Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 10 Number 11, August 2005 ISSN 1531-7714

Growth Models in Action: Selected Case Studies

Jessica Goldstein and Peter Behuniak
University of Connecticut

This paper provides an overview of the issues related to the use of growth to model student achievement. Case studies from nine states and two cities are presented to demonstrate the diversity in the definitions of growth across the nation as well as the complexities of incorporating growth into pre-existing state systems. Several suggestions for future research are also provided

In recent years, many in the educational measurement community have argued for the use of longitudinal analyses to account for changes in student achievement (Barton & Coley, 1998; Bryk & Raudenbush, 1988; Linn & Haug, 2002; Stevens, Estrada, & Parkes, 2000). This year, U.S. Department of Education Secretary Margaret Spellings announced that she will convene a panel to discuss the use of growth models to measure student achievement. In systems that include growth, schools receive credit for making progress towards a goal. This idea has great appeal for educators as it rewards schools for impacting students, even if the school does not make adequate yearly progress. Growth models contrast with the "status model" mandated by the No Child Left Behind legislation, in which the unadjusted percentage of students who have reached proficiency in a given grade is compared to pre-determined annual targets, set by the state.

Growth models are preferred by the educational measurement community for a number of reasons. One key issue is the accuracy of the estimates gleaned from these models. Linn and Haug (2002) suggest that status models that rely on the comparison of successive cohorts (the performance of third grade students in the current year compared to the performance of third grade students the prior year) rests on the implicit assumption that student characteristics in a given school remain stagnant from year to year. Yet it is known that variability in initial group differences and student mobility result in random data fluctuations that may cause certain schools to be inaccurately recognized as outstanding or in need of improvement (Carlson, 2003; Goldstein, 1991). There is little value in holding schools accountable for mean achievement levels when students enter schools with differing levels of mean achievement (Ballou, Sanders, & Wright, 2004).

Another key issue is whether it is more appropriate to use performance standards or growth to measure changes in student achievement. While it can be argued that current status measures offer the advantage that all students and schools are universally held to the same standards, measures of growth are less likely to be influenced by sociodemographic variables and, as a result, they may be more appropriate for schools that serve students with low initial achievement levels (Linn,

2001; Linn, 2005; Stevens, Estrada, & Parkes, 2000). Further, Linn (2005) suggests that a large percentage of proficient students within a school does not necessarily lead to the conclusion that high achievement is the result of good instruction as it may also be the result of prior achievement or other relevant characteristics of the student body. This conflict forms the foundation of the controversy surrounding the consideration of growth in accountability models. One other issue worth noting is that the requirements for annual testing in grades 3 through 8 by 2005-2006 explicated in NCLB will finally provide the infrastructure for states to conduct these analyses as part of their accountability systems.

It is clear that growth models are receiving increased attention at the state and national levels as well as from the measurement community. However, there is relatively little documentation of the growth models that have been implemented in state educational systems. The purpose of the present review is to provide such documentation by examining the use of growth in selected state and city systems including: Tennessee, Dallas, North Carolina, New Mexico, Louisiana, Kentucky, California, Ohio, Chicago, and Utah. A general description of each state's model is presented, followed by a discussion of its impact on the state.

TENNESSEE VALUE-ADDED ACCOUNTABILITY SYSTEM

Description

The Tennessee Value-Added Accountability System (TVAAS) is one of the most widely recognized value-added systems and the first to be used for an entire state. Developed by Dr. William Sanders, then a statistician at the University of Tennessee-Knoxville, the system allows for the analysis of data on student outcomes over time at the individual student level. Data is collected annually for students in grades three through eight in five subject areas: math, reading, language arts, social studies, and science. Using mixed model equations, TVAAS uses the covariance matrix from this multivariate, longitudinal data set to evaluate the impact of the educational system on student progress in comparison to national norms, with data reports at the district, school, and teacher levels (Sanders, 2003). The system requires a proprietary piece of software developed from the SAS PROC MIXED procedure.

In the model, each student acts as his or her own control and no other covariates are used. Sanders, Saxton, and Horn (1997) purport that the high correlation between pretest and posttest scores sufficiently accounts for non-school related variables and further, that the task of collecting all the relevant covariates is a "hopeless impossibility." Additionally, they assert that their own previous research has demonstrated that school, system, and teacher effects can be reasonably estimated without the inclusion of data pertaining to these student-level variables. The model also assumes that a teacher's effect on student achievement persists through the years that follow their time together. The inherent advantage of the model, as suggested by the authors, is that it focuses on academic gains rather than raw achievement scores, directing accountability at student improvement rather than on absolute levels of achievement.

Discussion

In 1995, the Tennessee Office of Accountability set forth on an effort to evaluate the effectiveness of the system (see Baker & Xu, 1995). The report presented a number of criticisms. First, the national norms against which schools and teachers are judged require uneven gains from students across grade levels and between subjects; i.e., students in lower grades are expected to have greater gains than students in higher grades. Still, teachers, schools, and districts were all required to meet these gain levels to avoid sanctions. Additionally, school and even district-level results varied widely from year to year without explanation. Finally, there were large changes in value-added scores from year to year within districts and schools. Ultimately, the authors were unable to explore the system in enough detail to render a judgment of its effectiveness:

> "*Additional evaluation of the value-added assessment model might lay to rest many of the questions and concerns people have raised about the TVAAS—but, as yet, no such comprehensive evaluation has been performed. Both statisticians and educational measurement experts need the opportunity to test the model. Without further evaluation, the state—and its educational practitioners—cannot determine the*

> *validity of the value-added model*" (Baker & Xu, 1995, p. 9).

They further emphasized the importance of widely available documentation of the system for the state's constituents as well as the need for efforts to be made to clarify the system to teachers, principals, and superintendents. Ten years later, there is little published documentation of the TVAAS model.

In Sanders' rebuttal to the assertions in the evaluation (see Baker & Xu, 1995, Appendix G), he explored the issue of uneven scores across years within schools and districts. He suggested that changes in the educational program and instruction that is inconsistent with prior levels of achievement can affect value added scores. Some teachers may receive assistance within their schools as a result of scores from TVAAS. Assuming this assistance is effective, students would enter the following grade with higher prior levels of achievement than students from the previous year. Unless the receiving grade's teachers changed their instructional strategies, these students' gain scores would not be as high as students from the previous year, because they entered with higher levels of prior achievement. Instead of being a fault of the system, Sanders asserts, these scores are actually evidence of the functionality of the system. Evidence to demonstrate this phenomenon is not provided in his rebuttal.

The future of the TVAAS is no longer certain in the state of Tennessee. In 2003, TVAAS results were removed from the state's accountability criteria because the No Child Left Behind law required the use of standards-based assessments, rather than norm-reference assessments to determine school effectiveness. In 2004, the Tennessee House Education committee considered a bill to eliminate the program, but it did not make it past the committee after positive testimony from Dr. Sanders and other superintendents (Commins, 2004). However, the TVAAS results are still in use for diagnostic purposes on the local level and for teachers to demonstrate they meet the "highly qualified teacher" standard explicated in NCLB. Recent discussion of the system has shown that data from TVAAS helps teachers quickly understand their students' histories but can penalize schools with higher performing students. One principal explained, "The problem is when you get to the upper end of the scale, the more difficult it is to meet the standard that shows one year's growth. You can have a student scoring in the 90 percentile miss one question more than he did the year before and that could knock him below a year's growth" (Carroll, 2005, p. 1). Further, the Tennessee State Department of Education and Dr. Sanders are discussing the conversion of TVAAS to use criterion-referenced test data (Morgan, 2004). Dr. Sanders is currently employed by SAS, which produces a software product entitled SAS EVAAS designed to evaluate student progress over time.

In the fall of 2002, the commonwealth of Pennsylvania began work with SAS to create the Pennsylvania Value-Added Assessment System (Pennsylvania Department of Education, n.d.). Pilot data collection will be collected through fall 2006 and comprehensive reporting will begin with the fall 2007 administration. The PVAAS will be used in conjunction with existing student data reporting systems and will not be used for teacher accountability. Student-teacher linkages are not maintained in the PVAAS system.

DALLAS INDEPENDENT SCHOOL DISTRICT

Like the TVAAS, the Dallas Independent School District (DISD) also tracks students longitudinally but uses a two-stage HLM analysis as to measure student growth.

Description

The value-added accountability system in place in the DISD is rooted in a strong philosophy of community consensus and fairness to all stakeholders in the educational system. Essentially, the system compares a student's observed test score to the prediction of how well that student was expected to achieve, given their performance the previous year. The threshold for effectiveness is defined by an Accountability Task Force, comprised of stakeholders including parents, teachers, and representatives from the community and local businesses, who decide on the selection and weighting of variables used within the system and the performance awards associated with it (Webster & Mendro, 1997). Thum and Bryk (1997) suggested that the buy-in from community stakeholders in the

DISD system has contributed strongly to its longevity. The goal of the Task Force is to ensure that schools do not derive unfair advantages from starting with high or low-performing students, minority or white students, high or low-socioeconomic level students, or limited English or non-limited English proficient students (Webster, 2004). While many variables are used in determining school effectiveness including attendance rates, dropout rates, retention rates, enrollment in honors courses, graduation rates, and the percentage of students taking college entrance tests, test scores on the Iowa Test of Basic Skills and the Texas Assessment of Academic Skills are weighted most heavily. Schools are only required to test students who have been enrolled in the school by the end of the first six weeks of the academic year and remain at that school until the end of the year.

Webster and Mendro (1997) provided a detailed description of the two-stage approach. At the first stage, three variables are used to control for fairness: a dichotomous gender variable, a dichotomous free-lunch status variable, and combination ethnicity/language proficiency variable with four levels: black, Hispanic, limited-English proficient, and other. These variables are collectively referred to as "fairness variables" are regressed on both pretest and posttest measures to demonstrate that there is no bias at the student level. The residuals from these regressions are then referred to as the Fairness Adjusted Pre-Scores for the prior year's scores and the Fairness Adjusted Post-scores for the current year's scores and are used all subsequent analyses. These residuals are an estimate of a student's deviation from what was expected and therefore form the basis for the judgment of the "value added" by the teacher or by the school. Next, regression is used to determine which Fairness Adjusted Pre-Score(s) are the best predictors of each Fairness Adjusted Post-Score. After the fairness variables have been accounted for, Hierarchical Linear Modeling is used in a second stage of analysis to control for school level variables such as mobility, crowding, percentage minority and SES. The residuals from the first stage are the outcome variables for each test in each grade. Value-added gains are then standardized to a mean of 50 and a standard deviation of 10; gains for students with excessive absences are not included.

Discussion

Thum and Bryk (1997) offered a critique of the DISD framework. First, they proposed that standardizing the residuals leads to biased estimates for teachers and/or schools who serve a disproportionate number of "deviating" subgroups. Additionally, they suggested that the two-stage approach is not necessary and that controlling for the student-level fairness variables can be accomplished in a multilevel HLM framework. Most models of school effects consider only the intercept term as random and all of the coefficients are fixed with their corresponding predictors grand mean centered (Thum & Bryk, 1997). Webster (2004), however, offered that the two-stage model was implemented because it reflected the values of the Accountability Task Force, and not for its statistical parsimony. Finally, Thum and Bryk advised that the use of pretest score as a covariate is problematic in that analysts cannot prove that all relevant background variables have been considered and controlled for in the analysis and further, that the use of residuals from previous regression analyses does not control for this problem. A more appropriate measure would be to use students' gain scores where each student acts as his or her own control in the analysis as is done in the Tennessee Value-Added Accountability System.

It is important to note that Texas also uses a status based approach to accountability based on performance on the TAAS and dropout rate with four levels: Exemplary, Recognized, Academically Acceptable, and Unacceptable or Low Performance. Webster (2004) demonstrated that of the top 25 Dallas elementary schools in the value-added system, only those with a high percentage of white students, a low percentage of black students, and a relatively low deprivation index (a measure of SES) ranked in the top 25 schools in the state system.

NORTH CAROLINA'S ABC'S OF PUBLIC EDUCATION

In North Carolina, groups of students are tracked longitudinally and achievement is calculated using the residual regression approach, as described below.

Description

The North Carolina State Board of Education created its accountability system, the ABCs of Public Education, in the late 1990's. The ABCs accountability program employs End-of-Course (EOC) and End-of-Grade (EOG) examinations in a residual regression approach to measure both growth and performance at the school level (North Carolina Department of Public Instruction [NCDPI], 2002). The system is quasi-longitudinal (Linn, 2001) in that students' aggregated scores in a given year are compared to aggregated scores from the same group of students in the earlier grade from the previous year; in other words growth is measured by the comparison of scores from all fourth grade students in 2004-2005 to scores from all third grade students in 2003-2004. Students are tested annually and scores are equated on a vertical scale to correspond with student development. In the system, a year of schooling is expected to produce a year of growth.

The amount of growth a school is expected to make is calculated using simple linear regression based on the North Carolina average rate of growth (b_0), an estimate of the true proficiency of the students in the school (b_1 * Index for True Proficiency [ITP]), and an estimate for variability in scores due to regression to the mean (b_2 * Index for Regression to the Mean [IRM]) (NCDPI, 2004a). Thus, the formula for calculating expected growth is: Expected Growth = b_0 + (b_1 * ITP) + (b_2 * IRM). The Average Rate of Growth (b_0) is determined using data from a longitudinal study of the actual growth rates of students in the system that were followed from 1992-1993 to 1993-1994 at each grade level beginning in third grade. A baseline for third grade growth was determined using a pre-test at the beginning of the 2000-2001 year. These growth rates have been used in analyses for every subsequent year; i.e. current analyses of growth are always conducted in reference to students' growth from the spring of 1993 to the spring of 1994. Values for b_1 and b_2 coefficients are calculated using data from grades four through eight from the 1994-1995 year and from third grade students in the 2000-2001 year. To calculate the ITP, both the reading and mathematics scale scores for the end-of-grade tests are added together to give a total overall score. The index for true proficiency (ITP) is computed by subtracting the North Carolina averages (from the 1994-1995 year) from the local test scores. The IRM is computed by subtracting the approved North Carolina averages from the local test scores (reading and mathematics respectively).

Expected Growth can be achieved when a school's students meet 100% of their expected growth based on all of the following parameters that can apply to a given school:

- Growth in EOG reading and mathematics for grades 3-8,
- Growth based on EOC tests,
- Change over a two-year baseline in the percent of students completing certain courses of study (college university prep/college tech prep),
- Change in the competency passing rate (from grade 8 to grade 10)
- Change in the ABCs dropout rate (compared to a two-year baseline) (NCDPI, 2004a).

Achievement Level III is used as the baseline for proficiency and is defined as students who "consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level" (NCDPI, 2002, p. 29). Additionally, the system allows for ten demarcations of schools based on the growth, performance, and attainment of Adequate Yearly Progress; financial awards are given to certified staff members in schools where students achieve high or expected growth standards (NCDPI, 2004a).

Discussion

There were 2232 public schools that were assigned an ABC status in the 2003-2004 academic year. In all, 75.1% of the schools met either expected or high growth standards and 70.8% of schools made Adequate Yearly Progress (NCDPI, 2004b). This high proficiency rate has led to some discussion of the need for an overhaul of the testing system (Waggoner, 2005). The challenge will be to make it more demanding while maintaining attainable standards for Adequate Yearly Progress and without losing the support of teachers whose

bonuses are tied to score increases. Still, the current system is not without controversy. There was an issue with data from 2003-2004 school year. In that year, only two of the state's 388 middle schools had achieved the progress expected for sixth-graders in reading. Although teachers, administrators, and a state Board of Education advisory committee on testing issues questioned the statistical accuracy of the formula used to measure performance and suggested that accountability measures be recalculated without the sixth grade data, the Board of Education decided not to change its system (Keung Hui, 2004). The specific issues the stakeholders had with the statistical accuracy of the formula are unclear. While the state does not provide analyses by socioeconomic status or other demographic variables, an analysis of data from 37,000 fifth grade students by Ladd and Walsh (2002) found that schools that serve a disproportionately high proportion of high income and white students fare better under the system than schools serving students from disadvantaged or racially diverse backgrounds and suggest that teachers serving in higher income and less diverse schools have a higher probability of earning a bonus based on student performance.

NEW MEXICO

Description

Prior to NCLB, New Mexico used norm-referenced achievement tests in reading, language arts, mathematics, science, and social studies to measure both student status and growth (New Mexico State Board of Education [NMBE], 2002). Growth was determined as the change in scale score for a cohort of students (followed in a similar manner as students in North Carolina, a quasi-longitudinal approach) and was analyzed in comparison to "typical growth" as defined for each subject and each grade by the test publisher and NMBE. Performance standards were set based on the median percentile of students' scores and the rate of growth from one year to the next. Higher growth was expected of lower performing schools. For example, in a school where less than 40% of the students were at the median percentile, a growth rate of 1.75 times the typical growth rate was required to achieve the exemplary standard; if 60% of the students were below the median percentile a growth rate of 1.25 times the typical growth rate was required to achieve the exemplary standard. The accountability system also incorporated status scores for tenth grade students as well as school attendance and dropout indicators.

Discussion

This model is no longer in use. The transition to the NCLB model of accountability, including the use of criterion-referenced testing, began in the 2002-2003 school year (New Mexico State Department of Education, 2003). As of the 2004-2005 school year, all students in grades three through nine and eleven took the criterion-referenced tests. Assistant Secretary of the Accountability and Assessment Division Dr. Don Watson (personal communication, April 19, 2005) affirms while there is interest in the use of growth models, such a plan will not be implemented until there is adequate data from Standards Based Assessments.

CHICAGO PUBLIC SCHOOLS SCHOOL ACCOUNTABILITY SYSTEM

The accountability system in Chicago is of interest because it employs a blend of growth models to track achievement. Like in New Mexico, growth at the school level is compared to a standard while, at the student level, gain scores are used to evaluate student progress.

Description

Introduced in 1996, the Chicago Public Schools (CPS) School Accountability System evaluates both school progress and student growth over time using four elementary level measures: the Iowa Test of Basic Skills (ITBS) Total School Growth, ITBS Average Student Gain, Illinois Student Achievement Test (ISAT) Total School Growth, and Adequate Yearly Progress. Total School Growth is calculated by subtracting the current year's composite school average from the average of the previous three years and then comparing that average to citywide average growth (CPS, n.d.,a).

The system used in Chicago is in direct contrast to the North Carolina system, which uses the 1992-1993 year as a baseline and the former system in New Mexico, which compared growth to targets set

by local administrators and the test publisher. In Chicago, Average Student Gain is calculated as the scale score difference between a student's current and previous year's scores. These differences are averaged by grade and then converted to an index score based on comparisons to the city and national average. The index yields a value of 1.00 when a school's gain scores are equivalent to the city's average gain scores.

The four measures and the four performance standards used to define Progress Levels are detailed in Table 1, below (CPS, n.d.,b). These Progress Levels are used in combination with the percent of proficient students reading at or above national norms on the ITBS or percent of students meeting or exceeding standards on the ISAT to label schools as Schools of Distinction, Schools of Excellence, Schools of Merit, Schools of Opportunity, Schools on Probation, or Schools of Challenge. High school level accountability measures include dropout rates, progress towards graduation, school growth index on the Prairie State Achievement Examination (PSAE), and AYP.

Table 1: ***Overview of the Chicago Public Schools Accountability System***

		Performance Standards		
Test	Description	Exceeds	Meets	Does Not Meet
ITBS School Growth	Measures growth in the % of students at or above national norms on the ITBS	2.0+ percentage points above citywide growth	-2.0 to 1.9 percentage points	-2.0+ percentage points below citywide growth
ITBS Average Student Gain	Measures the average one-year gain in ITBS reading	1.10+ (Higher than normed gains)	.0 - .109 (Equal to normed gains)	<1.10 (Lower than normed gains)
ISAT School Growth	Compares change in % of students meeting state standards on the ISAT to citywide growth	2.0+ percentage points above citywide growth	-2.0 to 1.9 percentage points	-2.0+ percentage points below citywide growth
Adequate Yearly Progress	State evaluation based on established annual targets	Made AYP and not on the School Improvement List	Made AYP and on the School Improvement List	Did not make AYP and on the School Improvement List

Discussion

When administrators in Chicago revised the accountability system in mid-to-late nineties, their aim was to end social promotion and encourage school and teacher accountability for student achievement. Under the system, students in third, sixth and eighth grades are required to meet minimum standards in reading and mathematics on the Iowa Test of Basic Skills (ITBS) in order to advance to the next grade and schools in which fewer than 15% of students scored at or above national norms were placed on probation (Jacob, 2005). A provision exists for the reconstitution of schools on probation that do not exhibit sufficient improvement, which involves the reassignment of teachers and school administrators. In an analysis of successive cohort data from 1993 to 2000, Jacob demonstrated that the new accountability policy led to substantial increases in reading and math achievement on the state test, with gains of .20-.30 standard deviations. Smaller gains were seen with younger students. As of September, 2004, the CPS system served 426,812 students in 486 elementary schools and 107 high schools (CPS, n.d.,c).

LOUISIANA ACCOUNTABILITY SYSTEM

More common models of accountability than the models described above are those that track change in successive cohorts. Since NCLB, a number of state systems have been revised to incorporate annual measurable objectives (AMOs) into their reporting. Louisiana is an example of such a system.

Description

The state of Louisiana uses both growth and performance in its accountability system, but each measure is considered separately (Louisiana Administrative Code, Title 28, 2005). As with Chicago, both norm and criterion-referenced tests are used. The criterion reference tests, known as the Louisiana Educational Assessment Program for the 21st Century (LEAP 21) and the Graduation Exit Examination for the 21st Century (GEE 21), are administered in grades four, eight, ten, and eleven respectively; the Iowa tests are administered in all other grades beginning in grade three and ending in grade nine. Students must pass the LEAP 21 in grades four and eight to be considered for advancement to the next grade. An index is calculated for the CRT based on the sum of all students' total scores divided by the number of students taking each test. For the NRT, a school's score is calculated as the average of its students' scores.

School Performance Scores (SPS) are calculated using a weighted average of students' performance on the state's criterion-referenced test (60%), the norm-referenced Iowa Tests (30%), student attendance in grades K-12 (10% for K-6, 5% for 7-12), and student dropout rates (5% for 7-12). Growth is measured over a two-year period and targets are set at the amount of growth a school must make each year to reach a target score of 120 in 2014 for grades four, eight, and ten. These targets are adjusted according to the percent of special education students and limited English proficient students; growth targets can vary from 2.0 to 10.0 points per year. Growth and Performance Labels are assigned to the schools each year (see Tables 2 and 3, below). School Improvement (SI) services, which range from directive assistance to reconstitution, are offered to schools labeled Academically Unacceptable and/or schools that do not meet their growth targets.

Table 2: ***Growth Labels in Louisiana in Use Since 2004***

Growth Label	Description
Exemplary Academic Growth	A school that makes its growth target, all subgroups grow at least two points, and the school is not in SI.
Recognized Academic Growth	A school that makes its growth target but any subgroup does not grow at least two points and/or the school is in SI.
Minimal Academic Growth	A school improving (at least 0.1 points) but not meeting its growth target.
No Growth	A school with a change in SPS of 0 to –2.5 points.
School In Decline	A school with a declining SPS (more than –2.5 points).

Table 3: *Performance Labels in Louisiana*

Performance Label	Description
Academically Unacceptable	Below 45.0
Academic Warning	45.0 – 59.9
*	60.0 – 79.9
* *	80.0 – 99.9
* * *	100.0 – 119.9
* * * *	120.0 – 139.9
* * * * *	140.0 and above

Discussion

In the 2003-2004 school year, 5.7% of all schools (elementary, middle, and high schools) did not make Adequate Yearly Progress while 21.2% were labeled as Schools in Decline, 12.7% as No Growth, 19.2% as Minimal Academic Growth, 15.6% as Recognized Academic Growth, and 30.1% as Exemplary Academic Growth (Louisiana Department of Education, 2004). Seventy-eight schools scored below 45, the threshold for Academically Unacceptable in 2004, compared to 182 schools in 1999. In 2006, the state of Louisiana will replace the Iowa tests with an extended version of the criterion-referenced LEAP, called the iLEAP, in grades three, five, six, seven and nine (Sentell, 2004).

KENTUCKY ACCOUNTABILITY INDEX

As in Chicago and Louisiana, Kentucky's Accountability Index (KAI) is calculated from a combination of criterion and norm-referenced tests. Growth is tracked by the attainment of annual measurable objectives.

Description

In response to Kentucky Supreme Court ruling in 1989 pertaining to school financing, the Kentucky General Assembly passed the Kentucky Education Reform Act of 1990 (KERA) that seeks to provide all students in the state with an adequate education. A cornerstone of the legislation is an accountability system that grants decision-making authority to local schools, which are held responsible for meeting statewide standards-based goals (Roeder, 1999). A large increase in state taxes allowed for a structure where schools exceeding those goals are offered financial rewards while schools that fall below those goals are offered technical assistance.

The Commonwealth Accountability Testing System (CATS) measures school effectiveness via the Kentucky Accountability Index (KAI), a numeric composite reflecting student performance with reference to Kentucky Performance Standards, a nationally norm-referenced test (CTBS/5 Survey Edition), and nonacademic indicators such as attendance rate, retention rate, dropout rate, and "transition to adult life." Transition to adult life is assessed annually via a survey that measures plans to enter college, the military, or an alternative vocation. The academic component of the KAI is calculated as a weighted average of the percentages of students scoring at each level of the standard (Novice, Apprentice, Proficient, and Distinguished) in reading, writing, mathematics, science, social studies, arts and humanities, practical living and vocational studies (Kentucky Department of Education [KDE], 2004). The grades in which each subject is tested are detailed in Table 4. A key feature of the CATS is the use of open-ended response questions to distinguish between Proficient and Distinguished performance standards. A separate "academic index" is calculated for each content area and those indices are combined with data from the other indicators using a weighted average formula unique to each level of schooling (elementary, middle, and high). In that

formula, scores from the norm-referenced test account for only 5% of the total score. Much like the system in Louisiana, Kentucky's Accountability Index measures progress towards the goal of total Proficiency by the year 2014 in two-year increments. Schools are labeled as Meeting Goal, Progressing, or Assistance.

Table 4: *Number of Test Items by Content and Grade*

	Kentucky Core Content Test							Portfolio	
Grade	Reading	Math	Science	Social Studies	On-Demand Writing	Arts & Hum	PL/VS	Writing	Alternate Portfolio
4	6 OR*		6 OR		X*			X	X
	24 MC*		24 MC						
5		6 OR		6 OR		2 OR	2 OR		
		24 MC		24 MC		8 MC	8 MC		
7	6 OR		6 OR		X			X	
	24 MC		24 MC						
8		6 OR		6 OR		2 OR	2 OR		X
		24 MC		24 MC		8 MC	8 MC		
10	6 OR						2 OR		
	24 MC						8 MC		
11		6 OR	6 OR	6 OR		2 OR			
		24 MC	24 MC	24 MC		8 MC			
12				/	X			X	X

* OR denotes Open Response; MC denotes Multiple Choice; X denotes On-Demand Writing or Writing Portfolio.

Discussion

As of April, 2005, Kentucky legislators were considering changes to the accountability system; among them is a component that will allow educators to track individual student's progress over time (Biesk, 2005). The state Senate also voted to allow select school districts to develop alternatives to the statewide test that gauges school performance (Schreiner, 2005). It is believed that these pilot projects that could potentially improve the overall state system. In the two-year cycle ending in 2004, 661 schools met their accountability goals, 467 schools finished in the progressing area and 48 schools needed some form of assistance; 76% of the state's schools made AYP in 2004.

CALIFORNIA'S ACADEMIC PERFORMANCE INDEX

California's accountability system is quite similar to the systems in Louisiana and Kentucky: an index is calculated at the school level and growth is measured using annual measurable objectives.

Description

The Academic Performance Index (API) was established by California's Public Schools Accountability Act (PSAA) of 1999 (California Department of Education [CDE], 2005). The index reflects the growth of a school's academic performance based on annual results of statewide testing. Each year, every public school receives an API Base report (released in the beginning of the calendar year) that indicates the school's current level of academic performance and sets growth targets as well as an API Growth report (released in the fall) that shows how well those targets were met.

The API Base report is calculated with results of the prior year's Standardized Testing and Reporting (STAR) Program and California High School Exit Examination (CAHSEE). The STAR Program includes the California Standards Tests (CSTs), the norm-referenced California Achievement Tests, Sixth Edition Survey (CAT/6 Survey), and the California Alternate Performance Assessment (CAPA).

Currently, the state is reporting both scores as it changes the system of indicators used to calculate the API. Like Louisiana and Kentucky, the index is calculated as a weighted average of students' performance scores, ranging from 200 to 1000. A separate weighting scheme is used to calculate the index for grades two through six, seven through eight, and nine through eleven. Similarly, growth is defined as 5% of the distance from the school's API to the statewide target of 800 or a minimum of one point of growth. Growth targets are set both for the school and for subgroups within the school, which include African American or Black, American Indian or Alaskan Native, Asian, Filipino, Hispanic or Latino, Pacific Islander, White, and Socioeconomically Disadvantaged. The Socioeconomically Disadvantaged subgroup includes students whose both parents have not received a high school diploma or a student who participates in the free or reduced-price lunch program.

Plans exist to provide awards to schools that meet or exceed their targets; however, funds are not presently available to do so. The API Base report includes both a Statewide Rank and a Similar Schools Rank. Statewide Ranks are calculated by ordering the state's schools by API scores and then dividing them into ten equal groups. Similar Schools Ranks offer a comparison of a given school to other schools of similar demographic characteristics (based on measures of pupil mobility, pupil ethnicity, pupil socioeconomic status as defined by parental education and free or reduced-price lunch status, percentage of teachers who are fully credentialed, percentage of teachers who hold emergency credentials, percentage of pupils who are English learners, average class size per grade level, and whether the school operates a multi-track year-round educational program). Alternative Accountability Systems also exist for small schools (with 11-99 valid test scores), special education schools, and schools that serve a majority of high-risk students.

Discussion

A number of researchers have investigated California's accountability system. Thum (2002) points out that the methodology used to calculate the API is not valid for schools with scores approaching 800 and that the standard errors of the estimates can be quite large and variable from one year to the next. Powers (2003) analyzed 1999 and 2000 API scores for two large urban districts and found that 75% of the variability in API scores could be accounted for by the percentage of students eligible for free or reduced-price lunch, mobility, and the percent of English learners. The data also showed a positive relationship between teacher training and API scores.

In 2000, Eliezer Williams, et al., vs. State of California, et al. case was filed in San Francisco County Superior Court on behalf of nearly 100 San Francisco County students, claiming that they were denied equal access to instructional materials, safe and decent school facilities, and qualified teachers (CDE, n.d.). The case was settled in 2004, which resulted in additional funding for schools in the first, second, and third decile of the state ranking system (based on the 2003 API Base testing data). Additionally, the state now requires that reports of overall condition of their facilities, the number of teacher mis-assignments and vacant teacher positions, and the availability of textbooks or instructional materials be included in the School Accountability Report Card.

In 2004, more than 60% of California schools improved their 2003-04 Academic Performance Index (API) scores, leading to an overall 10-point growth in statewide API scores this year and that more than 60% of schools met their 2004 federal Adequate Yearly Progress (AYP) benchmarks (Slater, 2004).

NEW YORK'S SYSTEM OF ACCOUNTABILITY FOR STUDENT SUCCESS (SASS)

Description

In 2000, New York State introduced the System of Accountability for Student Success

(SASS) (Kadamus, 2000). SASS requires the calculation of a Performance Index for each school, which is the simple sum of the percentage of students who demonstrate proficiency on some of the learning standards (Level 2) and students demonstrating all of the learning standards (Levels 3 and 4) in both English language arts and mathematics. The Performance Index ranges from 100 (where 100% of the students perform at Level 2) to 200 (where 100% of the students perform at or above Level 3); schools at which all of the students perform at Level 1 receive a Performance Index score of zero. At the high school level, schools were expected to have 90% of the students meet the graduation requirements for English language arts and mathematics within four years of starting ninth grade. For the 2004-2005 school year, the state standard was set at a Performance Index of 150; that index will gradually increase to 200 by the year 2013-2014. Annual measurable objectives are determined as prescribed in the NCLB legislation, with the state standard as the target goal for proficiency.

It should also be noted that there is a similar system in Massachusetts, called the Composite Performance Index (CPI) (Massachusetts Department of Education, n.d.). For the CPI calculation, each student is awarded a set number of points based on their attainment one of five performance standards: 100 points for the proficient and advanced proficiency down to zero points for the lowest level of scores approaching proficiency, in 25-point increments. Separate scales are used for the primary state assessment, the Massachusetts Comprehensive Assessment System (MCAS), and the alternate assessment, the MCAS-Alt. Proficiency Index scores for the MCAS and MCAS-Alt are combined to create the CPI.

Description

The use of an index score that incorporates more than one level of proficiency provides schools with greater opportunity to move students past target cut scores than the current NCLB legislation (Linn, Baker, & Betebenner, 2002). Thus, SASS has the advantage of offering partial credit for students who perform below the proficient level. Still at issue though, is whether it is more effective to award points for only two or more than two cut scores. Linn et al. suggest that empirical analyses are necessary to examine the ramifications of the number of regions that receive differential values for index scores.

OHIO'S REPORT CARD SYSTEM

A cornerstone of Ohio's Report Card System is the Performance Index, which is a weighted average of students' performance scores in a school. The Performance Index is used to classify schools as Excellent, Effective, Continuous Improvement, Academic Watch, or Academic Emergency. The details of that system will not be reviewed here, save for a discussion of a Growth Calculation that is used exclusively for schools in Academic Watch or Academic Emergency. The Growth Calculation is used to reward a school or district for improving its Performance Index by at least ten points over two years by allowing that school's overall rating to move up one designation. Such a change in designation requires that the school's scores increase in each of the prior two years, with the most recent year's scores at least three points higher than the prior year's scores (Ohio Department of Education, n.d.). Additionally, the state has established the Ohio Accountability Task Force to guide the implementation of value-added progress measures into the accountability system. Beginning in 2007-2008, Ohio will incorporate measures of individual student grade-to-grade achievement gains into its Report Card System. Currently, the Task Force is debating the whether the definition of one year of growth should be calculated as typical growth (e.g., New Mexico or Chicago) or as a graded measure that follows a trajectory towards a particular goal (e.g., Louisiana or Kentucky) (Ohio Accountability Task Force, 2005).

UTAH PERFORMANCE ASSESSMENT SYSTEM FOR STUDENTS

In May 2005, a bill was signed in to law that promotes Utah's Performance Assessment System for Students (U-PASS) over the federal NCLB Act. While annual measurable objectives will be used for NCLB accountability, a separate system is in development for U-PASS accountability. At present, the details of that system are not public, but a presentation on the plan to state legislators provides some insight into the system (Park, 2005). Multiple assessments and indicators will be

incorporated, including criterion-referenced tests in English language arts, math, and science; a direct writing assessment; a high stakes basic skills exam in reading, writing, and math in administered tenth grade; attendance and graduation rates. Status will be measured as the percent proficient on each assessment; school status will be labeled acceptable or unacceptable based on a weighted average across subjects. Progress will be assessed by comparing the achievement levels of the same student from one year to the next year on the criterion-referenced assessments. Successive cohorts will be tracked using math and science achievement data at the secondary level, scores from the tenth grade competency exam, the writing assessment, and graduation rates. Progress will be defined as high, medium, or low, based on a weighted average of progress rates across subjects. Additional details on the system are currently unavailable.

CONCLUDING REMARKS

This review examined the ways in which nine states and two large cities approach the issue of measuring growth in student achievement. As other states and large school districts contemplate implementing similar programs, it should be evident that transitioning to such a system is not a simple undertaking. The variety of approaches discussed in this review provides evidence of the complexity of issues and circumstances that need to be addressed. In order to facilitate comparisons, an overview of the growth models used in the states and cities included in this review is presented in Table 5, along with selected characteristics of the school systems.

This review provides support for several observations regarding the approaches to measuring growth. One of the most apparent complications found in the different applications is the variety of ways that each jurisdiction uses their measures of student achievement to determine growth. Some measure growth longitudinally by following the same students from grade to grade. Other states track successive cohorts by comparing, for example, one year's grade four students with the next year's grade four students. Some states employ a mix of both of these approaches. This variation across states affects many aspects of how these programs function and could certainly change the decisions made on the basis of results.

The choice of a specific growth model from the available options presents another challenge. All of the jurisdictions included in this review used one or more of the following methods: change scores, residual regression, mixed models and HLM analyses. These methods differ considerably in terms of their ease of implementation and administration. Each method requires somewhat different levels of sophistication in the procedures related to data collection, database development, and data analysis. Also, vertical scales across grade levels are necessary and the development of these scales pose a number of difficulties, depending on the characteristics of the student achievement measures being employed. Additionally, test scores must be available in an electronic database with unique identification numbers for students (and teachers, if desirable); further, minimum sample sizes must be available for analysis (Doran, 2003; Rubin, Stuart, & Zanutto, 2004; Stevens, 2004).

It is also necessary to consider which educational and demographic variables to include in some types of models. The list of possible covariates typically includes measures of prior student achievement and socioeconomic status. However, there are often choices that must be made among the existing options to determine which proxies for these variables will be used, based on availability, cost, accuracy or other factors. This decision will often affect not only the results produced by the system but also the degree to which it is perceived to support unbiased and fair decisions.

Additional research is needed to clarify the consequences for schools and districts under different models of accountability. It is possible, even likely, that the application of one model over another will lead to different conclusions regarding the growth in achievement of the same group of students. This may be due to certain growth models advantaging schools (or districts or students) with a specific profile of characteristics. As states and other jurisdictions adopt these systems of measuring growth, it is critical to understand the functioning of the models and how each interacts with the characteristics of the school system adopting it.

Table 5: ***Overview of States' and Cities' Use of Growth for Accountability***

	Tennessee	Dallas, Texas	North Carolina	New Mexico	Chicago, Illinois	Louisiana	Kentucky	California	Ohio	Utah	New York
Overview											
# Of Public Schools*	1,628	219	2,245	801	613	1,522	1,381	9,087	3,815	803	4,470
# of Pre-K to 12 Students*	928,000	158,059	1,335,954	320,234	426,812	730,464	660,782	6,356,348	1,838,285	489,072	2,888,233
Minority Students*	28.7%	94.2%	40.8%	66.4%	91.2%	51.5%	13.1%	66%	20.30%	15.90%	45.8%
Children in Poverty*	18%		17%	26%	85.2%	24%	19%	19%	14%	11%	19%
Growth Measures											
School Grades	3 though 8	3 though 11	3 though 8, High School	3 though 9	3 through 8, and 11	4, 8, and 10	4, 5, 7, 8, 10, 11, 12	2 though 11	3 through 8, 10	3 through 8, 10	3-8, High School
First year Growth was used for Accountability	1992-1993	n/a	1996-1997	n/a	1996-1997	1999-2000	2001-2002	1999-2000	2002-2003	2003-2004	2000-2001
Still in Operation?	Diagnostic purposes only	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES
Financial Incentives	YES, For teachers	NO	YES, For teachers	NO	NO	YES, For schools	YES, For schools	YES, For schools (unfunded at present)	NO	NO	NO
Type of test	NRT	CRT	CRT	NRT	BOTH	BOTH**	BOTH	BOTH	CRT	CRT	
Method of student tracking	Longitudinal	Longitudinal	Quasi-longitudinal	Quasi-longitudinal	Successive cohorts & Longitudinal	Successive cohorts	Successive cohorts	Successive cohorts	Successive cohorts	Longitudinal and Quasi-Longitudinal	Successive cohorts
Model Details	Value-added mixed model analysis at the student, teacher, and school levels	Value-added HLM analysis	Residual regression analyses using data from 1992-1993 as a baseline	Change in scale score as compared to "typical growth"	Two school growth measures compared to citywide average growth, one measure of student gain, and AYP	Annual measurable objectives at the school level based on an index score	Annual measurable objectives at the school level based on an index score	Annual measurable objectives at the school level based on an index score	Growth as change in Performance Index for schools labeled Academic Watch or Academic Emergency	Details are unavailable	Annual measurable objectives at the school level based on an index score with partial credit for partial proficiency

* (State Report Cards, Education Week, 2005)
** In 2006, Louisiana will replace the NRT with a CRT.

REFERENCES

Barton, P.E. (1999). Too much testing of the wrong kind, too little of the right kind in K-12 education. Princeton, NJ: Educational Testing Service.

Baker, A., & Xu, D. (1995). The measure of education: A Review of the Tennessee Value Added Assessment System. Nashville, TN: Office of Education Accountability.

Biesk, J. (2005, April 27). Lawmakers hear overview of school accountability system. The Associated Press State and Local Wire.

Bryk, A.S., & Raudenbush, S.W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In R.D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 159-204). San Diego: CA: Academic Press.

California Department of Education. (n.d.). Williams case. Retrieved May 11, 2005 from: http://www.cde.ca.gov/eo/ce/wc/wmslawsuit.asp.

California Department of Education. (2005). 2004 Academic Performance Index Base Report: Information guide. Retrieved May 6, 2005 from: http://www.cde.ca.gov/ta/ac/ap/.

Carlson, D. (2003, April). Four methods of judging school quality and progress. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Carroll, A. (2005, January 14). Scores show progress, help find weaknesses. *Chattanooga Times Free Press*, p B1.

Chicago Public Schools. (n.d.a). Frequently asked questions about the Chicago public schools accountability system. Retrieved from: http://acctplan.cps.k12.il.us/pdfs/Frequently%20Asked%20Questions-2004.pdf.

Chicago Public Schools. (n.d. b). Overview of the Chicago public schools accountability system. Retrieved from: April 24, 2005 from http://cps.k12.il.us/AtAGlance.html.

Chicago Public Schools. (n.d. c). CPS at a glance. Retrieved from the Chicago Public Schools Web site: April 24, 2005 from http://cps.k12.il.us/AtAGlance.html.

Commins, J. (2004, March 4). Revisions to state test stall in committee. Chattanooga Times

Goldstein, H. (1991). Better ways to compare schools? *Journal of Educational Statistics*, 16(2), 89-92.

Jacob, B.A. (2005). Accountability, incentives, and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89, 761-796.

Kadamus, J.A. (2000). System of accountability for student success. Retrieved July 27, 2005 from http://www.emsc.nysed.gov/nyc/SASS/SASS-fieldmemo.PDF.

Kentucky Department of Education. (2004a). 2004 CATS interpretive guide: Detailed information on using your score reports. Retrieved from: http://www.education.ky.gov/KDE/Administrative+Resources/Testing+and+Reporting+/CATS/Accountability+System/2004+CATS+Interpretive+Guide.htm.

Keung Hui, T. (2004, December 20). Board won't alter scores: Middle school teachers won't get bonuses although tests were scored incorrectly. *The News and Observer*, p. B1.

Ladd, H. F., & Walsh, R. P. (2002). Implementing value-added measures of school effectiveness: getting the incentives right. *Economics of Education Review*, 21, p. 1-17.

Linn, R.L. (2001). The design and evaluation of educational assessment and accountability systems. CSE Technical Report 539. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, UCLA.

Linn, R.L. (2005). Issues in the design of accountability systems. CSE Technical Report 650. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, UCLA.

Linn, R. L., & Haug, C. (2002). Stability of school building accountability scores and gains. *Educational Evaluation and Policy Analysis*, 24(1), 29-36.

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

Louisiana Administrative Code Title 28. (2005). Louisiana school, district, and state accountability. (Part LXXXIII, Bulletin 11). Baton Rouge, LA: Author.

Louisiana Department of Education. (2004). 2003-2004 Accountability summary results: State and district table. Retrieved from: http://www.doe.state.la.us/lde/pair/ReportCards04/districts/2004%20SPR.pdf.

Massachusetts Department of Education. (n.d.). AYP and accountability history. Retrieved July 27, 2005 from http://www.doe.mass.edu/sda/ayp/.

Morgan, J.G. (2004). The Education Improvement Act: A progress report. Nashville, TN: Tennessee Comptroller of the Treasury, Office of Research and Educational Accountability.

New Mexico State Board of Education. (2002). Primary and secondary education public school accountability: General provisions. Retrieved April 19, 2005 from: http://www.ped.state.nm.us/dl/nmac6191.pdf.

New Mexico State Department of Education. (2003). New Mexico State Department of Education consolidated state application accountability workbook. Retrieved April 19, 2005 from: http://www.ped.state.nm.us/div/acc.assess/csi/dl/vol.1.acctblty.wkbk.pdf.

North Carolina Department of Public Instruction. (2004a). Accountability brief. Retrieved April 17, 2005 from: http://www.ncpublicschools.org/docs/accountability/reporting/2004memo/standards2004.pdf.

North Carolina Department of Public Instruction. (2004b). Report to the joint legislative education oversight committee on the implementation of the ABCs. Raleigh, NC: North Carolina Department of Public Instruction.

North Carolina Department of Public Instruction. (2002). State of the state: Educational performance in North Carolina 2002. Raleigh, NC: North Carolina Department of Public Instruction.

Ohio Accountability Task Force. (2005). Ohio Accountability Task Force meeting minutes. Retrieved May 4, 2005 from: http://www.ode.state.oh.us/accountability/PDF/FEB05.PDF.

Ohio Department of Education. (n.d.). Guide for Ohio's report card system 2003-2004. Retrieved May 4, 2005 from: http://www.ode.state.oh.us/accountability/general_policy_documents.asp.

Park, J.W. (2005). U-PASS Accountability system. Retrieved May 12, 2005 from http://www.schools.utah.gov/eval/u_pass_randa/Legislative%20%20U-PASS%20Presentation%204-18.ppt.

Pennsylvania Department of Education. (n.d.). Understanding Value-Added Analysis: An Introductory Guide for Pennsylvania Educators. Retrieved July 17, 2005 from: http://www.pde.state.pa.us/a_and_t/site/default.asp.

Powers, J.M. (2003). An analysis of performance-based accountability: Factors shaping school performance in two urban districts. *Educational Policy*, 17(5), 558-585.

Roeder, P.W. (1999). Educational reform and equitable excellence: The Kentucky experiment. Retrieved April 10, 2005 from: http://www.uky.edu/~proeder/keraweb.htm.

Sanders. W. (2003, April). Beyond No Child Left Behind. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Sanders, W., Saxton, A., & Horn, S. (1997). The Tennessee Value-Added Accountability System: A quantitative, outcomes-based approach to educational assessment. (pp.137-162). In J. Millman (Ed.). *Grading teachers, grading schools.* (pp. 81-99). Thousand Oaks, CA: Corwin Press.

Schreiner, B. (2005, February 23). Senate passes bill to allow alternatives to CATS testing. The Associated Press State & Local Wire.

Sentell, W. (2004, August 18). La. Students face new test in spring '06. *The Advocate*, p. 1-B.

Stevens, J. J., Estrada, S. & Parkes, J. (2000, April). Measurement issues in the design of state accountability systems. Roundtable presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Thum, Y.M. (2002). Measuring student and school progress with the California API. CSE Technical Report 578. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, UCLA.

Thum, Y.M., & Bryk, A.S. (1997). Value-added productivity indicators: The Dallas system. In J. Millman (Ed.). *Grading teachers, grading schools.* (pp. 100-109). Thousand Oaks, CA: Corwin Press.

Waggoner, M. (2005, January 14). NC panel debates raising standards for end-of-grade test. The Associated Press State and Local Wire, BC Cycle.

Webster, W. (2004, October). The Dallas school-level accountability model: The marriage of status and value-added approaches. Paper presented at the University of Maryland, Department of

Measurement, Statistics, and Evaluation annual invitational conference on Value-Added Modeling: Issues with Theory and Applications, College Park, MD.

Webster, W., & Mendro, R. L. (1997). The Dallas value-added accountability system. In J. Millman (Ed.). *Grading teachers, grading schools.* (pp. 81-99). Thousand Oaks, CA: Corwin Press.

Citation

Goldstein, Jessica & Behuniak, Peter (2005). Growth Models in Action: Selected Case Studies. *Practical Assessment Research & Evaluation*, 10(11). Available online: http://pareonline.net/getvn.asp?v=10&n=11

Authors

Jessica Goldstein is a doctoral candidate at the University of Connecticut. Her dissertation title is "Measuring Growth in Student Achievement: Can Different Statistical Models Lead to Different Consequences for Schools?" Peter Behuniak is a Professor in Residence at the University of Connecticut.

Correspondence concerning this article should be addressed to Jessica.Goldstein@uconn.edu.