School Conditional Growth Model: How to Make an "Apples to Apples" Comparison Possible?

Yun Xiang and Carl Hauser

Yun Xiang (yun.xiang@nwea.org) is a Research Specialist for the Kingsbury Center at NWEA.

Carl Hauser (carl.hauser@nwea.org) is a Senior Research Specialist in the Measurement and Data Services division of NWEA.

Table of Contents

INTRODUCTION

The increasing demand to hold schools accountable for student outcomes has drawn a great deal of attention. It is generally acknowledged that an effective school and school system provides a positively stimulating environment for academic learning. Student learning in many studies is quantified, and to some extent, simplified as a test score. The use of standardized test data as a means for assessing student progress and evaluating school effectiveness is quite controversial (Kane, & Staiger, 2002; Linn, 2000, 2003, 2004). Moreover, varying definitions of school effectiveness and the systems based on these different beliefs complicate the discussion of school evaluation and school accountability. Carlson (2001) introduced a four-quadrant matrix to illustrate the different definitions of school quality and progress currently applied in many school systems (see Table 1).

Table 1: Carlson's matrix of the four methods of judging school quality and progress (2001)

	How good is this school? (status)	**Is it getting better? (change)**
Achievement	Q 1: What is the achievement level of the students in this school? (achievement status of one cohort)	Q2: Is the achievement level of this school improving? (change in status between successive cohorts)
Effectiveness	Q3: Is this an effective school? Given the achievement level of students when they enter, how much do they learn while they are in the school? (grade-to-grade slope of one cohort)	Q4: Is this school becoming more effective? How much more, or less, are the students learning this year than they did the year before? (change in slopes between successive cohorts)

Under NCLB, percentage of students above a certain state-defined performance level (commonly labeled proficiency) is widely used in the state-level reporting system. When school quality refers to what percentage of students who have performed above proficiency (see Quadrant 1), school improvement then becomes the rate at which this percentage increases (see Quadrant 2). Researchers have pointed out, however, that the core element of this accountability system—the cut-score used to define student proficiency - is misleading and easy to manipulate for different purposes (Lion 2000, 2004). Another argument against the status-based accountability system is that schools with a large proportion of academically disadvantaged students are unfairly judged since the non-school factors (i.e. student socio-economic status) can be confounded with achievement status (Braun, 2008; Linn, Baker, & Betebenner, 2002; Kane, & Staiger, 2002; Raudenbush, 2004). To help offset these effects, growth-based measures of school performance have gained approval from the U.S. Department of Education as a supplement to status-based accountability models.

Changing the measure of school performance is on the table. One option is the growth model which emphasizes individual students' progress. There are generally two approaches used to measure school effectiveness in the systems: (1) progress—the test scores of at least three consecutive or nonconsecutive years based on different cohorts of students; (2) growth—the test scores from at least three consecutive or nonconsecutive years based on the same cohort of students. Conceptually, the second approach, modeling of student achievement growth better represents the time-dependent process of academic learning and allows a degree of control over student background characteristics to be exercised (Seltzer, Choi & Thum, 2003; Willett, 1988).

More and more states are starting to apply different growth models as a supplement to their status-based accountability models. The new Elementary and Secondary Education Act (ESEA) blueprint proposed by the Obama administration also includes the idea of measuring student growth over time (U.S. Department of Education, 2010).

Growth models are increasingly seen as a statistically valid and readily understood approach in educational accountability systems. But are growth measures a panacea for our school accountability system? There are a few questions that we may want to explore before we so confidently rely on these new measures.

What is growth? Growth can mean different things. Growth can be the tracking of a single group of students' achievement or looking at trends based on different groups. Growth can be changes in percentile rankings or performance categories. Growth can be based on an absolute score gain or score difference relative to a norm population. The various definitions and models of growth that are currently being applied in different states can lead to large discrepancies in accountability results.

Do high-performing and low-performing students grow in the same fashion? If not, what are the implications for school evaluations? One criticism of the current system is that, in order to avoid sanctions, schools must expend disproportionate effort to improve the achievement of marginally performing students at the sacrifice of those who are high above or far below proficiency targets. In the growth-based accountability system, however, high achievers may still be weighed disproportionately due to the reality that there might be less room for them to grow.

Does growth modeling give adequate consideration to variability of student achievement? Measures of average performance, such as the average student achievement score or average growth rate, tend to be the most common indicators for tracking school performance. But averages tend to oversimplify what is going on within a school. A great deal more information about school performance can be learned by examining the full range (in other words, the distribution, or variability) of scores and growth rates.

Is growth a too rigid reflection of school effectiveness? As opposed to the AYP status-based measure, growth measures are believed to effectively control for contextual characteristics by taking prior achievement into account. However, non-curriculum factors - such as student socioeconomic status, teacher-student ratio, or school size - can still be associated with student growth. When school growth rate is used to evaluate school effectiveness, the growth model relies on a very important assumption—the aggregated school rate of change is not confounded with school characteristics since growth itself can have a degree of control over students' initial achievement. For example, schools with a large proportion of disadvantaged students are assumed to grow as fast as other schools. The growth-based accountability system is thus subjected to serious scrutiny.

The last aspect of using growth-based accountability measure is the focus of this study. To understand the complex school environment when schools are evaluated and compared, the concept of conditional growth study is presented to examine how educational settings can be designed to address the interrelated social demands of learning. Specifically, to further address the possibility that schools with a large proportion of academically disadvantaged students are unfairly evaluated and judged, we take into account school characteristics and develop conditional school growth to make the "apples to apples" comparison possible. Unconditional growth models depict school growth that depends solely on time, while conditional growth models explicitly account for school characteristics.

Another approach to enhancing fairness when evaluating schools is to examine how well a school increasingly improves the rate of change for successive cohorts (reflected in Quadrant 4, Table 1). When changes in growth rates of successive cohorts, instead of one single cohort, are compared, school characteristics will be less confounded with school outcomes. This approach also addresses the issue of cohort effects where results of one cohort are generalized as representative of an entire school. This cohort-to-cohort approach provides a sound alternative for measuring school improvement. This approach is not discussed in this study.

The purpose of this paper is to offer an analytic perspective to policy makers and educational practitioners regarding how to use longitudinal achievement data to evaluate schools. We further discuss the potential practical applications of our models for superintendents, researchers, and policy makers. The premise of the study is that the complexity of the school context can be leveraged within longitudinal growth models to account for more variance than the unconditional counterparts of these models. The following research questions were considered:

1. Do school growth rates differ when school characteristics are taken into account in growth modeling?
2. Do school growth rates differ when school initial status is taken into account in growth modeling?
3. How are schools evaluated differently based on the application of an unconditional model and a conditional model?

Methods

Data and Measurement

The study includes test records from almost 50,000 students in 476 schools located in a single state. The study focuses on one cohort of students by tracking their mathematic achievement growth from Grade 3 in term Fall 2006 to Grade 5 in term Spring 2009. The data came from the Growth Research Database (GRD®) developed and maintained by the Northwest Evaluation Association (NWEA).

The basic data structure:

- Was limited to beginning of year 2006 forward (see Figure 1).
- Excluded summer term tests (June 15 – August 15).
- Included only school-grade combinations when N ≥ 16.

One important criterion to conduct such a growth study is that the measurement scales remain stable over time. Hence, the skills possessed by a student with a score of 210 of different years are the same as the skills possessed by a student achievement that score at different times. It is that constancy of the MAP™ (Measures of Academic Progress™) scale in NWEA that allows us to measure growth and to compare student performance across time.

All of the test records used in this study were from computerized adaptive tests. All tests within a common subject were of similar length and content structure. All test items within a subject are calibrated to a common scale using item response theory. Thus, scores from different tests in the same subject can be compared and interpreted in the same manner.

Table 2 shows student math achievement on average grows across the six selected terms. The general trend of achievement growth does not really give us details about how schools grow at different rates and how school growth rates interact with different school characteristics.

Table 2: Mean and standard deviation of math achievement scores by term

Term	Grade	Mean	SD
Fall 2006	3	193.15	11.09
Spring 2007	3	202.40	12.06
Fall 2007	4	203.12	11.99
Spring 2008	4	211.10	13.13
Fall 2008	5	212.09	13.37
Spring 2009	5	220.69	14.41

Methods and Procedures

To use achievement growth trajectories to evaluate schools, we conducted a two-step analysis:

The first step examined the general trends in school mean growth. The unconditional growth provides insight into how students in each school change over time in general. The models are unconditional in the sense that growth is modeled as a function of time; common school predictors, such as percent Free-Reduced Lunch (FRL) students, percent minority students, school size, etc, are not used. The schools are ranked based on their estimated mean growth rates out of 476 schools.

In the section below we illustrate the role of hierarchical linear models in screening for different school growth patterns. The models outlined are three-level hierarchical models:

Level One: Test events (Repeated measures)

We begin at level 1 with an individual growth model of the academic achievement at time t of student i in school j:

$$Y_{tij} = \pi_{oij} + \pi_{1ij}(Term)_{tij} + e_{tij}, \qquad e_{tij} \sim N(0, \sigma^2) \quad (1)$$

Level Two: Students (individual growth trajectory)

$$\pi_{0ij} = \beta_{00j} + r_{0ij}, \quad (2)$$
$$\pi_{1ij} = \beta_{10j} + r_{1ij}, \quad (3)$$

Level Three: Schools (Unconditional)

$$\beta_{0ij} = \gamma_{000} + \mu_{00j}, \quad (4)$$
$$\beta_{1ij} = \gamma_{100} + \mu_{10j}, \quad (5)$$

The second step models rate of change on school-level variables. In this model, we only focus on what school factors are associated with school growth rate. Hence, we only modeled school level factors as predictors to school rate of change. We then discuss how schools can be fairly evaluated when their school characteristics are taken into account.

We first examined what factors could be significantly associated with school growth rates. Then we ranked schools based on estimated mean growth rates after controlling for selected school contextual variables and school mean initial status. School rankings were compared and correlated with part one.

Level Three: Schools (Conditional)

$$\beta_{0ij} = \gamma_{000} + \sum_{q}^{Q} \gamma_{00q} X_{qj} + \mu_{00j}, \quad (6)$$

$$\beta_{1ij} = \gamma_{100} + \sum_{q}^{Q} \gamma_{10q} X_{qj} + \mu_{10j}, \quad (7)$$

The selected school variables (q) include:

- School size
- Teacher-pupil ratio
- Percentage of FRL students
- Percentage of minority students
- Title one school or not
- Urbanicity

In this part we also examined the relationship between school growth rates and their initial status. The relationship was analyzed only at the school level. School initial status is employed as a predictor of school rate of change in the model. The purpose is to investigate how much of a change in rate of change is associated with 1 RIT point increase in school mean initial score. Thus we expand our school rate of change model as follows:

$$\beta_{0ij} = \gamma_{000} + +\mu_{00j} \text{ , (8)}$$

$$\beta_{1ij} = \gamma_{100} + \sum_{q}^{Q} \gamma_{10q} X_{qj} + \boldsymbol{b}(\boldsymbol{\beta_{0ij}}) + \mu_{10j} \text{ , (9)}$$

where b is the parameter that captures the relationship between school initial status and school growth rate. Note that the regression is termed as a latent variable regression, and this coefficient b is termed latent variable regression coefficients (Raudenbush & Bryk, 2002; Seltzer, Choi, & Thum, 2003).

Results and Conclusions

Unconditional School Growth Model

The unconditional growth model depicts how student achievements change over time without considering at what type of schools students are. This model does not require any other data structure except tracking individual students across time. Table 3 displays the fixed and random effects of the three-level unconditional model. It shows that on average all schools started at 192.9 in term 2006 fall and increased at the rate of 5.2 points per term. The small standard errors for the two growth parameters (0.21 & 0.03) indicate that the true initial status and rate of change fell into a relatively narrow range.

The second panel of Table 3 shows that the variance in initial status and rate of change was decomposed into within- and between-school components. Significant variation was found within schools (among students) for individual initial status and individual growth rates as well as between schools for school mean initial status and school mean growth rates. One can see that the variation in initial status between schools (19.1) was much smaller than the variation within schools (91.6), while the variation in rate of change between schools (0.5) was close to the one within schools (0.6).

Based on the variance component estimates, we can compute the percentage of variation that lies between schools in both initial status and growth rate. Specifically, percent of variance between schools in initial status is

$$\frac{\tau_{\beta 00}}{\tau_{\beta 00} + \tau_{\pi 00}} = \frac{19.1}{91.6 + 19.1} = 0.17,$$

and percent of variance between schools in rate of change is

$$\frac{\tau_{\beta 11}}{\tau_{\beta 11} + \tau_{\pi 11}} = \frac{0.5}{0.5 + 0.6} = 0.45.$$

The results above indicate that about 17% of the variance in initial status (achievement scores in Fall 2006) lies between schools. This is consistent with school effects in previous studies where 10% to 30% of the achievement variability was found between schools. The notable percent of variability (45%) in growth rates was found between schools. This indicated that schools differed a lot more in their mean growth rates. School characteristics were included later to investigate what can account for such a big proportion of variability. In this particular application, the variance component decomposition highlighted an important feature of the data: the high percentage of variation in growth rates that lies between schools.

Another approach to examining the within-school and between-school variability is to decompose the correlation between initial status and rate of change into within-school (level 2) and between-school (level 3) components. We found that within a typical school, the estimated correlation between the two growth parameters (initial status and growth rate) is 0.006. Student initial achievement scores have almost no association with their growth rates. The correlation is only slightly stronger at the school level (0.002) indicating a weak correlation between the school mean growth rates and school mean initial status in this sample of students.

The conditional growth model examines how student achievements change over time with the consideration of school contextual characteristics. Six school-level variables are included in the model. Only three of them were found to be significantly related with school mean growth rate. Hence, our final conditional models for rate of change is:

$$\beta_{1ij} = \gamma_{100} + r_{101}(PCT_{FRL}) + r_{102}\left(PCT_{Minority}\right) + r_{103}(Teacher\_Pupil_{Ratio}) + \boldsymbol{b}\left(\boldsymbol{\beta}_{0ij}\right) + \mu_{10j}\ , (10)$$

The random coefficients of this model were almost the same as the ones in the unconditional model, suggesting that adding the three school-level variables and school initial status as predictors to the models does not explain a significant proportion of variability in initial status and growth rates across schools.

Table 3: Three-level unconditional growth model in math achievement across schools

Fixed Effects	**Coefficient**	**Standard Error**	**p value**
Average initial status across schools, γ_{000}	192.9	0.21	.000
Average rate of change across schools,	5.2	0.03	.000

Random Effects	**Variance**	**df**	χ^2	**p value**
Level 1 (scores)	31.4			
Level 2 (between students)				
Individual initial status, r_{0ij}	91.6	41725	187912.3	.000
Individual growth rate, r_{1ij}	0.59	41725	50905.6	.000
Level 3 (between schools)				
School average initial status, μ_{00j}	19.1	475	7517.6	.000
School average growth rate, μ_{10j}	0.47	475	5965.3	.000

Conditional School Growth Model

Table 4 displays the fixed effects of the three-level conditional school growth model. The average school initial score and average school rate of change remain almost the same (192.8 & 5.2). The three selected school variables were all negatively related with school mean growth rates: the higher percentage of FRL or minority students or the higher the teacher pupil ratio in a school, the lower the mean rate of change the school has. The results imply that schools with more students from disadvantaged backgrounds, or schools with fewer resources tend to grow less in their math achievement from grade 3 to 5.

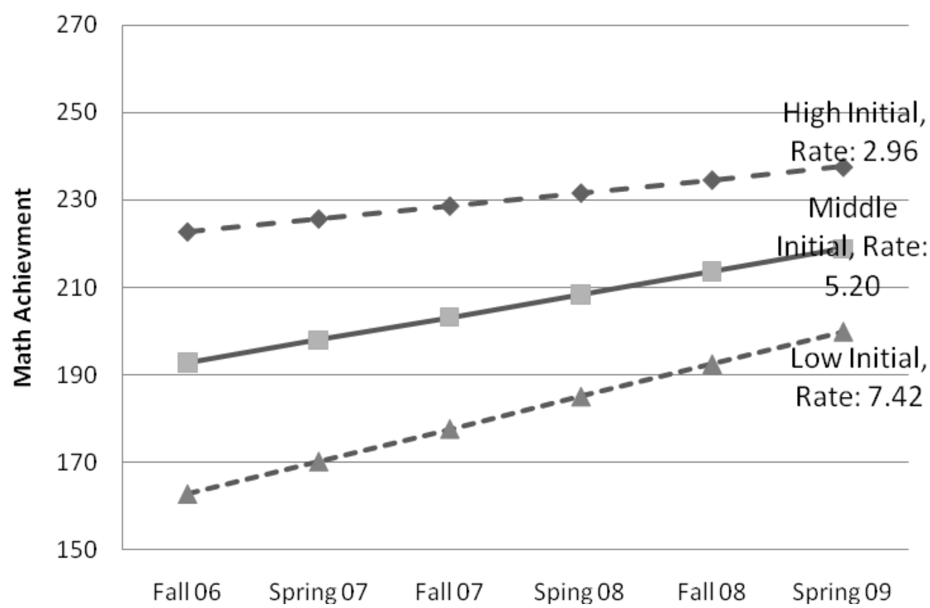
Table 4: Three-level conditional model of growth in math achievement across schools

Fixed Effects	Coefficient	se	p value
Average initial status across schools, γ_{000}	192.8	0.21	.000
Average rate of change across schools, γ_{100}	5.2	0.03	.000
Percent_FRL	-0.012	0.2	.000
Percent_Minority	-0.004	0.002	.015
TeacherPupil_Ratio	-0.037	0.0002	.034
Intercept	-0.068	0.008	.000

Although we don't find strong correlation between initial status and rate of change in the unconditional model, we found that the latent regression coefficient b is a significant predictor in the conditional model (b=.068). It means that when initial status increases by, say 10 points, we expect a decrease in rate of change in mathematics achievement of 10 x (0.068) = 0.68 points per term.

Figure 1 shows that when student achievements in a school start lower they tend to grow more, and that when student achievements start higher they tend to grow less. For example, when schools' mean achievement score was 50 points below the average achievement of all schools in fall 2006, their student achievement scores increased 7.42 points every term. When schools' whose mean achievement score was 50 points above the average achievement of all schools in fall 2006, their student achievement scores increased only 2.96 points per term1.

Figure 1: Conditional school growth trajectories based on different school initial status

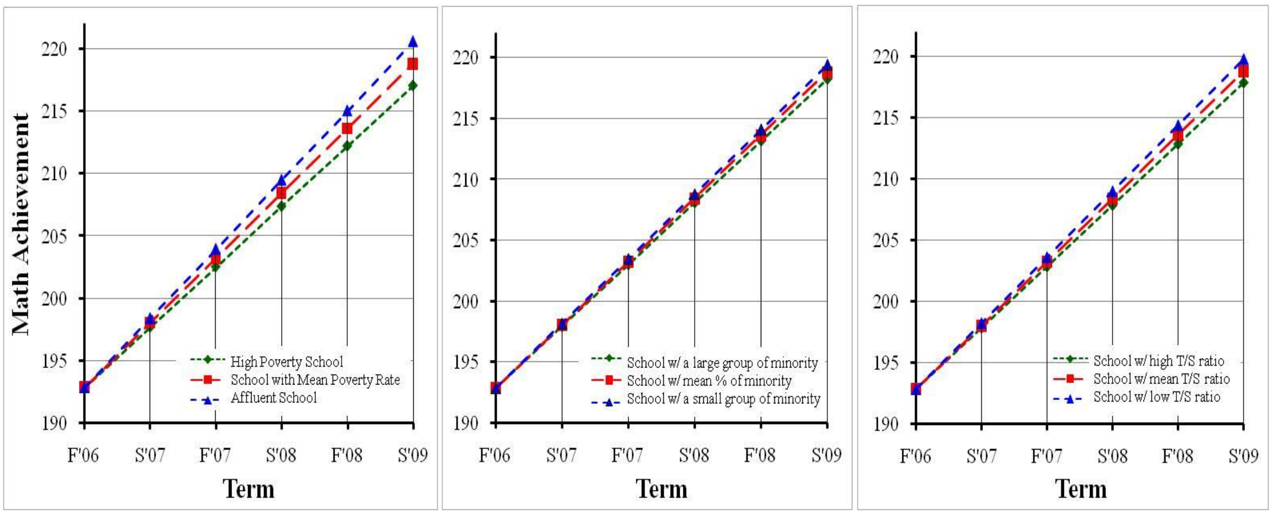


By plotting growth trajectories for different types of schools in the figure below (see Figure 2), we can see that schools don't always grow in a same pattern and growth can still be confounded with other school contextual characteristics even when their students on average perform at the same level at the beginning. For example, the left figure in the panel shows that for a high poverty school (with 89% of FRL students), the school grew 4.84 points per term when a school with 29% of FRL students grew 5.56 points per term2. Thus, even the students in both types of schools start with the same score 192.8 in fall 2006, high-poverty schools tend to grow less and the achievement gaps between high-poverty and low-poverty schools widens across 6 terms/3 years even they started at the same level of achievement (4.32 point gap in spring 2009).

[1] Note calculations can be found in Table 1 in Appendix I.
[2] Note calculations can be found in Table 2, 3, & 4 in Appendix I.

The middle figure in the panel shows that high-minority schools tend to grow less and the achievement gaps between high-minority and low-minority schools widens across 6 terms/3 years even they started at the same level of achievement (1.44 point gap in spring 2009). The right figure in the panel shows that schools with high teacher-pupil ratio schools tend to grow less and the achievement gaps between schools with high teacher-pupil ratio and low teacher-pupil ratio widens across 6 terms/3 years even they started at the same level of achievement (2.28 point gap in spring 2009).

Figure 2: Conditional school growth rate based on different school characteristics



Based on the unconditional model and the conditional model, we ranked all 476 schools, and then compared the difference between the two rankings. Below are the findings:

- One-third of schools ranked differently more than 4 positions, 10% of schools ranked differently more than 10 positions, 5% of schools ranked differently more than 20 positions
- Smaller schools, schools with a large proportion of academically disadvantaged students, and schools with fewer resources tended to rank differently between the unconditional and the conditional models.

The second finding is particularly thought provoking. We often hear questions from school districts: we are a school with a large group of minority students, or a school with limited resources, or a small school in rural area, can we be compared to schools that are similar to us? Our study shows that the request for an apples-to-apples comparison needs to be addressed since disadvantaged schools tend to rank differently in a model that considers school contexts.

What is the implication of this study to individual schools if schools are evaluated solely by their mean growth rates? To illustrate this question, we randomly selected 18 schools out of 468 schools in our sample and ranked them based on their mean rate of change in two models (see Figure 3 & 4). Each bar represents a school's mean rate of change with the upper line and the lower line representing the 95% confidence interval of where the mean rate of change may fall in between. First, we found that three schools (S11, S12 and S16) changed rankings based on their mean rates of change. However, when we also look at their standard error of estimation, we found that the two bars for school no. 11 and school no. 12 mostly overlap. It indicates that even they switch positions based on different models, the change is not significant. Their growth rates do not significantly differ from each other. For school no. 16, it is another story. This school did change its ranking since there is no overlap between its 95% confidence interval and school no. 17's. It tells us that based on the unconditional model school no. 16 had a lower ranking than school no. 17, but after considering school contextual characteristics and initial achievement, this school ranked higher than school no. 17.

Another notable finding is that if you compare the size of the bars in the two figures, you can see that the bar sizes in the second figure are larger. It is because the HLM approach is based on large-sample theory. When a school has a small size, the latent variable regression coefficient in the conditional model reflects more uncertainty about all unknowns in the model (Seltzer, Wong, & Bryk, 1996). That is why we also found that schools with a small student sample tend to rank differently when they are evaluated by the conditional model. It implies that the latent regression conditional model is more sensitive to the cohort size in a school.

Figure 3 and Figure 4 show that some schools rank differently based on the unconditional model and the latent regression conditional model.

Figure 3: School ranking based on the unconditional model

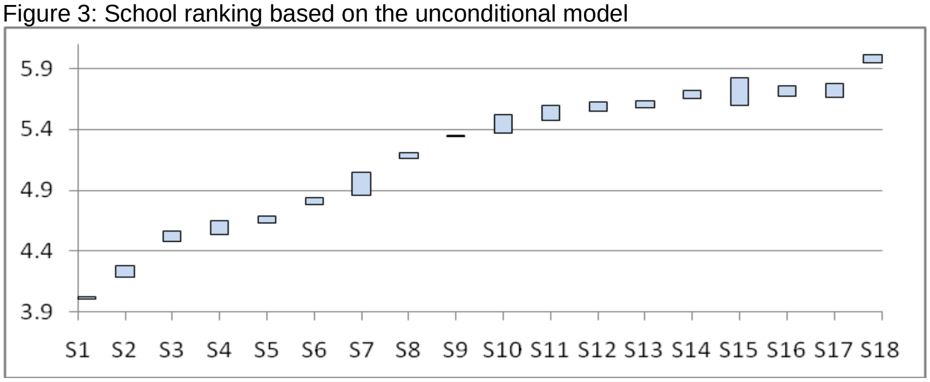
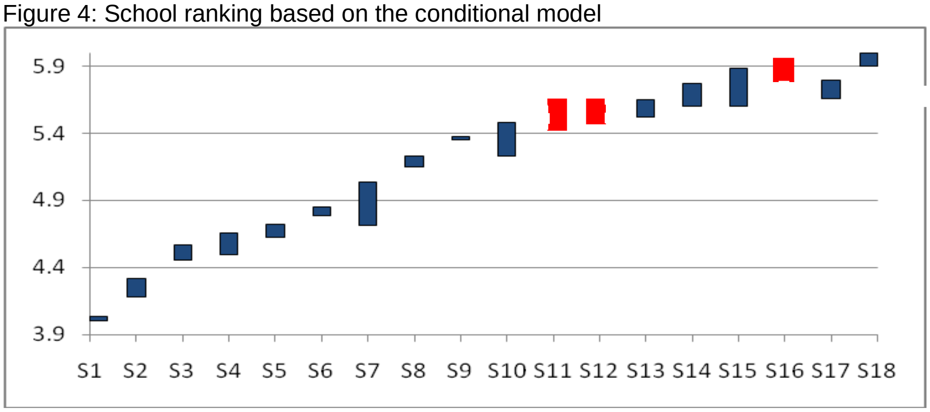


Figure 4: School ranking based on the conditional model



Discussion

Many believe that schools with a large proportion of academically disadvantaged students may be unfairly judged when they fail to make Adequate Yearly Progress (AYP), and it is inevitable that we will be changing the way we measure school performance. One option for consideration is a growth model that emphasizes individual students' progress. Growth models are increasingly seen as a statistically valid and readily understood approach in educational accountability systems. But is growth measure a panacea for our school accountability system? Not necessarily: growth only provides different information than status measures of achievement. The chart below recommend by CCSSO's 2005 report, Policymakers' Guide to Growth Models for School Accountability: How Do Accountability Models Differ, shows a two-dimension matrix (see Table 5) that captures growth and status. In this matrix, schools in Group IV produce both high growth and high status and schools in Group I will be identified as schools need to improve. There are also many schools that have mixed results (in Groups II and III). Only looking at one dimension of the matrix, either status or growth, will rush us to a misleading conclusion about school performance.

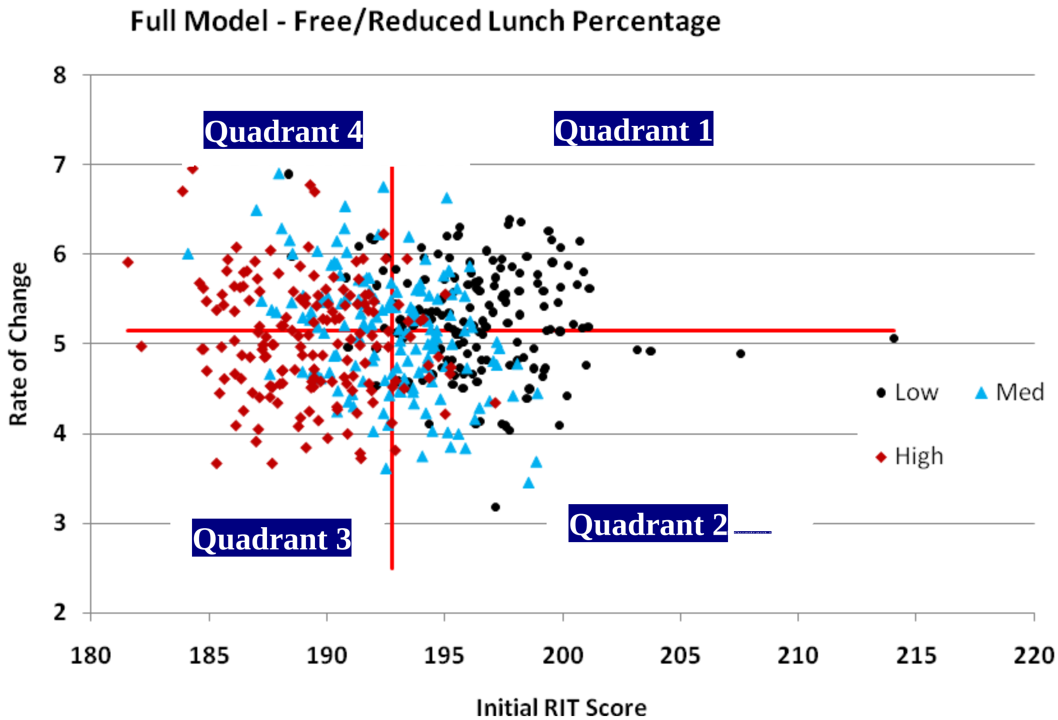
Table 5: Two-dimension matrix of school accountability models (CCSSO, 2005)

High Growth **Group III** **Low Status**	**High Growth** **Group IV** **High Status**
Low Growth **Group I** **Low Status**	**Low Growth** **Group II** **High Status**

When researchers and policy makers start to recommend the two-dimension matrix of initial score by rate of change and begin to evaluate schools based on it, this study suggests a three-dimension perspective that considers school contextual characteristics based on their initial status and rate of change.

Figure 5 is an example of expanding the matrix of initial score by rate of change by adding relevant school factors, in our example, percent of FRL students. The figure shows that schools with high percentage of FRL students tend to cluster in Quadrant 3—low initial status with low growth while schools with low percentage of FRL students mostly cluster in Quadrant 1—high initial status with high growth. It would be misleading if we simply evaluated schools based on their mean rate of change without considering what kind of students attend in a school and what resources a school obtains. The figure shows that growth cannot really eliminate all the confounding factors and make an apples-to-apples comparison between schools. We hope to provide schools and school districts with evaluation modeling tools to make better informed decisions about school performance. Looking at student achievement growth is one step forward, but growth cannot be used as a solitary measure of school effectiveness.

Figure 5: matrix of initial score by rate of change with P Free/Reduced Lunch Percentage



The study is also a demonstration of how schools can be evaluated under the context of a school accountability system in one state. This study can be used to provide a reference of school performance when an amount of growth for a particular school is compared to similar schools in a larger context. School districts, states, or educational funding organizations can use the conditional growth rates to evaluate schools based on the three-dimension matrix shown above.

It is not our intention to show what specific school characteristics should be considered in school evaluation. We recommend that researchers and administrators to explore how schools are different from each other in their system. For example, in one state, schools may differ greatly as to their socio-economic status; in another state, schools may only differ greatly as to the level of students' English language proficiency.

A question may rise as to what is the difference between the latent regression conditional growth model and the value added models (VAM) that are already being applied in some districts and states. First, the two models address different questions. For example, a value-added model can ask the question 'How much student growth can be attributed to a teacher or a school?' By including student characteristics (i.e., ELL status), the model can also ask, 'How much growth by a specific group of students can be attributed to a teacher or a school?' The latent regression conditional growth models ask another type of question, such as 'If a school has a large group of disadvantaged students or has limited resources, how can we take into these factors when we evaluate those schools?' Addressing different questions, the two models then focus on different levels of data. The value-added models more focus on student characteristics while the latent regression conditional models focus on school contextual characteristics. By doing so, the latter models require less data and the data collection process is less intrusive since we can always link student achievement with school variables from the existed Common Core data system. Another important difference is that the value-added models usually deal with one year growth while the school conditional models look at achievement on a longitudinal basis, which gives a relatively stable estimation of student growth. This feature is especially important when we use low-stake assessments to evaluate schools where student motivation or effort is in question and could potentially affect their growth scores in one year.

For our next step, we will apply cross-classified hierarchical models as a solution to the student mobility issue. We may also adopt three other cohorts from different years, grades and subjects to further demonstrate the difference between simply looking at mean rate of change and examining school rate of change with their contextual characteristics.

After investigating how to precisely estimate school mean growth rate, we hope to focus on the variability of student achievement within schools, including looking at within-school achievement gap. Achievement growth gaps between high performing and low performing students can be included into the accountability model as another important indicator of school performance.

REFERENCES

Braun, H. J., Wang, A., Jenkins, F., & Weinbaum, E. (2006). The Black-White achievement gap: Do state policies matter? Education Policy Analysis Archives, 14(8). Retrieved [4-20-06] http://epaa.asu.edu/epaa/v14n8/.

Carlson, D. (2001). Focusing state educational accountability systems: Four methods for judging school quality and progress. (Retrieved June 8th, 2009 from http://www.nciea.org)

Council of Chief State School Officers. (2005). Policymakers' Guide to Growth Models for School Accountability: How Do Accountability Models Differ? Retrieved [4-15-09] http://www.ccsso.org/projects/Accountabilty_Systems/.

Hauser, C., & Kingsbury, G.G. (2007). Using growth models and a large longitudinal database to predict individual student achievement: A modest demonstration. Paper presented at the 2007 Annual Meeting of the American Educational Research Association in Chicago, IL.

Kane, T. J., & Staiger, D. O. (2002). Volatility in school test scores: Implications for test-based accountability systems. Brooking Papers on Educational Policy, 1, 235-283.

Linn, R. L. (2000). Assessments and accountability. Educational Researcher, 29(2), 4-16

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. Educational Researcher, 32(7), 3-13.

Linn, R. L. (2004). Rethinking the No Child Left Behind accountability system. (Paper prepared for a forum on No Child Left Behind sponsored by the Center on Education Policy, Washington DC, July 28, 2004).

Lion, L. G. (2000). Standardized tests: The interpretation of racial and ethnic gaps. La Griffe du Lion, 2(3). Retrieved September 23, 2007 from http://www.lagriffedulion.f2s.com/.

Lion, L. G. (2004). Closing the racial learning gap. La Griffe du Lion, 6(1). Retrieved September 23, 2007 from http://www.lagriffedulion.f2s.com/.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. Sociology of Education, 59, 1-17.

Seltzer, M., Wong, W. & Bryk, A. (1996). Bayesian analysis in applications of hierarchical models: issues and methods. Journal of Educational and Behavioral Statistics, 18, 131-167.

Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. Educational Evaluation and Policy Analysis, 25, 263-286.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ESEA Blueprint for Reform, Washington, D.C., 2010.

Willet, J. B. (1988). Questions and answers in the measurement of change. In E. Rothkopf (Ed.), Review of research in education (1988-1989) (pp.345-422). Washington, DC: American Educational Research Association.

APPENDIX I

Table 1: schools have different initial status

School Type	Mean Initial Score	Rate of change (per term)
High Mean Initial Status	222.8	5.2+30*(-.068)=7.44
High Mean Initial Status	192.8	5.2
High Mean Initial Status	162.8	5.2-30*(-.068)=2.96

Table 2: schools have different percentages of FRL students

School Type	FRL %	Rate of change (per term)
High-poverty School	89%	5.2-30*(-.012)=4.84
Mid-poverty School	59%	5.2
Low-poverty School	29%	5.2-30*(-.012)=5.56

Table 3: schools have different percentages of Minority students

School Type	Minority %	Rate of change (per term)
High-minority School	79%	5.2+30*(-.004)=5.08
Mid-minority School	49%	5.2
Low-minority School	19%	5.2-30*(-.004)=5.32

Table 4: schools have different teacher student ratios

School Type	Teacher-student ratio	Rate of change (per term)
Low Teacher-pupil Ratio	10	5.2+5*(-.037)=5.01
Middle Teacher-pupil Ratio	15	5.2
High Teacher-pupil Ratio	20	5.2-5*(-.037)=5.39