated statistical techniques but don't quantify how much growth is made by a student(Betebenner, 2009c).

There seems to be a lack of information between what seem like distinct elements, establishment of performance standards, the development of status based accountability systems, and research on growth analysis techniques, but in reality these are actually closely linked and lack a normative basis (Betebenner, 2009b).

Difference in data levels also needs to be taken into consideration. When data is aggregated to the school such as school proficiency rate the measure contains factors other than school performance. The average effects of socio-economic factors and issues of attribution with regard to student mobility are associated with school level performance measures (Meyers R., 2003). In addition, complexity is added even when the level is school, when attempting to summarize across all grades and/or all subject areas (Meyers, Lindsay, Condon, & Wan, 2012). Though student growth models are data intensive, it is believed that student effects can be accounted for more readily and interpretation of the model results are more specific.

Methodology

This study included the 52 public Maine middle schools with grade 6 – 8 school configuration. School and student academic achievement data were collected from the Maine Department of Education, covering the 2004-05 through the 2008-09 school years. Academic achievement was measured, during this time period, according to the Maine Educational Assessment (MEA), the test used to measure the performance of all Maine public school students against the state learning standards. Based on performance on the test, each student receives a standardized score, a scale score, and is assigned to one of four proficiency levels relative to the state standards—Does Not Meet, Partially Meets, Meets, or Exceeds—in each subject area tested. The scale score and the percentage of students at each proficiency level were averaged across subject areas, reading, writing, mathematics, and science. However, only reading and math are tested in all grades tested. In the period studied, the MEA was given in the 4^{th}, 8^{th}, and 11^{th} grades in 2004-05, and in 3^{rd} through 8^{th} grades starting in 2005-06. National School Lunch Program data identifying the number and percent of students eligible for free or reduced lunch within a school from 2006-07 through 2008-09 was supplied by the Maine State Department of Education. The U.S. Census 2000 provided data on the number and percentage of adults with a bachelor's degree or higher within a school district.

School level measures developed:

1. *Composite Scale Score Grade 8 0809* (Comp. SS Grade 8 0809) were defined as the average scale score of reading, math, and science on the MEA within the given grade and year.
2. *Composite Scale Score Grade 5 0607* (Comp. SS Grade 5 0607) was defined as the average scale score in reading and math on the MEA within the given grade and year.
3. *Composite Scale Score Grade 4 0405* (Comp. SS Grade 4 0405) was defined as the average scale score in reading, math, science and writing on the MEA within a school if the school could be identified as the only feeder school for the middle school otherwise the district average composite grade 4 scale score was assigned.
4. *Percent Meets or Better Grade 8 0809* (Meets or Better % Grade 8) was defined as the percentage of students achieving a proficiency level of Meets or better in reading, math, and science.

5. *Percent Partially Meets or Better Grade 8 0809* (Partially Meets or Better % Grade 8) was defined as the percentage of students achieving a proficiency level of Partially Meets or better in reading, math, and science.
6. *Percent Free or Reduced Lunch* (Free/Reduced Lunch) is the school three year average percent eligible for free or reduced lunch, 2006-07 through 2008-09.
7. *Percent Bachelor's degree or higher* (BA or Higher) is the 2000 percent of 18 year olds or older within the school district with a bachelor's degree or higher.

Table 1 illustrates basic descriptive statistics for all school level variables also, included in the table are school level descriptive for grades 6 – 7, though no school level aggregate data for these grades was used within this study. Scale scores are comprised of two parts, the first digit indicates grade and the remaining two digits are the score. A score above 40 is considered Meeting or Exceeding the learning standards for the grade. This is the case across all grades and time encompassed within this study, though the scales changed after 2004-05, a scale score above 540 for grade 4 in 2004-05 is considered Meeting or Exceeding. Average school composite scale scores are within the Meets or better range and variation looks consistent across grades with the exception of grade 4 2004-05. This could be due to the composite scale score including all subjects, science and writing tending to be more difficult and generally having lower scores; a more stringent scoring scale, and or the data being school or district depending on the correspondence between the elementary school and middle school.

Table 1. School Level Descriptive Statistics for 52 Maine Middle Schools

	N	Min	Max	Mean	Std.Dev.
Comp. SS Grade 8 0809	52	838	859	847	3.97
Comp. SS Grade 7 0708	52	738	758	747	4.25
Comp. SS Grade 6 0607	52	639	655	646	3.46
Comp. SS Grade 5 0506	52	537	557	544	3.85
Comp. SS Grade 4 0405	52	526	542	534	2.84
Meets or Better % Grade 8	52	45.45	83.33	62.12	9.13
Partially Meets or Better % Grade 8	52	71.86	95.83	84.68	5.81
Free/Reduced Lunch	52	6.76	62.54	36.53	13.94
BA or Higher %	52	8.11	48.53	24.89	9.68

Due to concerns about the different definitions of the school grade level composite scale scores a correlation was run to assess the relationships between the school variables. The correlation results are in Table 2, showing significant strong positive correlations among the school grade level composite scale scores, moderately strong positive relationship with percent bachelor's or higher and moderate negative correlation associated with percent free and reduced lunch.

Table 2. School Level Correlations for 52 Maine Middle Schools

N=52	Comp. SS Grade 8 0809	Comp. SS Grade 7 0708	Comp. SS Grade 6 0607	Comp. SS Grade 5 0506	Comp. SS Grade 4 0405	Meets or Better % Grade 8	Partially Meets or Better % Grade 8	Free/ Reduced Lunch	BA or Higher %
Comp. SS G8 0809	1								
Comp. SS G7 0708	.872**	1							
Comp. SS G6 0607	.832**	.891**	1						
Comp. SS G5 0506	.799**	.808**	.876**	1					
Comp. SS G4 0405	.690**	.711**	.721**	.754**	1				
Meets or Better% G8	.927**	.778**	.771**	.754**	.680**	1			
Partially Meets or Better % G8	.803**	.650**	.605**	.614**	.646**	.891**	1		
Free/Reduced Lunch	-.669**	-.618**	-.637**	-.640**	-.574**	-.678**	-.662**	1	
BA or Higher %	.700**	.636**	.678**	.677**	.585**	.604**	.446**	-.680**	1

**. significant at the 0.01 level

Student level data was obtained for 5,913 students within the 52 middle schools for the school years 2006-07 through 2008-09 and only included data for students who had data for all years of the study period. This is a conservative cohort construction for the grade 8 class of 2008-09, including every repeated measure of academic performance on the MEA from grade 5 (prior to entering middle school) through grade 8. Table 3 provides basic descriptive statistics for all student level variables. Student composite scale scores on average are within the meets range across all grades and have wider variation in scores as grade increases.

Student level measures of performance developed:

1. *Composite Scale Score Grade 5 - 8* (Comp. SS Grade 5 - 8) were defined as the average scale score of reading and math on the MEA for the student within the given grade and year.
2. *Free or Reduced Lunch anytime* (Free/Reduced Lunch anytime) is a student indicator variable (0/1), indicating if the student was eligible for free or reduced lunch at any point between grades 5 – 8.

Table 3. Student Level Descriptive Statistics for 52 Maine Middle Schools

	N	Min	Max	Mean	Std.Dev.
Comp. SS Grade 8 0809	5913	800	880	847.48	14.85
Comp. SS Grade 7 0708	5913	700	780	747.86	14.17
Comp. SS Grade 6 0607	5913	600	680	645.99	13.13
Comp. SS Grade 5 0506	5913	500	580	544.90	11.76
Free/Reduced Lunch anytime	5913	0	1	0.38	0.49

Table 4 displays the student variables correlation results. Student level correlations were similar to school level correlations, significant strong positive relationship across grade composite scale scores and significant negative correlations associated with free or reduced lunch though less than at the school level.

Table 4. Student Level Correlations for 52 Maine Middle Schools

N=5913	Comp. SS Grade 8 0809	Comp. SS Grade 7 0708	Comp. SS Grade 6 0607	Comp. SS Grade 5 0506	Free/Reduced Lunch anytime
Comp. SS Grade 8 0809	1				
Comp. SS Grade 7 0708	.872**	1			
Comp. SS Grade 6 0607	.840**	.856**	1		
Comp. SS Grade 5 0506	.794**	.800**	.824**	1	
Free/Reduced Lunch anytime	-.333**	-.323**	-.316**	-.309**	1

**significant at the 0.01 level

Performance Criteria

To be identified as a higher performing middle school the following four criteria evaluated in grade 8 must be meet:

1. The average cumulative scale score on the state exams (MEA) is at least one-third of a standard deviation higher than state average,
2. The percentage of pupils at or above the *Meets* proficiency level is higher than the state average,
3. The percentage of pupils at or above the *Partially Meets* proficiency level is higher than the state average, and
4. The average cumulative scale score on the MEA is higher than would be predicted based on pupil characteristics and student scores in previous grades.

A middle school is designated as lower performing if its average cumulative scale score on the state MEA is at least one-third of a standard deviation lower than state average and it fails to meet *each* of criteria 2 through 4 above. If the middle school is neither higher performing nor lower performing then the school is classified as having mixed performance.

The first three criteria are status driven criteria and were evaluated by standardizing school level composite scale score grade 8 2008-09, percent meets or better grade 8 2008-09, and percent meets or better grade 8 2008-09. Accepting and applying all three status criteria, Table 5. shows the count and percent of schools falling into each performance category. Based on all three status criteria (All Status) the largest percent (42.3%) of middle schools are classified as lower performing, followed by 34.6% classified as higher performing.

Table 5. Performance Classification Based on Three Status Criteria

	Count	Percent
Higher Performing	18	34.6
Mixed	12	23.1
Lower Performing	22	42.3

Value-Added Criterion

The fourth criterion is the value-added criterion that measures the effect of the school accounting for pupil characteristics, community characteristics and prior academic performance. The customary Maine model used to ascertain residuals and evaluate this criterion has been a single level school aggregated ordinary least square model (OLS). This method has been used due to the lack of student level longitudinal data. As of 2005-06, a unique student identification number has been assigned to every public school student within the state of Maine and attached to all subsequent MEA assessments. The development of the student identification number, standards assessments given for all students grade 3 – grade 8, and the four year time period that has past, allows for schools to be evaluated on a value-added criterion based on students nested within schools (HLM2) two-level model and students nested within schools across time (HLM3) a three-level repeated measure, which evaluates schools based on growth in student scores (Raudenbush, Bryk, Congdon 2004).

Three models were used to explore the value-added criteria:

- OLS single level – School level composite scale score grade 8 2008-09 predicted by districts prior academic performance specifically, composite scale score grade 4 2004-05, school Free or reduced lunch and school district percent BA or higher

- HLM2 two level – Student level composite scale score grade 8 2008-09 predicted by districts prior academic performance specifically, student composite scale score grade 5 2005-06, and student free or reduced lunch status nested within School Level Free or reduced lunch and school district percent BA or higher.

- HLM3 three level repeated measures - Outcome all students composite scale scores grade 5 - grade 8; Level 1 being time, having a value of 0 for grade 5 and increasing by 1 for each additional year with a maximum value of 3 for grade 8 (the repeated measures component); Level 2 student free or reduced lunch status; and Level 3 school level Free or reduced lunch and school district percent BA or higher.

Table 6. displays the school level model results for all three models. All three models are displayed together for convenience, only the OLS and HLM2 are comparable in their coefficients and variance explained (Raudenbush, & Bryk 2002). They are both value-added status components, while the HLM3 evaluates value-added growth. Both the OLS and HLM 2 models have similar coefficients and significance levels. However HLM2 explained more of the variance 72.2% versus 61.4% for OLS. Complete model results are located in the appendix.

Table 6. Middle School Regression Results

School Level Results						
Dependent Variables:	Composite Grade 8 Scale Score 0809		Composite Grade 8 Scale Score 0809		Cohort Composite Scale Scores Grades 5 - 8 (0809 - 0506)	
Independent Variables:	**OLS** (single level)		**HLM2** (two level)		**HLM3** (three level repeated measures)	
	b	*p*	*b*	*p*	*b*	*p*
Percent Free or Reduced Lunch	**-0.093**	**0.014**	**-0.058**	**0.090**	-0.090	0.012
Percent BA or Better	8.933	0.100	8.475	0.100	16.609	0.002
Prior Academic Performance	**0.506**	**0.003**	**0.579**	**0.000**	-	-
Variance Explained	**61.4%**		**72.2%**		**52.8%**	

Results

Residuals from all three models for criterion four were standardized. For both HLM models OLS residuals were used for consistency and comparability as opposed to using empirical Bayes residuals. Correlations for all criteria were run. Table 7 illustrates the results of the correlation. All three status criteria have significant strong positive associations with one another. In relation to the status criteria, the OLS residuals are more strongly associated, especially for composite scale score and percent Meets or better, than the HLM residuals. The HLM3 residual association with the status criteria is positive but moderate suggesting that HLM3, growth, is different than what is being measured within the status criteria. Among the three possible measures for criteria four, the OLS residuals are strongly associated with HLM2 residuals but less strength is in the relationship between OLS residuals and HLM3 residuals. Not surprising, HLM2 residuals are strongly related to HLM3 residuals, which is not unexpected given that the HLM 2 model uses a limited subset of measurement points (prior achievement before middle school and final middle school year achievement, grade 8) and the HLM3 model uses all the measurement points.

Table 7. Criteria Measures Correlations

N=52	Comp. SS	Percent Meets or Better	Percent Partially Meets or Better	OLS Residual	HLM2 Residual	HLM3 Residual
Comp. SS	1					
Percent Meets or Better	.965**	1				
Percent Partially Meets or Better	.842**	.891**	1			
OLS Residual	.623**	.586**	.472**	1		
HLM2 Residual	.532**	.502**	.474**	.780**	1	
HLM3 Residual	.342*	.301*	.322*	.566**	.902**	1

** significant at the 0.01 level; * significant at the 0.05 level

To investigate how the different criterion four measures may influence performance categorization of middle schools scatter plots were created by various criteria, marking the schools by different colors as they are classified as higher performing, mixed, or lower performing by the three status criteria, composite scale score, percent meets or better, and percent partially meets or better. In all the plots presented quadrant II (the upper right) section of the graph represents the most favorable academic performance and quadrant IV (the lower left) represents the least favorable academic performance. Figure 1. illustrates schools classified by all status criteria comparing percent meets or better to percent partially meets or better. In quadrant II are all 18 higher performing schools (green circles) identified by the three status criteria with three mixed results schools. The three mixed results school missed classification of higher performing by not meeting the composite scale score criteria of at least 1/3rd greater than the state. Quadrant IV contains the 22 schools classified as lower performing (red circles) on all status criteria and two mixed results schools that had a composite scale score of greater than -1/3rd of the state average.

Figure 1. ALL Status by Percent Meets or Better & Percent Partially Meets or Better

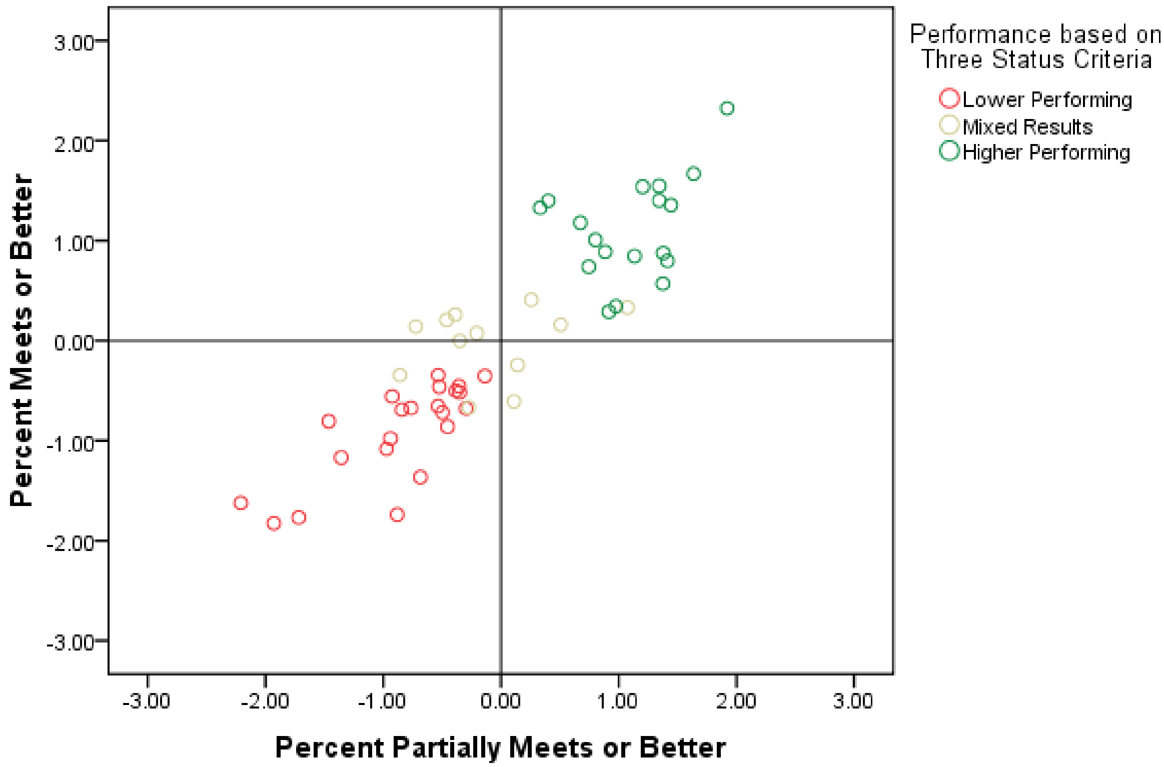
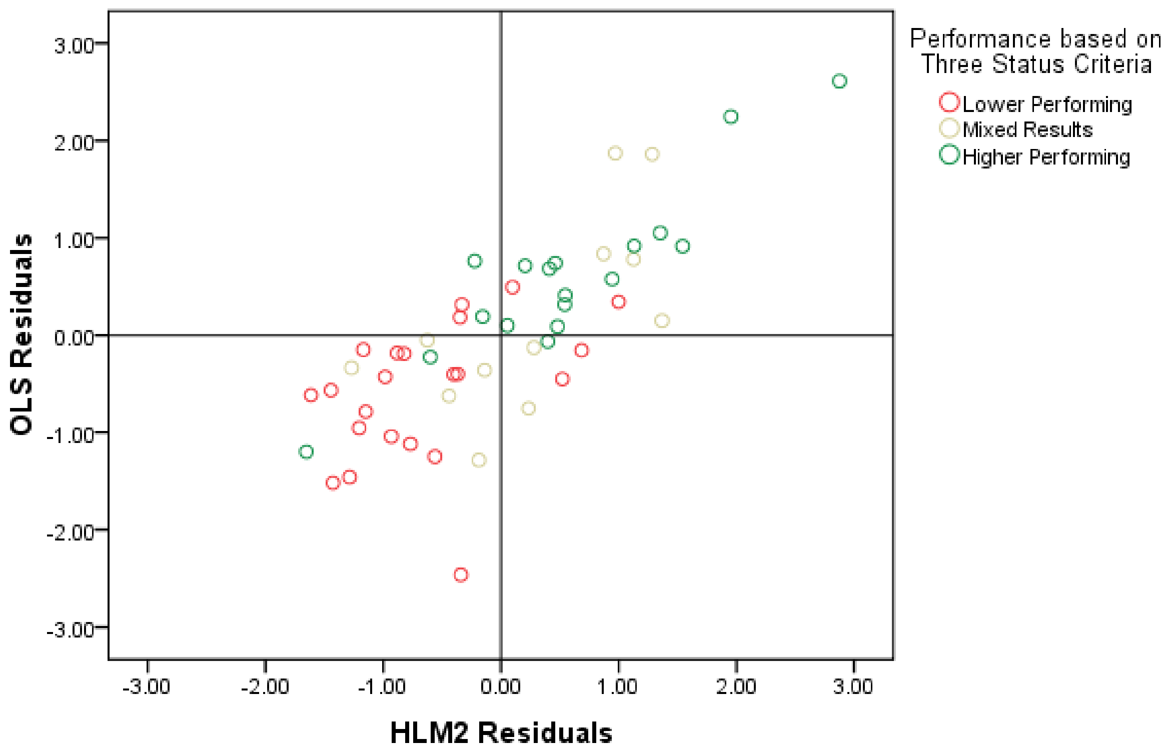


Figure 2. shows schools classified by three status criteria comparing OLS residuals to HLM2 residuals. Looking at just the OLS residual axis, 15 higher performing schools are above average and the remaining three all status higher performing schools are performing lower than predicted while four of the all-status lower performing schools are performing better than predicted. The results from the students nested within schools, HLM2 model show that 14 of the original higher performing schools would still be considered higher performing and 18 out of the original 22 lower performing schools are still classified as lower performing. Though using either of the residual measures identifies 18 lower performing schools it should be noted that only 16 of those schools are the same schools. The two red schools in quadrant I and two red schools in quadrant III switch inclusion in the lower performing category depending on the residuals used. OLS residuals as a measure developed from aggregated school data maybe slightly less discerning than matched individual student data nested within a school, however

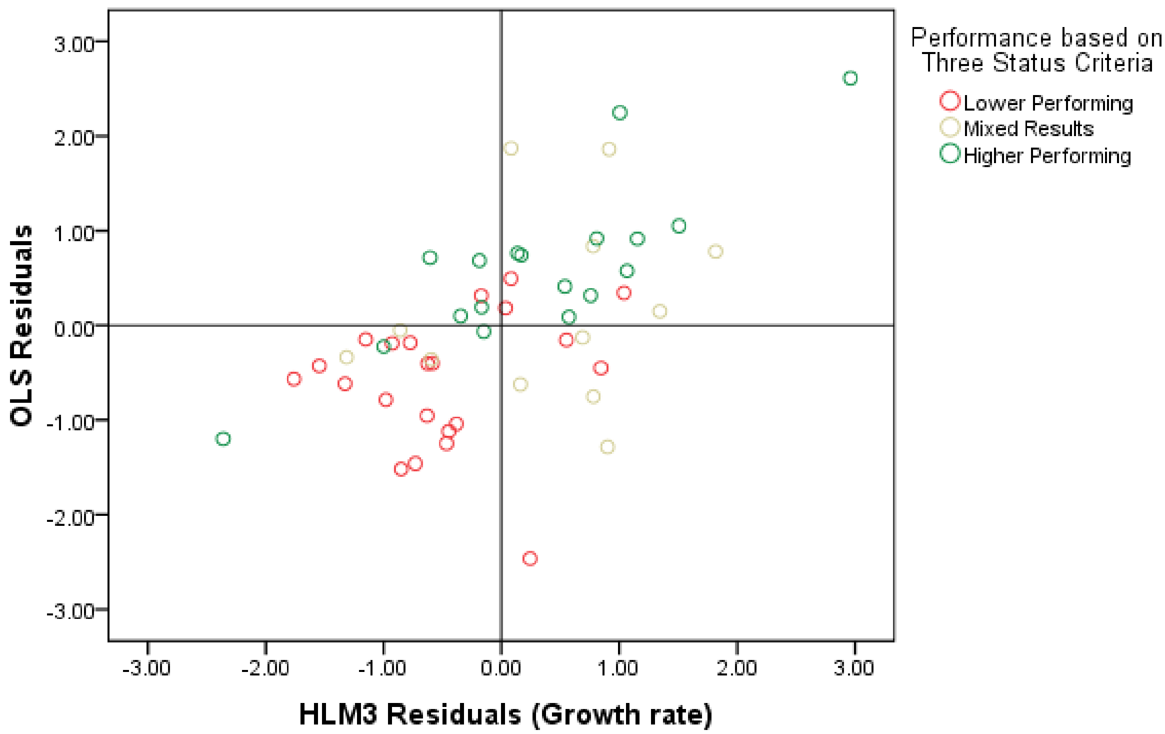
OLS residuals identified 83% of the all status higher performing school in the same quadrants as the HLM2 residuals and has 81% of school agreement with the HLM2 residuals.

Figure 2. All Status by OLS Residuals and HLM2 Residuals



A similar comparison was done comparing OLS residuals to HLM3 residuals, Figure 3 displays the comparison. The HLM3 residuals identify 11 of the 18 all status higher performing as higher performing and 16 schools as lower performing in conjunction with the all status criteria in comparison to the 15 higher performing and 18 lower performing schools identified using the OLS residuals, with 73% agreement on higher performing schools. If both OLS and HLM3 residuals were used the 11 green schools in quadrant II would be considered higher performing and the 15 red schools in quadrant IV would be considered lower performing.

Figure 3. All Status by OLS Residuals and HLM3 Residuals



A summary of categorization of the 52 middle schools based on the possible groupings of criteria is shown in Table 8. Using just status criteria identifies almost 77% of the schools as either higher or lower performing. However when you add a criteria that accounts for the value added by the school and measures how the predicted achievement of the school be it using aggregated school data (OLS residuals) or student data nested within a school (HLM2 residuals) approximately 62% of schools are identified as either higher or lower performing. When the possibility of using longitudinal data is available and a measure of school growth can be used (HLM3 residuals), 52% of schools are identified as either higher or lower performing. Alternatively, both OLS residuals and HLM3 residuals could be used classifying 50% of the schools as either higher or lower performing. In all cases, between the schools identified as higher or lower performing, more schools are identified as lower performing.

Table 8. Performance Classification Based on All Possible Criteria

	All Status Criteria	All Status + OLS	All Status + HLM2	All Status + HLM3	All Status + OLS & HLM3
Higher Performing	18 (34.6%)	15 (28.8%)	14 (26.9%)	11 (21.2%)	11 (21.2%)
Mixed	12 (23.1%)	19 (36.5%)	20 (38.5%)	25 (48.1%)	26 (50%)
Lower Performing	22 (42.3%)	18 (34.6%)	18 (34.6%)	16 (30.8%)	15(28.8%)

Discussion

In Maine a four criteria system is in place to identify higher performing schools. Three of the criteria are status based and the fourth criterion looks at the value-added by the school. Currently this fourth criterion, value-added, is modeled by a one level OLS using aggregated school data for several reasons, the main one being the lack of a unique student identifier far enough back in time to make use of prior grade achievement for each individual student. This study wanted to explore the current method of development for this value-added criterion by comparing it to measures developed using hierarchical linear modeling with a limited set of longitudinal student level data.

All models looked at the same general cohort of 8th grade students in 2008-09 in grade 6 – grade 8 configured schools. All models also had similar covariates within the models. The OLS method is the least complex model methods but uses aggregated school data for achievement making assumptions on the student data incorporated in the aggregate, such as the 8th grade 2008-09 students are the same students contributing to the prior achievement for the given school. HLM models use student data within a school for this study students are considered in the school associated with their 8th grade MEA and conservatively, include only students who have all four years of study period data. Some further work to be done with this current longitudinal data is to explore the HLM models with an additional variable on the students continuity in the school they are attributed to, were the students in the same school for grades 6 - 8 and what is the effect on the residuals based on this information. This additional information helps to inform accounting for a school based on a school being defined primarily by its terminus grade assuming responsibility for the grades contained within the school and the level of proficiency the student has when the student transfers to the following educational phase.

The OLS model is more generous in its identification of higher and lower performing schools. There is considerable overlap using OLS and the three status criteria in identification of higher performing schools, 81% with HLM2 and 73% with HLM3. In practice, if the gold standard is the use of longitudinal data and you have it to use, using HLM3 residuals are the measure that would be put in place, replacing OLS residuals in the criteria system. When a more comprehensive student longitudinal data system is in place the OLS method may be considered a first look approach to school performance identification with the HLM3 method used as a final criteria for identification. Maine's current performance criteria system for identification of higher performing schools, though it over identifies high performing schools based on the current data available, is sufficient for determining adequate funding allocations. No matter whether value-added status or value-added growth is incorporated in performance identification of schools, defining a higher performing district, though science will definitely inform, will be a matter of policy for it will determine how a district is defined.

References

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

Betebenner, D. (2009), *Growth, Standards and Accountability*. Dover, NH: Center for Assessment. Retrieved April 20, 2014 from http://www.researchgate.net/profile/Damian_Betebenner

Betebenner, D. (2009), *Norm- and Criterion-Referenced Student Growth.* Educational Measurement: Issues and Practice, 28: 42–51. doi: 10.1111/j.1745-3992.2009.00161.x Retrieved April 20, 2014 from http://www.researchgate.net/profile/Damian_Betebenner

Betebenner, D, & Linn R.L. (2009), *Growth in Student Achievement: Issues of Measurement, Longitudinal Data Analysis, and Accountability.* Retrieved April 20, 2014 from http://www.k12center.org/publications.html

Buttram, J., Cooksy, L, & Rubright, J. (2009). *Policies and Practices of Successful Delaware High Schools. Newark, DE: Delaware Education Research & Development Center.*

Figlio D, & Loeb S. (2011) *School Accountability* In Hanushek, E.A, Machin, S, Woessmann, L (2011), *Economics of Education*. Handbooks in Economics, Vol. 3, Amsterdam: Elsevier Science, 2011, pp.383-421. Retrieved April 20, 2014 from http://cepa.stanford.edu/sites/default/files/Accountability_Handbook.pdf

Ho, A. D. (2008). *The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind.* Educational Researcher, 37(6), 351-360. Retrieved April 28, 2014 from www.cosa.k12.or.us/.../EESC/**Ho**%202008%20on%20**Proficiency**%20Rates.pdf

Linn, R.L., Baker, E.L., Betebenner D.W. (2002) Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*,31(6), 3-16.

Linn, R.L. (2003) *Accountability: Responsibility and reasonable expectations* (Tech Rep.). Los Angeles, CA: Center for the Study of Evaluation, CRESST

Maine State Board of Education. (1997). *Essential Programs and Services*. Augusta, ME: Maine State Board of Education.

Maine Department of Education (n.d) *Accountability* Retrieved from http://www.maine.gov/doe/accountability/index.html

Maine Department of Education (n.d) *Report Cards* Retrieved from http://www.maine.gov/doe/schoolreportcards/resources/methodology.html)

Meyers, C, Lindsay, J, Condon, C, & Wan, Y (2012). *A Statistical Approach to Identifying Schools Demonstrating Substantial Improvement in Student Learning*. Journal of Education for Students Placed at Risk. (JESPAR) V.17(pages:70-91).

Meyers, R.H., (2003). *Value-added Indicators: Do They Make* an Important Difference? *Evidence from the Milwaukee Public Schools.* Report by Wisconsin Center for Education Research. University of Wisconsin-Madison, No. 2003-5

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Raudenbush, S. W., & Bryk, A.S. (2002).*Hierarchical Linear models: Applications and data analysis methods* (2nd ed). Newbury Park, CA: Sage Press.

Raudenbush, S.W., Bryk, A.S, & Congdon, R. (2004). HLM 6 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

Raudenbush, S. W. (2004). *What are value-added models estimating and what does this imply for statistical practice?* Journal of Educational and Behavioral Statistics, 29(1), 121– 129

Siegel, Dorothy, Hella Bel Hadj Amor, Ariel Zaltsman, and Norm Fruchter. 2005. Assessing *Success in New York City High Schools.* Report by New York University Institute for Education and Social Policy for New Visions for Public Schools. New York: New York University, Steinhardt School of Education, Institute for Education and Social Policy Retrieved May10, 2014 from www.annenberginstitute.org/pdf/BTO_report.pdf

Silvernail, D. L. (2007). *The identification of higher and lower performing Maine schools: School profiles and characteristics.* Gorham, ME: University of Southern Maine, Center for Education Policy, Applied Research, and Evaluation

Silvernail, D., & Stump, E. (2012). *More Efficient Public Schools in Maine: Learning Communities Building the Foundation of Intellectual Work.* Gorham, ME: Center for Education Policy, Applied Research, & Evaluation.

Stone,J., Bruce, G., & Hursh, D. (2007). *Effective Schools, Common Practices Twelve Ingredients of Success from Tennessee's Most Effective Schools*. Education Consumers Foundation.

Appendix

A. OLS Regression

Descriptive Statistics

	Mean	Std. Deviation	N
CompSSg8_0809	846.6612	3.97192	52
Avg.FRLunch07-09	36.5328	13.94150	52
%BAorBetter	.248941	.0967961	52
COMPSSG4_0405	533.6386	2.83822	52

Correlations

		CompSSg8 0809	Avg.FRLunch 07-09	%BAorBetter	COMPSSG4 0405
Pearson Correlation	CompSSg8_0809	1.000	-.684	.652	.677
	Avg.FRLunch07-09	-.684	1.000	-.680	-.574
	%BAorBetter	.652	-.680	1.000	.585
	COMPSSG4_0405	.677	-.574	.585	1.000
Sig. (1-tailed)	CompSSg8_0809	.	.000	.000	.000
	Avg.FRLunch07-09	.000	.	.000	.000
	%BAorBetter	.000	.000	.	.000
	COMPSSG4_0405	.000	.000	.000	.
N	CompSSg8_0809	52	52	52	52
	Avg.FRLunch07-09	52	52	52	52
	%BAorBetter	52	52	52	52
	COMPSSG4_0405	52	52	52	52

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method

1	COMPSSG4_0405, Avg.FRLunch07-09, %BAorBetter[a]	.	Enter

a. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.782[a]	.612	.587	2.55142

a. Predictors: (Constant), COMPSSG4_0405, Avg.FRLunch07-09, %BAorBetter

b. Dependent Variable: CompSSg8_0809

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	492.115	3	164.038	25.199	.000[a]
	Residual	312.469	48	6.510		
	Total	804.583	51			

a. Predictors: (Constant), COMPSSG4_0405, Avg.FRLunch07-09, %BAorBetter

b. Dependent Variable: CompSSg8_0809

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	577.618	86.696		6.663	.000
	Avg.FRLunch07-09	-.093	.037	-.328	-2.554	.014
	%BAorBetter	8.933	5.322	.218	1.679	.100
	COMPSSG4_0405	.506	.162	.362	3.117	.003

a. Dependent Variable: CompSSg8_0809

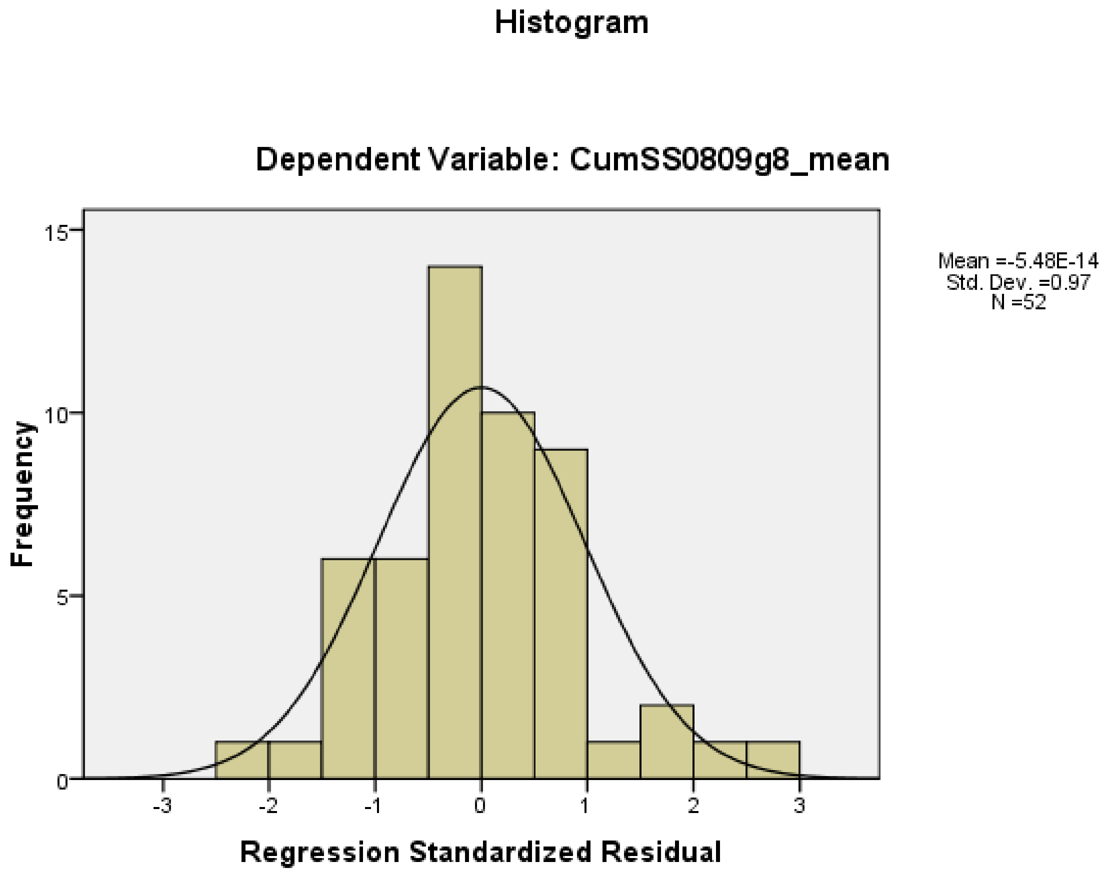
Residuals Statistics[a]

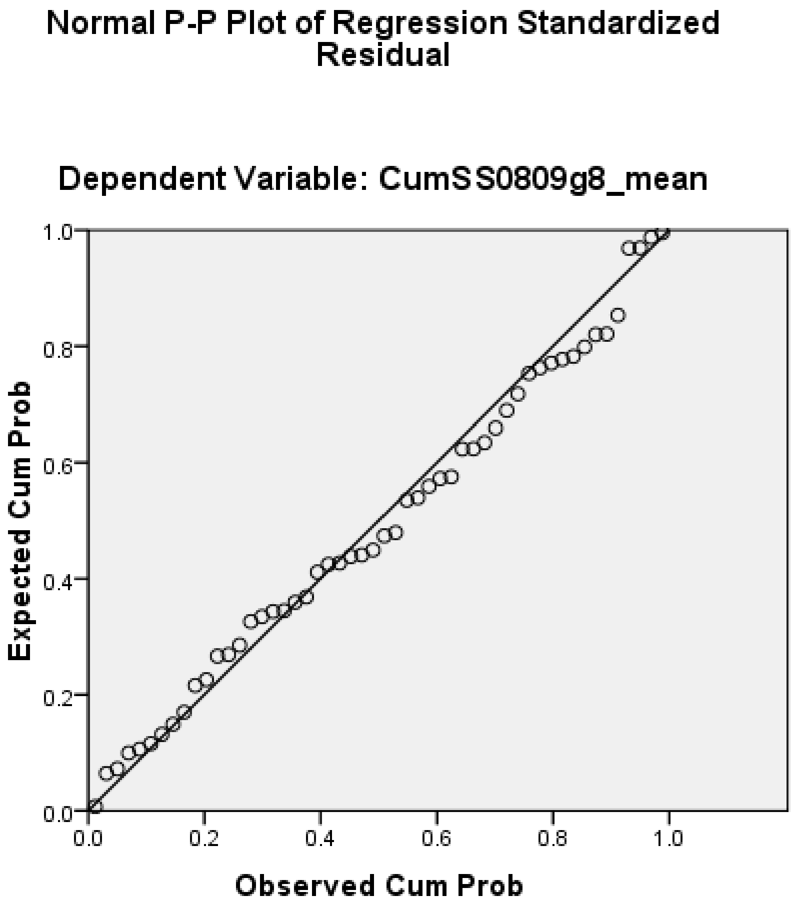
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	839.2379	854.6393	846.6612	3.10633	52

Residual	-6.28594	6.66230	.00000	2.47524	52
Std. Predicted Value	-2.390	2.568	.000	1.000	52
Std. Residual	-2.464	2.611	.000	.970	52

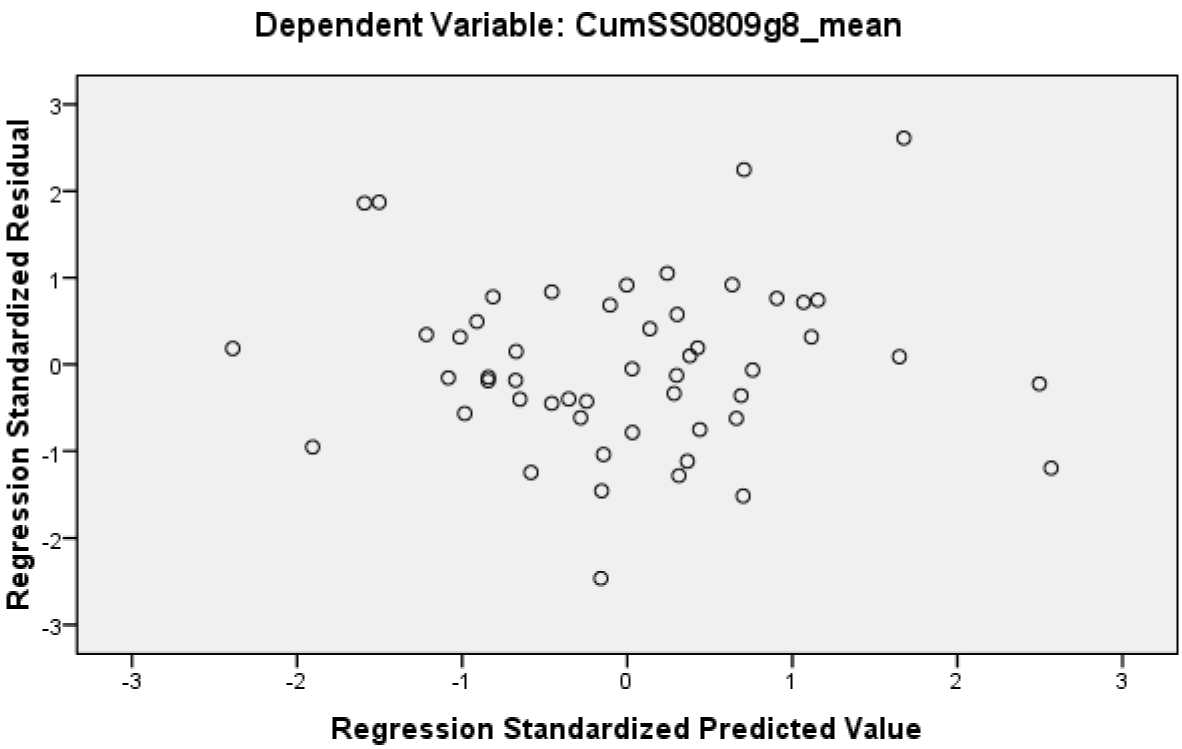
a. Dependent Variable: CompSSg8_0809

Charts




Normal P-P Plot of Regression Standardized Residual
Dependent Variable: CumSS0809g8_mean
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

Scatterplot



Results StatusModel 2L conditional

```
 Program:                       HLM 6 Hierarchical Linear and Nonlinear Modeling
 Authors:                       Stephen Raudenbush, Tony Bryk, & Richard Congdon
 Publisher:                     Scientific Software International, Inc. (c) 2000
                                                    techsupport@ssicentral.com
                                                            www.ssicentral.com
 -------------------------------------------------------------------------------

 SPECIFICATIONS FOR THIS HLM2 Conditional Model RUN


  Problem Title: Status Model 2 Level Cond 030310

  The data source for this run  = Stmdl2L
  The command file for this run = whlmtemp.hlm
  Output file name              = C:\Documents and
Settings\ibatista\Desktop\AEFA2010_Growth\HLM_SPSS\HLM_StatusModel\New\hlm2.txt
  The maximum number of level-1 units = 5913
  The maximum number of level-2 units = 52
  The maximum number of iterations = 100
  Method of estimation: restricted maximum likelihood

 Weighting Specification
 -----------------------
                        Weight
                        Variable
           Weighting?   Name        Normalized?
 Level 1       no
 Level 2       no
 Precision     no

  The outcome variable is compssg80809

  The model specified for the fixed effects was:
 ----------------------------------------------------

   Level-1                  Level-2
   Coefficients             Predictors
 ---------------------   ---------------
         INTRCPT1, B0      INTRCPT2, G00
$                            AVE%FR, G01
$                             %BAPL, G02
$                       cmpSSg50506, G03
#*   FRANY slope, B1       INTRCPT2, G10
#* STdcompSSg50506 slope,B2 INTRCPT2, G20
$                            AVE%FR, G21
$                             %BAPL, G22
$                       cmpSSg50506, G23

'#' - The residual parameter variance for this level-1 coefficient has been set
      to zero.
'*' - This level-1 predictor has been centered around its group mean.
'$' - This level-2 predictor has been centered around its grand mean.

 The model specified for the covariance components was:
 ---------------------------------------------------------

         Sigma squared (constant across level-2 units)

         Tau dimensions
               INTRCPT1
```

Results StatusModel 2L conditional

Summary of the model specified (in equation format)

Level-1 Model

Y = B0 + B1*(FRANY) + B2*(STdcompSSg50506) + R

Level-2 Model

B0 = G00 + G01*(AVE%FR) + G02*(%BAPL) + G03*(schcompSSg50506) + U0
B1 = G10
B2 = G20 + G21*(AVE%FR) + G22*(%BAPL) + G23*(schcompSSg50506)

Iterations stopped due to small change in likelihood function
♀******* ITERATION 5 *******

Sigma_squared = 74.57375

Tau
INTRCPT1,B0 4.55876

Tau (as correlations)
INTRCPT1,B0 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.847

The value of the likelihood function at iteration 5 = -2.120289E+004
♀ The outcome variable is compssg80809

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	847.090814	0.321866	2631.811	48	0.000
AVE%FR, G01	-0.057939	0.033531	-1.728	48	0.090
PCBAPL, G02	8.474521	5.063256	1.674	48	0.100
schcompSS0506, G03	0.578953	0.121415	4.768	48	0.000
For FRANY slope, B1					
INTRCPT2, G10	-2.795011	0.254048	-11.002	5904	0.000
For STdcompSSg50506 slope, B2					
INTRCPT2, G20	0.994586	0.010734	92.654	5904	0.000
AVE%FR, G21	0.000835	0.001023	0.816	5904	0.414
PCBAPL, G22	-0.188356	0.153684	-1.226	5904	0.221
schcompSS0506, G23	0.006487	0.003844	1.688	5904	0.091

The outcome variable is compssg80809

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	847.090814	0.308715	2743.923	48	0.000
AVE%FR, G01	-0.057939	0.030959	-1.871	48	0.067

Results StatusModel 2L conditional

PCBAPL, G02	8.474521	5.589548	1.516	48	0.136
schcompSS0506, G03	0.578953	0.139354	4.155	48	0.000
For FRANY slope, B1					
INTRCPT2, G10	-2.795011	0.311827	-8.963	5904	0.000
For STdcompSSg50506 slope, B2					
INTRCPT2, G20	0.994586	0.017780	55.938	5904	0.000
AVE%FR, G21	0.000835	0.001893	0.441	5904	0.659
PCBAPL, G22	-0.188356	0.255289	-0.738	5904	0.461
schcompSS0506, G23	0.006487	0.007520	0.863	5904	0.389

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	2.13512	4.55876	48	355.89749	0.000
level-1,	R	8.63561	74.57375			

Statistics for current covariance components model

Deviance = 42405.771744
Number of estimated parameters = 2

Results StatusModel 3L conditional

```
 Program:                       HLM 6 Hierarchical Linear and Nonlinear Modeling
 Authors:                       Stephen Raudenbush, Tony Bryk, & Richard Congdon
 Publisher:                     Scientific Software International, Inc. (c) 2000
                                                      techsupport@ssicentral.com
                                                              www.ssicentral.com
 -------------------------------------------------------------------------------

 SPECIFICATIONS FOR THIS HLM3 Conditional Model RUN


  Problem Title: growthmdl_ All err in model

  The data source for this run  = GrowthMdl3L
  The command file for this run = whlmtemp.hlm
  Output file name              = C:\Documents and
Settings\ibatista\Desktop\AEFA2010_Growth\HLM_SPSS\HLM_3L_Time\New\hlm3.txt
  The maximum number of level-1 units = 23652
  The maximum number of level-2 units = 5913
  The maximum number of level-3 units = 52
  The maximum number of iterations = 100
  Method of estimation: full maximum likelihood

  The outcome variable is   COMPSSallgrades

  The model specified for the fixed effects was:
 ----------------------------------------------------

   Level-1                 Level-2          Level-3
   Coefficients            Predictors       Predictors
 ---------------------  ---------------   ----------------
        INTRCPT1, P0      INTRCPT2, B00    INTRCPT3, G000
                                         %  AVE%FR, G001
                                         %   %BAPL, G002
                        #*FRANYTIM, B01    INTRCPT3, G010
      TIME slope, P1      INTRCPT2, B10    INTRCPT3, G100
                                           %AVE%FR, G101
                                         %   %BAPL, G102
                        #*FRANYTIM, B11    INTRCPT3, G110

 '#' - The residual parameter variance for the parameter has been set to zero
 '*' - This variable has been centered around its group mean
 '%' - This variable has been centered around its grand mean


 Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model

        Y = P0 + P1*(TIME) + E

Level-2 Model

        P0 = B00 + B01*(FRANYTIM) + R0
        P1 = B10 + B11*(FRANYTIM) + R1


Level-3 Model

        B00 = G000 + G001(AVE%FR) + G002(%BAPL) + U00
        B01 = G010
        B10 = G100 + G101(AVE%FR) + G102(%BAPL) + U10
```

```
                          Results StatusModel 3L conditional
        B11 = G110
♀ For starting values, data from 23652 level-1 and 5913 level-2 records were used

Iterations stopped due to small change in likelihood function
♀******* ITERATION 7 *******

 Sigma_squared =     26.82829

 Standard Error of Sigma_squared =      0.34889

Tau(pi)
 INTRCPT1,P0     96.11296      8.38974
     TIME,P1      8.38974      2.56447


Tau(pi) (as correlations)
 INTRCPT1,P0  1.000  0.534
     TIME,P1  0.534  1.000

Standard Errors of Tau(pi)
 INTRCPT1,P0      2.13624      0.40794
     TIME,P1      0.40794      0.16226


 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, P0                        0.837
      TIME, P1                        0.323
 ----------------------------------------------------

Tau(beta)
 INTRCPT1         TIME
 INTRCPT2,B00 INTRCPT2,B10
    5.27449     -0.85328
   -0.85328      0.60564

Tau(beta) (as correlations)
 INTRCPT1/INTRCPT2,B00  1.000 -0.477
     TIME/INTRCPT2,B10 -0.477  1.000

Standard Errors of Tau(beta)
 INTRCPT1         TIME
 INTRCPT2,B00 INTRCPT2,B10
    1.27098      0.31681
    0.31681      0.13564

 ----------------------------------------------------
  Random level-2 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1/INTRCPT2, B00               0.808
      TIME/INTRCPT2, B10               0.873
 ----------------------------------------------------

The value of the likelihood function at iteration 7 = -8.211252E+004
♀ The outcome variable is   COMPSSallgrades

 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                      Standard                Approx.
    Fixed Effect        Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, P0
```

For INTRCPT2, B00					
INTRCPT3, G000	544.596144	0.354251	1537.316	49	0.000
AVE%FR, G001	-0.090092	0.034537	-2.609	49	0.012
%BAPL, G002	16.608698	5.033522	3.300	49	0.002
For FRANYTIM, B01					
INTRCPT3, G010	-6.621222	0.303147	-21.842	5911	0.000
For TIME slope, P1					
For INTRCPT2, B10					
INTRCPT3, G100	100.997646	0.115513	874.342	49	0.000
AVE%FR, G101	-0.007900	0.011313	-0.698	49	0.488
%BAPL, G102	0.405188	1.642293	0.247	49	0.806
For FRANYTIM, B11					
INTRCPT3, G110	-0.900964	0.079643	-11.313	5911	0.000

The outcome variable is COMPSSallgrades

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, P0					
For INTRCPT2, B00					
INTRCPT3, G000	544.596144	0.352718	1544.000	49	0.000
AVE%FR, G001	-0.090092	0.033343	-2.702	49	0.010
%BAPL, G002	16.608698	5.241063	3.169	49	0.003
For FRANYTIM, B01					
INTRCPT3, G010	-6.621222	0.456713	-14.498	5911	0.000
For TIME slope, P1					
For INTRCPT2, B10					
INTRCPT3, G100	100.997646	0.115105	877.436	49	0.000
AVE%FR, G101	-0.007900	0.009607	-0.822	49	0.415
%BAPL, G102	0.405188	1.978469	0.205	49	0.839
For FRANYTIM, B11					
INTRCPT3, G110	-0.900964	0.104717	-8.604	5911	0.000

Final estimation of level-1 and level-2 variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	R0	9.80372	96.11296	5860	35858.05086	0.000
TIME slope,	R1	1.60140	2.56447	5860	8661.36645	0.000
level-1,	E	5.17960	26.82829			

Final estimation of level-3 variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1/INTRCPT2, U00	2.29663	5.27449	49	324.90054	0.000
TIME/INTRCPT2, U10	0.77823	0.60564	49	492.02834	0.000

Statistics for current covariance components model

```
                         Results StatusModel 3L conditional
 --------------------------------------------------
Deviance                     = 164225.035054
Number of estimated parameters = 15
```