



Education Policy Analysis Archives/Archivos
Analíticos de Políticas Educativas

ISSN: 1068-2341

epaa@alperin.ca

Arizona State University
Estados Unidos

Zvoch, Keith; Stevens, Joseph J.
Successive Student Cohorts and Longitudinal Growth Models: An Investigation of Elementary School
Mathematics Performance
Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, vol. 14, 2006, pp. 1-25
Arizona State University
Arizona, Estados Unidos

Available in: <http://www.redalyc.org/articulo.oa?id=275020543002>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal

Editor: Sherman Dorn

College of Education

University of South Florida

Volume 14 Number 2

January 20, 2006

ISSN 1068-2341

Successive Student Cohorts and Longitudinal Growth Models: An Investigation of Elementary School Mathematics Performance

Keith Zvoch

University of Nevada, Las Vegas

Joseph J. Stevens

University of Oregon

Citation: Zvoch, K., & Stevens, J. J. (2006). Successive student cohorts and longitudinal growth models: An investigation of elementary school mathematics performance. *Education Policy Analysis Archives*, 14(2). Retrieved [date] from <http://epaa.asu.edu/epaa/v14n2/>.

Abstract

Mathematics achievement data from three longitudinally matched student cohorts were analyzed with multilevel growth models to investigate the viability of using status and growth-based indices of student achievement to examine the multi-year performance of schools. Elementary schools in a large southwestern school district were evaluated in terms of the mean achievement status and growth of students across cohorts as well as changes in the achievement status and growth of students between student cohorts. Results indicated that the cross and between-cohort performance of schools differed depending on whether the mean achievement status or growth of students was considered. Results also indicated that the cross-cohort indicators of school performance were more reliably estimated than their between-cohort counterparts. Further examination of the performance indices revealed that cross-cohort achievement status estimates were closely related to student demographics while between-cohort estimates were associated with cohort enrollment size and cohort initial performance status. Of the four school performance indices studied, only student growth in achievement (averaged across



SOME RIGHTS RESERVED Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-nd/2.5/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published jointly by the Colleges of Education at Arizona State University and the University of South Florida. Articles are indexed by H.W. Wilson & Co. Send commentary to Casey Cobb (casey.cobb@uconn.edu) and errata notes to Sherman Dorn (epaa-editor@shermamdorn.com).

cohorts) provided a relatively reliable and unbiased indication of school performance. Implications for the No Child Left Behind school accountability framework are discussed.

Keywords: school accountability, longitudinal growth models, No Child Left Behind Act.

Over the past several years, states have developed educational accountability systems as a means for improving the achievement outcomes for students (see Fuhrman & Elmore, 2004; Ladd, 1996). Educational accountability systems have been built on an implicit theory of action that assumes a public airing of student achievement results and a structured program of rewards and sanctions is requisite to motivate school personnel to constructively respond to evidence of substandard student outcomes (Forte-Fast & Hebbler, 2004; Furhman & Elmore, 2004; Marion, et al., 2002). For state policy makers, the substandard outcome most in need of redress by system stakeholders is student performance on standardized achievement tests. As reflected in the weighting of accountability outcomes, achievement test scores have been utilized as the key evidential component for determining the relative efficacy of schools in each state accountability system (Goertz & Duffy, 2001; Stevens, Parkes, & Estrada, 2000). Although widespread, the use of standardized test data as the primary or sole means for evaluating school performance is not without controversy. Questions regarding measurement precision, alignment with instructional content, and fairness in use for special student populations make the reliance on achievement tests a concern for many (e.g., AERA, APA, & NCME, 1999; Baker & Linn, 2004; Barton, 2004; Linn, 2000; Popham, 1999). Nonetheless, with passage of the No Child Left Behind federal legislation (NCLB: No Child Left Behind Act, 2002), testing is now more ubiquitous and of higher stakes than ever before. Under NCLB, states must revise their accountability systems to include annual testing of students in grades 3 through 8 in mathematics and reading/language arts. Consequences for substandard performance have also become more uniform and more stringent. Schools now face the clear prospect of a probationary designation, staff restructuring and/or state takeover if achievement standards are not met (NCLB, 2002).

The institutionalization of mandatory testing across content area and grade level and the concomitant performance pressures that schools now face place a special burden on the analytic methods used to measure school performance. For accountability systems to work fairly and effectively, school performance indices need to be reliable and valid (Baker & Linn, 2004; Forte-Fast & Hebbler, 2004; Marion, et al., 2002). The challenge presented by the need for scientifically credible school performance data has led to investigation of the assessment approaches that have been used in state accountability systems. State approaches to school assessment can be categorized into those that measure school performance as a function of student achievement at one point in time (i.e., status) or those that measure the change in student achievement across two or more occasions. Status approaches (e.g., percent proficient, mean achievement) have been most commonly used by states and have had wide appeal because of the relative ease with which these measures can be calculated and understood by system stakeholders. However, status measures tend to be problematic when used for evaluative or accountability purposes. As singular snapshots of student achievement, status measures capture both the influence of student background and prior educational experience as well as current school contributions to student performance (Raudenbush, 2004; Raudenbush & Willms, 1995). The confounding of different sources of achievement performance presents a particular challenge under conditions commonly found in public school districts. Student assignment

to schools is not random, but is instead influenced by social and economic-based selection processes. The non-random sorting of families into neighborhoods and students into schools tends to result in a differential accountability burden for those schools that happen to serve large numbers of disadvantaged students (Raudenbush, 2004). Relative to their more advantaged counterparts, schools situated in impoverished contexts typically are required to produce a disproportionate increase in student achievement levels if state achievement standards are to be met and low performance sanctions are to be avoided.

Perhaps in partial recognition of the challenge that schools with disadvantaged intakes face when status-type measures are used to evaluate school performance, states have also utilized measures that index the change in student achievement between testing occasions. Measures of student changes in achievement are seen as an alternative means by which schools, particularly those with challenging intakes, can demonstrate positive effects on students. Several states have measured student changes in achievement by comparing the grade level performance of successive student cohorts (e.g., the mean performance of 3rd graders in 2004 is compared to the mean performance of 3rd graders in 2005: “quasi” change) in an attempt to mitigate school differences in student intake (Stevens, et al., 2000). However, measuring school effectiveness by the change in successive student cohort performance levels can also be problematic for evaluative and accountability purposes (Hill & DePascale, 2003). Recent investigations of the successive cohort approach demonstrate that estimates of year-to-year changes in the mean achievement of students tend to be affected in large part by sampling variation, measurement error, and unique, non-persistent factors (e.g., construction noise) that affect test scores on only one of the testing occasions (Kane & Staiger, 2002; Linn & Haug, 2002). As a result, the observed change in school mean performance across student cohorts may be due in large part to the year-to-year fluctuation in student characteristics and testing conditions rather than actual changes in student performance (Carlson, 2002; Linn & Haug, 2002).

The observed difficulty of obtaining valid and precise estimates of school performance when school compositions differ non-randomly and/or when the mean performance of successive student cohorts is compared has led to interest in measuring the achievement progress of individual students as another alternative for evaluating school performance (Teddle & Reynolds, 2000; Willms, 1992; Zvoch & Stevens, 2003). In this approach, the test scores of individual students are linked across time. Individual growth trajectories are then estimated by fitting a regression function to the time series data obtained on each student. A measure of school performance follows from averaging the individual growth trajectories within each school. Tracking the achievement progress of individual students has certain advantages over the status and quasi-change models that states have used for school accountability purposes. Conceptually, longitudinal models of student achievement growth better represent the time-dependent process of academic learning (Bryk & Raudenbush, 1988; Seltzer, Choi, & Thum, 2003; Willett, 1988). Further, unlike status models, indices that capture the year-to-year changes in student achievement provide a degree of control over the stable background characteristics of students that otherwise complicate the evaluation of school effectiveness (Ballou, Sanders, & Wright, 2004; Sanders, Saxton, & Horn, 1997; Stevens, 2005). In addition, school performance measures that follow from estimates of the achievement progress of individual students tend to be more reliable than school performance measures that are based on the changes in achievement status between successive student cohorts (e.g., Kane & Staiger, 2002). Indices of student achievement growth may thus offer an alternative for monitoring school performance that avoids some of the inherent difficulties associated with the achievement status and the quasi-change approaches to school evaluation.

Despite the potential of using individual time series data as a basis for measuring and evaluating school performance, states have a current disincentive for incorporating indices of student achievement growth into their accountability systems. Under NCLB, states are required to

utilize a status-type measure (i.e., the percentage of students “proficient” or above on one testing occasion) as the primary means for evaluating school performance. Secondly, states are permitted to evaluate schools that fail to meet standard by the percent proficient methodology by indexing the changes in proficiency between successive student cohorts (i.e., quasi-change). States can also choose to track the achievement progress of individual students as a third approach for evaluating school performance, but under the provisions of NCLB, this methodology can only serve to further identify schools in need of improvement (Olson, 2004). In other words, schools that meet standards either by the percent proficient or quasi-change approaches can be identified as needing improvement if a growth target is not met, but demonstrating strong student growth is not sufficient to avoid a low performance sanction if the school does not have an adequate percentage of students proficient by either of the two primary methodologies endorsed by NCLB.

The disincentive currently associated with using individual time series data to measure and evaluate school performance has not allowed states to take full advantage of the annual testing of students required under the NCLB legislation. At present, only a couple of states and a handful of school districts have examined school performance as a function of student achievement growth (e.g., Kiplinger, 2004; Sanders, et al., 1997; Webster & Mendro, 1997; Zvoch & Stevens, 2003). Even less common are examinations of the multi-year performance of schools using longitudinal data on successive student cohorts (see Ponisciak, & Bryk, 2005; Bryk, Thum, Easton, & Luppescu, 1998; Bryk, Raudenbush, & Ponisciak, 2004, for examples). The limited application of longitudinal growth modeling methods to achievement data collected on students over time has left unanswered questions about the viability of using these techniques in state accountability systems. Although the studies conducted to date suggest that indices of student achievement growth tend to provide a less biased and a potentially more stable estimate of school performance than some NCLB-endorsed alternatives, questions about the mechanics of implementation (e.g., cross-cohort or between cohort analyses, estimation of unadjusted or value-added models) and the feasibility of use remain to be clarified (Bryk, et al., 2004; Flicek, 2004; Raudenbush, 2004). In response, the present study was designed to provide one example of how longitudinal growth models can be used to assess school performance across multiple student cohorts. Of particular interest was ascertaining whether estimates of cohort-to-cohort changes in the achievement growth of students provide a sound alternative for measuring school improvement. Note however that the intent of the current investigation was only to provide a preliminary and exploratory examination of the behavior and viability of certain growth-based approaches to measuring school performance. As such, school performance estimates were examined in relation to student intake characteristics rather than being adjusted by them. The investigation was facilitated by the analysis of achievement data from three longitudinally matched elementary school student cohorts from a large school district in the southwestern United States. The following research questions were considered: 1) Does the cross-cohort performance of schools differ based on an examination of school mean achievement vs. an examination of school average rates of growth in achievement? 2) Are the cross-cohort school performance estimates related to selected school characteristics? 3) To what degree do estimates of the mean achievement status and achievement growth of schools change with each successive student cohort? 4) Are estimates of the cohort-to-cohort changes in school performance related to selected school characteristics? and, 5) How reliable, on average, are each of the school performance estimates?

Method

Participants

The multi-year performance of elementary schools was investigated by examining the mathematics achievement of students from three longitudinally matched cohorts. The school district that provided the test score data has 79 kindergarten through grade 5 elementary schools that serve over 30,000 students each year. The district serves a significant number of students from special populations. At the elementary school level, English Language Learners, students eligible for a free or reduced price lunch, and students from ethnic minority groups constitute approximately 20%, 50%, and 55% of the student body, respectively. Beginning in the 1999–2000 school year, all third, fourth, and fifth grade students were assessed annually on the TerraNova/CTBS5 Survey Plus, a norm-referenced achievement test (CTB/McGraw-Hill, 1997). Between 6,000 and 6,500 students in each grade were assessed each spring. Achievement data from the three most recent longitudinal cohorts were analyzed in the present study. Table 1 diagrams the data structure associated with the current investigation. In Table 1, it can be seen that third to fifth grade longitudinal matches were available for students who entered the third grade in 1999–2000 (cohort 1), 2000–01 (cohort 2), and 2001–02 (cohort 3). Cohort 1 thus consisted of students who were third graders in 1999–2000, fourth graders in 2000–01, and fifth graders in 2001–02. The second and third cohorts consisted of the two following elementary school third to fifth grade student cohorts (i.e., cohort 2 from 2000–01 to 2002–03, and cohort 3 from 2001–02 to 2003–04).

Table 1
Cohort Data Structure

Grade	Year				
	1999–2000	2000–01	2001–02	2002–03	2003–04
3	C1	C2	C3		
4		C1	C2	C3	
5			C1	C2	C3

Cohort 1 ($N = 3,325$), Cohort 2 ($N = 3,347$), Cohort 3 ($N = 3,322$); School $N = 79$

Within cohort matches were accomplished by the following set of procedures. For each cohort, students who participated in accountability testing in all three study years were selected ($N \sim 5,000$). To facilitate the study of school effects, students who attended the same elementary school in all three years were then identified. In each cohort, approximately 900 students transferred schools at least once during the respective three-year period studied. Next, students who did not have a mathematics score in any of the three study years ($N \sim 100$) were dropped from their cohorts. Finally, students who received one or more modified test administrations were eliminated from the working data files ($N \sim 600$). The sample exclusions resulted in the following within cohort sample sizes; cohort 1 ($N = 3,325$), cohort 2 ($N = 3,347$), cohort 3 ($N = 3,322$). The three cohorts were comprised of relatively equal numbers of students from special populations. The percentage of English Language Learners ranged between 11–13% per cohort while the percentage of students from economically disadvantaged backgrounds comprised 45 to 46% of the cohorts. The percentage of students from ethnic minority groups was also relatively constant at 54–55%

across cohorts. Note however that the exclusion of students who did not participate in all three test administrations, students who transferred schools, and students who received at least one modified test administration lowered the percentage of students from special populations below district averages. Implications associated with the disproportionate exclusion of students from special populations will be addressed in the discussion.

Measures

Outcome data analyzed in the current study were student scale scores on the mathematics subtest of the TerraNova/CTBS5 Survey Plus. The Survey Plus is a standardized, vertically equated, norm referenced achievement test. All items are selected-response. According to the publisher, the mathematics subtest measures a student's ability to apply grade appropriate mathematical concepts and procedures to a range of problem-solving situations. The publisher reports KR-20 estimates of reliability of .87 in grade 3, .89 in grade 4, and .87 in grade 5 (CTB/McGraw-Hill, 1997). Other measures utilized in the study were the five-year school average (i.e., 1999–2000 to 2003–04) of the percentage of students eligible for a free or reduced lunch ($M = .58$, $SD = .28$) and cohort enrollment size, averaged across the three student cohorts by school ($M = 42.27$, $SD = 18.81$).

Analytic Procedures

Three-level longitudinal models were estimated using the Hierarchical Linear Modeling (HLM) program, version 6.0 (Raudenbush, Bryk, Cheong, & Congdon, 2004). Models were estimated using student and school records that were collected in three data files. The first file (level-1) contained student and school identifiers, mathematics scale scores from students in each of the three cohorts, and a field for grade level. This file contained 30,051 records (i.e., three records for each of 10,017 students). The level-2 data file contained student and school identifiers and a field that designated cohort membership ($N = 10,017$). The level-3 data file contained only school identifiers ($N = 79$).

After preparing the data for analysis, an unconditional three-level model was first used to estimate a mathematics growth trajectory for each elementary school student, to partition the observed parameter variance into its within and between school components, and to estimate the average achievement score and average growth rate for each elementary school across the three student cohorts. The level-1 model was composed of a longitudinal growth model that fitted a linear regression function to each individual student's grade 3, 4, and 5 achievement scores. Equation 1 specifies the level-1 model,

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{Grade} - 3) + e_{tij} \quad (1)$$

where Y_{tij} is the outcome (i.e., mathematics achievement) at time t for student i in school j , π_{0ij} is the initial status of student ij (i.e., 3rd grade performance),¹ π_{1ij} is the linear growth rate across grades 3–5 for student ij , and e_{tij} is a residual term representing unexplained variation from the

¹ By subtracting a value of 3 from *GRADE*, initial status is defined as the expected achievement of student i in school j at the end of grade 3 [$\pi_{0ij} + \pi_{1ij}(3 - 3) = \pi_{0ij}$].

latent growth trajectory. Levels 2 and 3 in the HLM model estimate mean growth trajectories in terms of initial status and growth rate across all students (equations 2a and 2b) and across all schools (equations 3a and 3b).

$$\pi_{0ij} = \beta_{00j} + r_{0ij} \quad (2a)$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij} \quad (2b)$$

$$\beta_{00j} = \gamma_{000} + u_{00j} \quad (3a)$$

$$\beta_{10j} = \gamma_{100} + u_{10j} \quad (3b)$$

In equations 2a and 2b, it can be seen that the initial achievement status and growth of students is conceived as a function of school average achievement (β_{00j}) or school average growth (β_{10j}) and corresponding residuals (r_{0ij} , r_{1ij}). Similarly, the initial status and growth by school in equations 3a and 3b is conceived as a function of the grand mean achievement (γ_{000}) or the grand mean slope (γ_{100}) and corresponding residuals (u_{00j} , u_{10j}). Equations 3a and 3b were used to calculate the pooled estimates of school mean achievement (i.e., the mean performance of 3rd graders across the three cohorts) and school mean growth (i.e., the average 3rd to 5th growth rate of students across the three cohorts).

The second model estimated included a term to represent changes over time in the performance of successive cohorts. As with the unconditional model, student growth trajectories were estimated at level 1 (see equation 1), but in this model the achievement and growth of students was conceived to also vary at level 2 as a function of the temporal span from one cohort to another (coded with a value of 0 for the first cohort, a 1 for the second cohort, and a 2 for the third cohort). The linear cohort term represents the federal expectation, outlined in the NCLB legislation, that regular, annual progress in student proficiency be made from one cohort of students to the next.² Equations 4a and 4b specify the level-2 model.

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{Cohort}) + r_{0ij} \quad (4a)$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}(\text{Cohort}) + r_{1ij} \quad (4b)$$

Using the above coding scheme for cohort membership, the intercept status parameter, school average achievement (β_{00j}) becomes the expected mean performance of 3rd graders in cohort 1 (2000–02) whereas the intercept growth parameter, school mean growth (β_{10j}) becomes the expected growth in achievement across grades 3 to 5 for the first cohort (2000–02). In addition, the cohort term (β_{01j}) can be interpreted as the expected change in the 3rd grade mean achievement of schools across the three cohorts and the cohort term (β_{11j}) can be interpreted as the expected change in school mean growth rates across cohorts.

At level-3, between-school variation in the initial achievement status and growth rate of schools and the school-to-school differences in the cohort changes in achievement and growth were first modeled either in terms of the grand mean achievement (γ_{000}) or the grand mean slope (γ_{100}) of schools and corresponding residuals (u_{00j} , u_{10j}) or the grand mean achievement change (γ_{010}) or the

² The expectation of regular annual progress most often assumes a linear increase in school performance over succeeding student cohorts. This assumption may not always hold. The performance of schools could, for example, change across student cohorts in a non-linear fashion. In the present study, the time trend was modeled with a linear function as the time series was relatively short (three data points). When the time series is of longer duration, it may be necessary to represent the data with a more complex function.

grand mean growth change (γ_{110}) of schools (across cohorts) and corresponding residuals (u_{01j} , u_{11j} ; see equations 5a through 5d). Note that estimation of the residual variances enables assessment of the degree to which schools vary in the 3rd grade mean achievement in the first cohort (2000–02), u_{00j} ; the changes in 3rd grade mean achievement between the three cohorts, u_{01j} ; the achievement growth of elementary school students in the first cohort (2000–02), u_{10j} ; and the changes in the achievement growth of elementary school students between the three cohorts, u_{11j} . Equations 5a through 5d were used to calculate the within and between-cohort school performance estimates.

$$\beta_{00j} = \gamma_{000} + u_{00j} \quad (5a)$$

$$\beta_{01j} = \gamma_{010} + u_{01j} \quad (5b)$$

$$\beta_{10j} = \gamma_{100} + u_{10j} \quad (5c)$$

$$\beta_{11j} = \gamma_{110} + u_{11j} \quad (5d)$$

Results

Mathematics Achievement across Cohorts

Table 2 presents the results of model 1, the pooled HLM model. In the upper panel of Table 2, the results of the fixed effects regression model are presented. The first estimate shown, the grand mean (γ_{000}), is the average 3rd grade mathematics scale score across all students. The second estimate, the grand slope (γ_{100}), is the average yearly growth rate for those students. Across the three student cohorts, the average 3rd grade mathematics scale score was estimated as 616.97 while the average yearly growth rate across grades 3 to 5 was estimated to increase by 16.74 scale score units per year. In the next panel of Table 2, estimates of the student-to-student and school-to-school variation in achievement and growth rates are presented. Chi-square tests of the model's variance components indicated that students and schools differed significantly in achievement levels and the rate of achievement growth. The other estimates presented in the middle of Table 1 are the parameter reliabilities associated with each outcome measure. As can be seen in the table, most of the observed variability in the cross-cohort parameter estimates was true parameter variance (school mean achievement = .95, school mean growth = .84). The proportion of variation in student outcomes attributable to schools is presented in the bottom panel of Table 2. Twenty-one percent of the variation in student achievement level and 38% of the variation in student achievement growth was due to school-to-school differences.

To illustrate the school-to-school differences in mathematics achievement averaged across the three cohorts, empirical Bayes (EB) estimates of the 79 elementary school mathematics mean achievement and mean growth rates are presented in the scatterplot in Figure 1. The horizontal line in the interior of the figure represents the cross-cohort grand mean achievement in mathematics. The vertical line in the interior of the figure represents the cross-cohort grand mean growth in mathematics. The two grand mean reference lines classify schools into four quadrants of school performance. The upper right quadrant contains schools with above average cross-cohort mean achievement in grade 3 and above average cross-cohort growth from grades 3 to 5. The lower right quadrant contains schools with below average cross-cohort mean scores but above average growth. The two quadrants on the left side of the figure contain schools with below average cross-cohort growth and either high or low mean achievement. The spread of points in Figure 1 demonstrates that schools with low mean scores were not always low performing schools in terms of student

growth in achievement. Similarly, above average school mean achievement at grade 3 did not always translate into above average growth across grades 3 to 5. Schools with low grade 3 mean scores had above or below average growth as did schools with relatively high grade 3 mean scores. The lack of a consistent relationship between the mean achievement and growth of schools is reflected in the correlation between the model's level-3 residual terms ($\tau_\beta = -.16$). In these data, knowing a school's initial achievement status offered little insight into the subsequent achievement progress of students.

Table 2

Three-Level Cross-Cohort Model for Mathematics Achievement

Variable	Parameter estimates		
Fixed Effects	Coefficient	SE	t
School Mean Achievement, γ_{000}	616.97	1.69	365.81*
School Mean Growth, γ_{100}	16.74	0.40	41.50*
Random Effects	Variance Component	df	χ^2
Individual Achievement, r_{0ij}	790.85	9938	24535.66*
Individual Growth, r_{1ij}	17.95	9938	10826.72*
Level-1 Error, e_{tij}	408.11		
School Mean Achievement, u_{00j}	214.19	78	2087.91*
School Mean Growth, u_{10j}	10.82	78	542.96*
Reliability Estimates			
School Mean Achievement		.95	
School Mean Growth		.84	
Level-1 Coefficient	Percentage of Variation Between Schools		
Individual Achievement, π_{0ij}		21.3	
Individual Growth, π_{1ij}		37.6	

Results based on data from 10,017 students distributed across 79 elementary schools.

* $p < .001$

To assess the degree to which the estimates of school mean achievement and school mean growth were associated with schools' social context (a measure of bias), correlations between the EB estimates of school performance and schools' percentage free lunch rate were calculated. Percentage free lunch was strongly related to the average performance level of schools, $r(77) = -.81, p < .001$. Schools with a larger percentage of students eligible for free or reduced price lunch had student achievement levels that were lower than schools with smaller rates of free or reduced price lunch eligibility. However, knowing the percentage of the student body eligible for a free or reduced price lunch provided little insight into the average rate at which students learned mathematics across the three cohorts. A systematic relationship between percent free lunch and school mean growth was not observed, $r(77) = -.17, p > .05$.

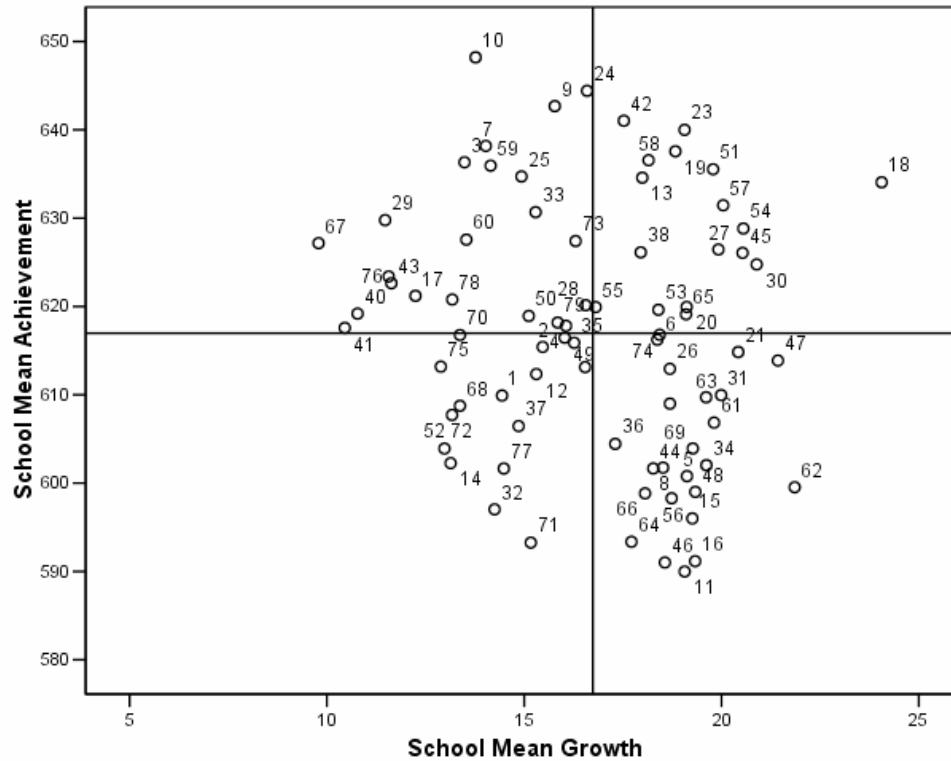


Figure 1. Cross-cohort relationship between school mean achievement and school mean growth in mathematics

Mathematics Achievement by Cohort

Table 3 presents the results of the second model that examined changes over time in the performance of successive cohorts. Estimates of the model's fixed effects are presented in the top panel of Table 3. The first estimate presented (γ_{000}) is the average 3rd grade mathematics scale score for the first student cohort (2000–02). The second estimate (γ_{010}) is the average cohort-to-cohort change in 3rd grade mean scale scores. These estimates indicate that the 3rd grade mean achievement of the first student cohort was 619.39 and that the 3rd grade mean achievement of schools decreased by 2.43 scale score points on average with each successive student cohort. The next estimates presented are the average growth rate across grades 3 through 5 for the first student cohort (γ_{100}) and the average cohort-to-cohort change in longitudinal growth rates (γ_{110}). These estimates indicate that the first student cohort grew an average of 15.75 scale score points per year and that the mean growth rate of schools across grades 3 through 5 was increasing by an average of 1.03 scale score units with each successive cohort. Variance estimates are presented next in Table 3. Chi-square tests demonstrated that in the first cohort of students, students and schools differed significantly with respect to achievement levels and rates of growth. Further, these tests also indicated that schools differed with respect to the changes in successive cohort performance. Statistically significant school-to-school variation was observed in the changes in 3rd grade mean achievement and the grade 3 to 5 changes in achievement growth between cohorts. Parameter reliability estimates are presented

in the bottom panel of Table 3. As with model 1, the mean achievement (.92) and mean growth (.78) of schools were estimated with relatively high parameter reliability, but note that these estimates were somewhat lower than their counterparts from the previous model that estimated the mean achievement and growth of schools across three student cohorts. In addition, the between cohort estimates of school performance were noticeably less reliable than within cohort mean achievement and growth estimates. Only half of the observed variability in the cohort-to-cohort changes in mean achievement status (.51) and two-thirds of the observed variability in the cohort-to-cohort changes in mean achievement growth (.68) was true parameter variance.

Table 3

Three-Level Between-Cohort Model for Mathematics Achievement

Variable	Parameter Estimates		
Fixed Effects	Coefficient	SE	t
School Mean Achievement, γ_{000}	616.97	1.69	365.81*
Mean Achievement Change, γ_{010}	-2.43	0.61	-3.96*
School Mean Growth, γ_{100}	16.74	0.40	41.50*
Mean Growth Change, γ_{110}	1.03	0.34	3.02*
Random Effects	Variance Component	df	χ^2
Individual Achievement, r_{0ij}	775.02	9859	23855.67*
Individual Growth, r_{1ij}	13.29	9859	10576.21
Level-1 Error, e_{tij}	408.11		
School Mean Achievement, u_{00j}	292.95	78	1212.08
Mean Achievement Change, u_{01j}	15.43	78	167.80
School Mean Growth, u_{10j}	10.82	78	542.96*
Mean Growth Change, u_{11j}	6.23	78	259.49*
Reliability Estimates			
School Mean Achievement			.92
Mean Achievement Change			.51
School Mean Growth			.78
Mean Growth Change			.68

Results based on data from 10,017 students distributed across 79 elementary schools.

* $p < .001$

The between-cohort change in school performance is illustrated in Figures 2 and 3. Fitted trajectories representing cohort-to-cohort changes in the 3rd grade mean achievement of schools are presented in Figure 2.³ In Figure 2, it can be seen that schools differed in terms of the mean achievement of the first student cohort and in terms of the change in mean achievement of 3rd graders over time. It can also be seen that while mean achievement of schools was generally decreasing over time, the cohort-to-cohort changes in the mathematics achievement of 3rd graders

³ To better demonstrate the directional change in school performance over successive cohorts, fitted trajectories are presented in Figures 2 and 3. The fitted trajectories mask the year-to-year fluctuations in cohort performance.

were relatively modest. The systematic association between the 3rd grade mean achievement of the first cohort and subsequent changes in cohort grade 3 mean performance is also evident in Figure 2. In Figure 2, it can be seen that the mean achievement of schools tended to regress toward the district mean and thus become more homogenous with each succeeding cohort. In other words, schools with a high-achieving 2000–02 cohort tended to demonstrate lower 3rd grade mean performance in subsequent student cohorts and schools with a low achieving 2000–02 cohort tended to demonstrate higher 3rd grade mean performance over subsequent cohorts. The correlation between the mathematics performance of 3rd graders in the 2000–02 cohort and the estimated change in the average mathematics performance of 3rd graders in subsequent cohorts was negative and relatively strong ($\tau_{00, 01} = -.70$).

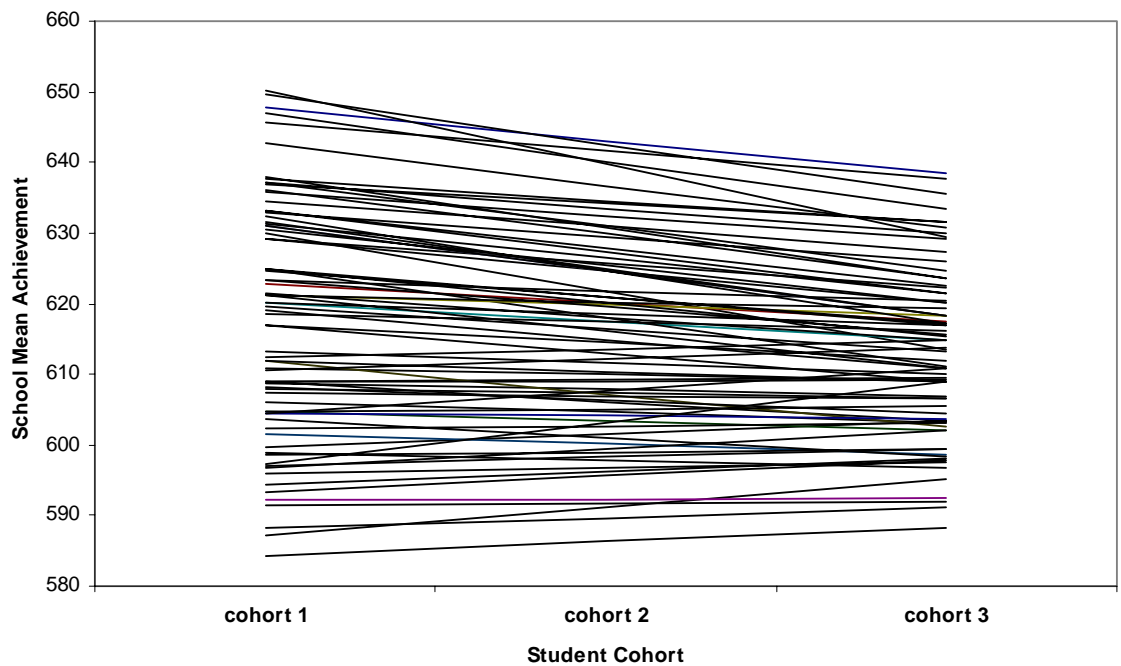


Figure 2. School mean achievement in mathematics as a function of student cohort

A similar picture emerged when cohort growth rates were examined. Figure 3 presents the cohort growth trajectories by school. In Figure 3, school-to-school differences in the growth rate of cohort 1 and the changes in cohort growth over time can be seen. School changes in cohort growth rates tended to be positive and somewhat more variable than the changes in school mean achievement displayed in Figure 2. However, the same overall pattern of relationship between initial status and subsequent change was again evident. Schools with a high performing 2000–02 student cohort (in terms of growth) had relatively less successful succeeding cohorts while schools that had an initially low performing cohort had higher growth rates with the following student cohorts. The relationship between the initial and subsequent growth of cohorts was negative but smaller in magnitude than the relationship between the mean achievement status and mean achievement change of cohorts ($\tau_{10, 11} = -.59$).

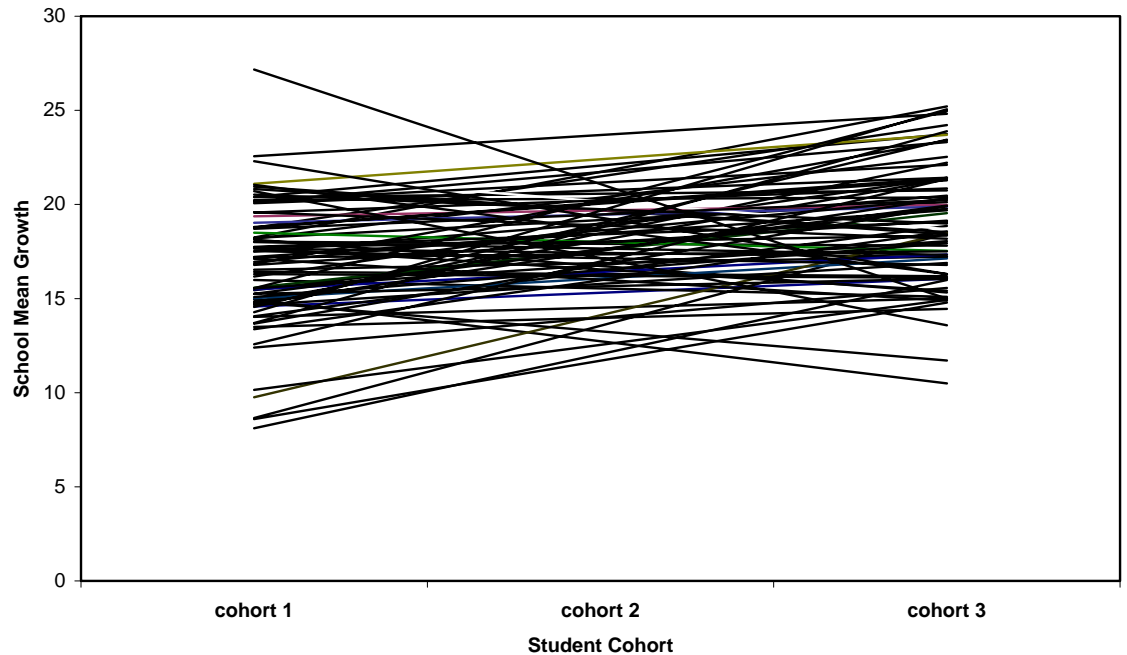


Figure 3. School mean growth in mathematics as a function of student cohort

To assess whether the cohort-to-cohort change scores were also associated with cohort enrollment size, school change estimates were plotted against the three-year cohort enrollment averages. Figure 4 presents the relationship between cohort-to-cohort changes in school mean growth and cohort enrollment size. In Figure 4, it can be seen that schools with small enrollments were more likely than schools with large enrollments to have above or below average changes in mean growth between cohorts.⁴ With the exception of one outlying school (school 23), large cohort-to-cohort changes in school mean growth tended to be concentrated in schools with relatively small enrollments. A similar pattern emerged when changes in school mean achievement were plotted against cohort enrollment size. The greater successive cohort change estimates for smaller schools suggest that relative to their larger counterparts, schools with small enrollments have greater potential for changes in the achievement outcomes of students. However, the differential impact (both positive and negative) of small enrollments is likely attributable to the heightened potential for differences in the composition of student cohorts, rather than any systematic differences in school policy or practice.

⁴ School changes in mean growth (between cohorts) were averaged across two change cycles, thereby reducing some of the variability in the change estimates.

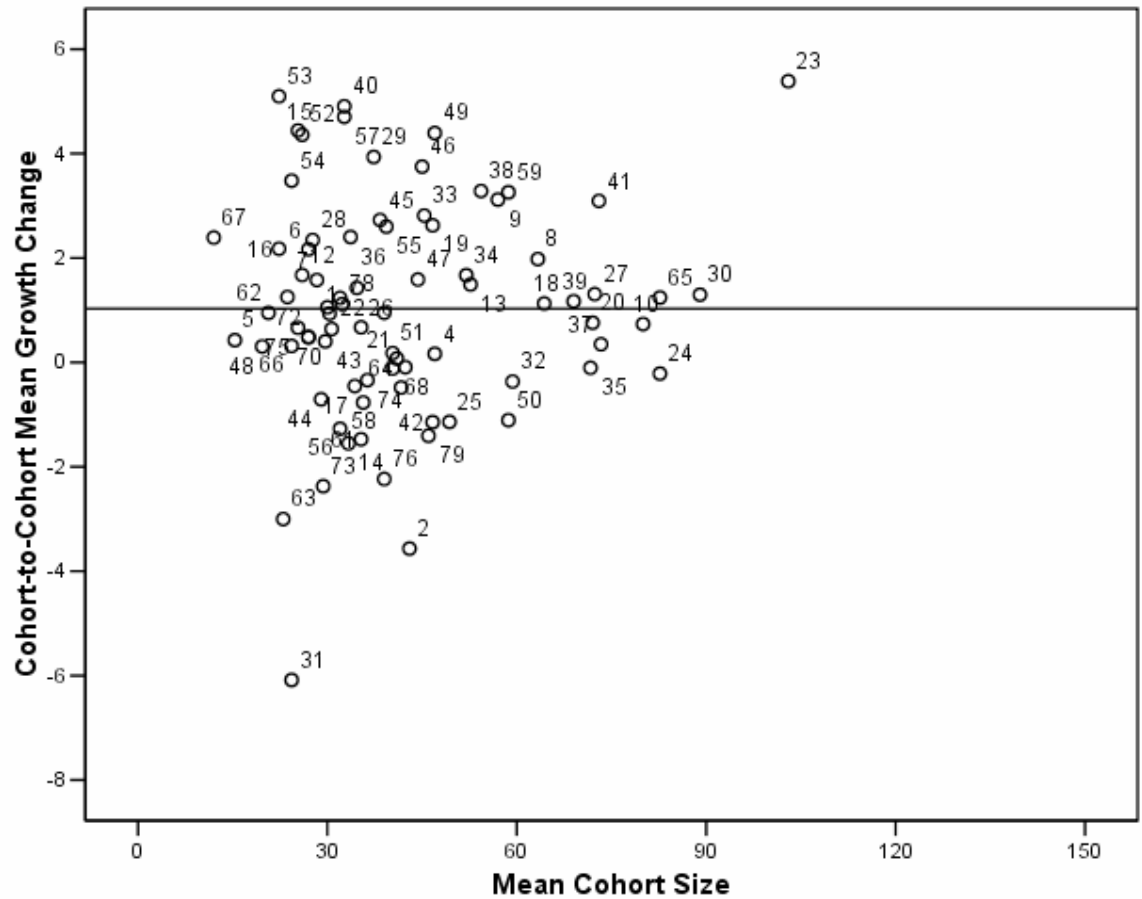


Figure 4. Relationship between cohort-to-cohort changes in school mean growth and cohort enrollment size by school

Discussion

With passage of NCLB, states are required to restructure their accountability systems to comply with a uniform set of federal guidelines. These guidelines outline the content areas (i.e., mathematics, reading/language arts) and students to be tested, the frequency of testing, the methodology to be used for evaluating school performance, and the set of consequences that befall schools failing to demonstrate adequate student achievement outcomes (NCLB, 2002). Of the changes NCLB has introduced to state accountability systems, one of the most far reaching stems from the manner in which school performance is to be evaluated. Under NCLB, states are required to annually evaluate schools in terms of the percentage of students who are at or above a particular cut-point or proficiency standard (i.e., status) and/or by the cohort-to-cohort change in the percentage of students who reach proficiency (i.e., quasi-change). The proficiency standard used to evaluate school performance is allowed to vary by state, but NCLB requires adequate yearly progress (AYP) toward the goal of having 100% of students reach the state-adopted standard in each content area by the year 2013–14.

Although NCLB is laudable in its aim to push schools toward providing an effective and equitable education for all students, concerns about the methodology used to evaluate school performance have been raised (Linn, 2003; Linn, Baker, & Betebenner, 2002; Ponsciak & Bryk, 2005; Raudenbush, 2004; Stevens, 2005). Of particular concern is whether NCLB methods for measuring school progress reliably capture the impact that schools have on students and whether these methods are biased against schools that serve students from disadvantaged backgrounds. Validity concerns about the linkage between treatment (i.e., instruction) and outcome (i.e., student achievement) and the differential accountability burden placed on schools with challenging intakes stem directly from the manner in which school performance is assessed. The achievement status and quasi-change approaches endorsed by NCLB monitor school effectiveness as a function of the absolute level of student performance and/or with respect to the change in student status across successive student cohorts. The use of measures that index the achievement status of a single cohort (relative to a proficiency target) or the change in status between two successive cohorts present a challenge for states as schools can be identified as in need of improvement on the basis of factors (e.g., student demographics) that are outside of the school's control (Linn & Haug, 2002; Kane & Staiger, 2002; Raudenbush, 2004). The potential for factors exogenous to the school to confound the measures of school performance endorsed by NCLB has led to calls for a reexamination of the school performance indices that are used to evaluate schools under NCLB (National Conference of State Legislatures, 2005; Olson, 2004). Of particular interest to system stakeholders is the potential for growth-based measures of school performance to enhance the fairness and equity of the federal accountability framework.

In the present study, the viability of using student growth rates as a means for evaluating the achievement progress of schools was investigated to ascertain whether indices of students' growth in achievement provide a reliable and valid alternative to the status and quasi-change approaches to school evaluation endorsed by NCLB. The investigation was based on the analysis of achievement data from three longitudinally matched elementary school student cohorts from a large school district in the southwestern United States. Results indicated that the cross-cohort performance of schools differed depending on whether the mean achievement status or growth of students was considered. Across the three cohorts studied, the relationship between the initial achievement status of students and students' subsequent achievement progress was quite weak as schools with high initial achievement (averaged across cohorts) were generally as likely as schools with low initial achievement to have low, average, or high mean achievement growth. The same was generally true of the relationship between school demographics and school mean rates of achievement growth. Knowing the percentage of the student body eligible for a free or reduced price lunch provided little insight into the average rate at which students learned mathematics across the three cohorts. However, the free-lunch percentage was strongly related to the average performance level of schools. Schools with a larger percentage of students eligible for free or reduced price lunch had student achievement levels that were lower than schools with smaller rates of free or reduced lunch eligibility.

Between-cohort estimates of school improvement (i.e., cohort-to-cohort changes in student achievement) provided an additional perspective on the effectiveness of schools by indexing the degree to which school performance changed with each succeeding student cohort. Over the study period, schools tended to have lower mean achievement scores but increased rates of student growth. On average, the mean achievement of third graders decreased by close to two and a half scale score points per cohort while the growth in mathematics achievement across grades 3 to 5 increased by slightly more than one scale score point per year with each succeeding cohort. The overall cohort-to-cohort changes in school performance were thus relatively modest and "equalizing" across cohorts. In other words, the decreases in third grade mean achievement were

small in magnitude and offset by increases in student achievement growth so that the grand mean performance of 5th graders remained relatively constant across the study period. The negative relationship between the mean achievement and growth of schools was reflected in the correlation between the model's residual change estimates ($\tau_{01, 11} = -.61$). Schools that had increases in the third grade achievement status of subsequent student cohorts were less likely to have an increase in cohort growth rates and vice versa. In fact, only a handful of schools had either simultaneous increases or decreases in the achievement and growth of student cohorts.

The difficulty schools face in delivering continual increases in student achievement outcomes was also reflected in the coefficients relating the achievement and growth status of cohort 1 to the changes in student achievement and growth between cohorts. In both instances, schools with high initial performance (either in terms of mean achievement or mean growth) were less able than schools with low initial performance to demonstrate positive changes in the achievement and growth of subsequent student cohorts. Clear evidence of the regression effect is displayed in Figures 2 and 3. In these figures, it can be seen that while the mean achievement of third graders was slightly decreasing and the growth in achievement across grades 3 to 5 was slightly increasing from cohort-to-cohort, school performance estimates were becoming more similar over time. For the majority of schools then, student performance changed very little from cohort-to-cohort on either outcome. However, for those schools with relatively extreme initial status performance estimates, the observed cohort-to-cohort changes in student achievement served to homogenize school performance as student achievement and growth tended to regress toward the district's achievement status and growth averages.

The relationship between the performance status of cohort 1 and the subsequent changes in achievement between cohorts is an indication that a school's ability to increase student achievement outcomes may be contingent upon how well students initially perform.⁵ However, it is worth noting that the changes in school performance were related to student cohort size as well. Schools with smaller student cohorts had greater changes in student outcomes than schools with larger cohorts. The greater volatility of the successive cohort change estimates for smaller schools follow in part from the heightened potential for differences in the make-up of student cohorts to occur when schools serve relatively small numbers of students (Kane & Staiger, 2002; Linn & Haug, 2002). The volatility of the cohort-to-cohort school improvement estimates was also reflected to some degree in the consistency with which these parameters were estimated. Relative to the consistency with which the cross-cohort mean achievement and growth of schools was estimated, cohort-to-cohort changes in school mean achievement and school mean growth were noticeably less reliable indicators of school performance.

In many respects, results of the current study were consistent with other recent investigations of the reliability and validity of various school performance indicators. As with findings from other recent studies, the level at which students in a school achieved (i.e., school mean achievement) was estimated with high reliability but was closely tied to the level of economic hardship experienced by students (Raudenbush, 2004). In addition, the modest changes in school mean achievement, the negative relationship between initial cohort mean achievement status and successive cohort mean change, the greater volatility of the mean change estimates for small schools,

⁵ The relationship between initial status and school changes in performance could also be due to district policies, including those aimed at school improvement, that are sufficiently uniform to draw achievement scores together. In other districts or in national samples, regression effects may not be as pronounced.

and the overall less reliable estimates of cohort-to-cohort changes in mean achievement also mirrored other recent findings (Hill & DePascale, 2003; Kane & Staiger, 2002; Linn & Haug, 2002; Schwarz, Yen, & Schafer, 2001). The current study was unique, however, in the focus on evaluating school performance from the perspective of changes in individual student achievement across and between cohorts. Estimates of the growth in student achievement across cohorts tended to provide a relatively reliable and unbiased measure of school performance. However, estimates of the cohort-to-cohort changes in student achievement growth shared similar properties with estimates of the successive cohort mean change score. For example, cohort-to-cohort changes in student achievement growth were estimated with less reliability than the cross-cohort student growth estimates. The average cohort-to-cohort changes in student achievement growth were also generally small in magnitude and tied closely to cohort enrollment size and the first cohort's initial growth status. These results further highlight the difficulties associated with comparing successive student cohorts, even those that are longitudinally matched over time. Changes in school performance between student cohorts tend to be quite modest when averaged across schools while the changes in cohort performance for any one school can result from idiosyncrasies associated with the composition of the cohort being evaluated rather than with any real change in the effectiveness of instruction at a school.

Results of the current study provide some indication of the strengths and weaknesses associated with four distinct measures of school performance. However, consideration of sample and data limitations is necessary for contextualizing the current findings. Specifically, it should be noted that the study was based on the norm-referenced mathematics achievement of students. The patterns seen in norm-referenced math achievement may not be the same in other subject areas or if scores were taken from a criterion-referenced instrument. Results were also based on achievement data from a select, non-transient student sample. The sample analyzed differed (in terms of student demographics) from the district's general student population and may have produced an upward bias on estimates of student and school achievement outcomes (Zvoch & Stevens, 2005). The study also focused on the analysis of entire student cohorts. The focus on the achievement performance of entire student cohorts differs from the NCLB requirement that achievement outcomes also be disaggregated by student subgroups. The achievement outcomes associated with disaggregated groups may or may not mirror the results reported here, although it is likely that due to the smaller size of student subgroups, estimates of year-to-year changes in school performance would be more volatile. Generalizability concerns also follow from the analysis of data from a single southwestern school district. As with other school districts located in the same geographic region, the district studied serves large numbers of Hispanic students and large numbers of English-language learners. The high percentage of students from these demographic groups distinguishes this district from many others in the United States and may limit the generalizability of results.

The study also may have been limited to some degree by constraints associated with the data structure. Of particular concern is that achievement data were not available until students were in grade 3. Not having data on students' kindergarten entry status and achievement growth from kindergarten to grade 2 makes it difficult to know the true school effect on students. For example, schools that appeared average in terms of student growth across grades 3 to 5 may have been either more or less effective for students across kindergarten to grade 2. In the former scenario (i.e., high kindergarten to grade 2 growth, average grade 3 to 5 growth), the school would be judged as less effective than warranted. A related concern follows from the number of cohorts available for analysis. In the current study, estimates of school improvement were based on the changes in performance between three student cohorts. The small number of cohorts available for estimating school trends in achievement along with the observed fluctuations in cohort performance led to relatively unreliable estimates of school improvement. Although not inconsistent with findings from

previous investigation of school achievement trends (Hill & DePascale, 2003; Kane & Staiger, 2002; Linn & Haug, 2002), current school improvement estimates may not have been as indicative of the true change in school achievement outcomes as would be required for high stakes decision-making (see Bryk, et al., 2004; Raudenbush, 2004).

Although the aforementioned limitations suggest a need for additional research on the multi-year performance of schools, both within and between cohorts and across different sampling conditions, the current study does provide a glimpse into the potential usefulness of various indicators of school performance. Of the four measures examined in the current study, estimates of the growth of students within cohorts (or averaged across cohorts) offered the most favorable combination of attributes for assessing the effectiveness of schools. Although slightly less reliable than estimates of school mean achievement, estimates of school mean growth were more reliable than either of the cohort change measures. School mean growth estimates were also less confounded by student demographics than their school mean achievement counterparts. In addition, by capturing the gains that students achieve over time instead of student performance on a particular testing occasion, school mean growth tends to be a more conceptually defensible indicator of school performance. The combination of attributes afforded by the achievement growth estimates coupled with the difficulties associated with the mean achievement and successive cohort change measures suggest that consideration should be given to incorporating growth measures (either within or averaged across cohorts) into state accountability systems. One approach to utilizing growth data for school accountability purposes would be to evaluate schools on the basis of the percentage of students meeting an annual growth target. Assessing school performance with respect to the percent of students meeting “expected” growth instead of the percent of students proficient at any one time would potentially enable schools serving disadvantaged student populations to demonstrate positive instructional impacts on students and simultaneously keep schools with advantaged intakes honest. Utilizing the growth of students as a measure of school performance would also enable states to avoid evaluating schools on the basis of inherently volatile short-term successive cohort comparisons. A change in accountability focus from status-based measures to student growth indices would not be without difficulty however. Issues surrounding student mobility, test alignment and equating, the setting of growth targets, demographic change, and incomplete time series data lead to a different set of challenges for the design of state accountability systems (Bryk, et al., 2004; Gong, 2004). Nevertheless, if the effectiveness of schools is to be determined on the basis of student performance on standardized tests, it seems reasonable to construct an accountability framework that enables schools to be evaluated on an outcome measure that more closely taps the school contribution to student learning.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.
- Baker, E. L., & Linn, R. L. (2004). Validity issues for accountability systems. In S. H. Fuhrman & R.F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 47–72). Teachers College Press: New York.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37–65.
- Barton, P. E. (2004). *Unfinished business: More measured approaches in standards-based reform*. Princeton, NJ: Educational Testing Service. Retrieved February 28, 2005, from http://www.ets.org/Media/Education_Topics/pdf/unfinbusiness.pdf.
- Bryk, A. S., Raudenbush, S. W. & Ponisciak, S. (2004). *A value-added model for assessing improvements in school productivity: Results from the Washington, DC public schools and an analysis of their statistical conclusion validity*. Chicago, IL: Consortium on Chicago School Research.
- Bryk, A. Thum, Y. M., Easton, J. Q., & Luppescu, S. (1998). Assessing school academic productivity: The case of Chicago school reform. *Social Psychology of Education*, 2, 103–142.
- 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.
- Carlson, D. (2002). The focus on state educational accountability systems: Four methods of judging school quality and progress. In W.J. Erpenbach (Ed.), *Incorporating multiple measures of student performance into state accountability systems: A compendium of resources* (pp. 285–297). Washington, DC: Council of Chief State School Officers.
- CTB/McGraw-Hill (1997). *TerraNova Technical Bulletin 1*. Monterey, CA: Author.
- Flicek, M. (2004). *Judging school quality using longitudinal methods that are comprehensible to stakeholders*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Forte-Fast, E., & Hebbler, S. (2004). *A framework for examining validity in state accountability systems*. Washington, DC: Council of Chief State School Officers.
- Fuhrman, S. H., & Elmore, R. F. (2004). *Redesigning accountability systems for education*. Teachers College Press: New York

- Goertz, M. E., & Duffy, M. C. (2001). *Assessment and accountability systems in the 50 states: 1999–2000* (Research Report No. RR-046). Consortium for Policy Research in Education, University of Pennsylvania.
- Gong, B. (2004). *Models for using student growth measures in school accountability*. Paper presented at the Council of Chief State School Officers' "Brain Trust" on Value-added Models, Washington, DC.
- Hill, R. K., & DePascale, C. A. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practice*, 22, 12–20.
- Kane, T. J., & Staiger, D. O. (2002). Volatility in school test scores: Implications for test-based accountability systems. *Brookings Papers on Educational Policy*, 1, 235–283.
- Kiplinger, V. L. (2004). Longitudinal study of student achievement growth in reading, writing, and mathematics achievement in thirty-six Colorado public school districts: Phase II final report. Colorado Springs, CO: Academic School District Twenty. Retrieved August 29, 2005, from <http://matrix10.d20.co.edu/lssg/phase2/Phase2Report.pdf>.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3–13.
- Linn, R. L., (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16.
- 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 requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3–16.
- Marion, S., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S, Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. Washington, DC: Council of Chief State School Officers.
- National Conference of State Legislatures (2005). *Task force on No Child Left Behind final report*. Retrieved August 9, 2005, from http://www.ncsl.org/programs/educ/nclb_report.htm.
- No Child Left Behind Act of 2001, Pub. L. No. 107–110 (2002).
- Olson, L. (2004). Value-added models gain in popularity. *Education Week*, 24(12), 1–4.
- Ponisciak, S. M., & Bryk, A. (2005). Value-added analysis of the Chicago Public Schools: An application of hierarchical models. In R. Lissitz (Ed.), *Value-added modeling in education: Theory and applications* (pp. 40–79). Maple Grove, MN: JAM Press.
- Popham, W. J. (1999). Where large scale educational assessment is heading and why it shouldn't. *Educational Measurement: Issues and Practice*, 18(3), 13–17.

- Raudenbush, S. W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Paper presented at the William H. Angoff Memorial Lecture Series, Princeton, NJ. Retrieved January, 22, 2005, from <http://www.ets.org/research/researcher/PIC-ANG9.html>.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.
- Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20, 307–335.
- Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added assessment system (TVAAS): A quantitative, outcomes-based approach to educational assessment in J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137–162). Thousand Oaks, CA: Corwin Press.
- Schwarz, R. D., Yen, W. M., & Schafer, W. D. (2001). The challenge and attainability of goals for adequate yearly progress. *Educational Measurement: Issues and Practice*, 20, 26–33.
- 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.
- Stevens, J. J. (2005). The study of school effectiveness as a problem in research design. In R. Lissitz (Ed.), *Value-added modes in education: Theory and applications* (pp. 166–208). Maple Grove, MN: JAM Press.
- Stevens, J. J., Estrada, S., & Parkes, J. (2000). *Measurement issues in the design of state accountability systems*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Teddlie, C. & Reynolds, D. (2000). *The international handbook of school effectiveness research*. New York: Falmer Press.
- Webster, W. J., & Mendro, R. L. (1997). The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 81–99). Thousand Oaks, CA: Corwin Press.
- Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Rothkopf (Ed.), *Review of research in education 1988–89* (pp. 345–422). Washington: American Educational Research Association.
- Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: Falmer Press.
- Zvoch, K., & Stevens, J. J. (2005). Sample exclusion and student attrition effects in the longitudinal study of middle school mathematics performance. *Educational Assessment*, 10(2), 105–123.

Zvoch, K., & Stevens, J. J. (2003). A multilevel, longitudinal model of middle school math and language achievement. *Educational Policy Analysis Archives*, 11(20). Retrieved July 8, 2003, from <http://epaa.asu.edu/epaa/v11n20>.

About the Authors

Keith Zvoch

University of Nevada, Las Vegas

Joseph J. Stevens

Affiliation information

Email: zvochk@unlv.nevada.edu

Keith Zvoch is an Assistant Professor in the Department of Educational Psychology at the University of Nevada, Las Vegas. His research interests include program evaluation, educational assessment, and school accountability.

Joseph J. Stevens is an Associate Professor in the Department of Educational Leadership at the University of Oregon. His research interests are assessment, validity, policy and accountability issues.

EDUCATION POLICY ANALYSIS ARCHIVES <http://epaa.asu.edu>

Editor: Sherman Dorn, University of South Florida

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Sherman Dorn, epaa-editor@shermamdorn.com.

Editorial Board

Michael W. Apple

University of Wisconsin

Robert Bickel

Marshall University

Casey Cobb

University of Connecticut

Gunapala Edirisooriya

Youngstown State University

Gustavo E. Fischman

Arizona State University

Gene V Glass

Arizona State University

Aimee Howley

Ohio University

William Hunter

University of Ontario Institute of Technology

Benjamin Levin

University of Manitoba

Les McLean

University of Toronto

Michele Moses

Arizona State University

Michael Scriven

Western Michigan University

John Willinsky

University of British Columbia

David C. Berliner

Arizona State University

Greg Camilli

Rutgers University

Linda Darling-Hammond

Stanford University

Mark E. Fetler

California Commission on
Teacher Credentialing

Richard Garlikov

Birmingham, Alabama

Thomas F. Green

Syracuse University

Craig B. Howley

Appalachia Educational Laboratory

Daniel Kallós

Umeå University

Thomas Mauhs-Pugh

Green Mountain College

Heinrich Mintrop

University of California, Berkeley

Anthony G. Rud Jr.

Purdue University

Terrence G. Wiley

Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES
English-language Graduate-Student Editorial Board

Noga Admon
New York University

Jessica Allen
University of Colorado

Cheryl Aman
University of British Columbia

Anne Black
University of Connecticut

Marisa Cannata
Michigan State University

Chad d'Entremont
Teachers College Columbia University

Carol Da Silva
Harvard University

Tara Donahue
Michigan State University

Camille Farrington
University of Illinois Chicago

Chris Frey
Indiana University

Amy Garrett Dikkers
University of Minnesota

Misty Ginicola
Yale University

Jake Gross
Indiana University

Hee Kyung Hong
Loyola University Chicago

Jennifer Lloyd
University of British Columbia

Heather Lord
Yale University

Shereeza Mohammed
Florida Atlantic University

Ben Superfine
University of Michigan

John Weathers
University of Pennsylvania

Kyo Yamashiro
University of California Los Angeles

Archivos Analíticos de Políticas Educativas

Associate Editors

Gustavo E. Fischman & Pablo Gentili

Arizona State University & Universidade do Estado do Rio de Janeiro

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez

Editorial Board

Hugo Aboites

Universidad Autónoma
Metropolitana-Xochimilco

Dalila Andrade de Oliveira

Universidade Federal de Minas
Gerais, Belo Horizonte, Brasil

Alejandro Canales

Universidad Nacional Autónoma
de México

Erwin Epstein

Loyola University, Chicago,
Illinois

Rollin Kent

Universidad Autónoma de
Puebla. Puebla, México

Daniel C. Levy

University at Albany, SUNY,
Albany, New York

María Loreto Egaña

Programa Interdisciplinario de
Investigación en Educación

Grover Pango

Foro Latinoamericano de
Políticas Educativas, Perú

Angel Ignacio Pérez Gómez

Universidad de Málaga

Diana Rhoten

Social Science Research Council,
New York, New York

Susan Street

Centro de Investigaciones y
Estudios Superiores en
Antropología Social Occidente,
Guadalajara, México

Antonio Teodoro

Universidade Lusófona Lisboa,

Adrián Acosta

Universidad de Guadalajara
México

Alejandra Birgin

Ministerio de Educación,
Argentina

Ursula Casanova

Arizona State University,
Tempe, Arizona

Mariano Fernández

Enguita Universidad de
Salamanca. España

Walter Kohan

Universidade Estadual do Rio
de Janeiro, Brasil

Nilma Limo Gomes

Universidade Federal de
Minas Gerais, Belo Horizonte

Mariano Narodowski

Universidad Torcuato Di
Tella, Argentina

Vanilda Paiva

Universidade Estadual Do
Rio De Janeiro, Brasil

Mónica Pini

Universidad Nacional de San
Martín, Argentina

José Gimeno Sacristán

Universidad de Valencia,
España

Nelly P. Stromquist

University of Southern
California, Los Angeles,
California

Carlos A. Torres

UCLA

Claudio Almonacid Avila

Universidad Metropolitana de
Ciencias de la Educación, Chile

Teresa Bracho

Centro de Investigación y
Docencia Económica-CIDE

Sigfredo Chiroque

Instituto de Pedagogía Popular,
Perú

Gaudêncio Frigotto

Universidade Estadual do Rio
de Janeiro, Brasil

Roberto Leher

Universidade Estadual do Rio
de Janeiro, Brasil

Pia Lindquist Wong

California State University,
Sacramento, California

Iolanda de Oliveira

Universidade Federal
Fluminense, Brasil

Miguel Pereira

Catedrático Universidad de
Granada, España

Romualdo Portella do

Oliveira

Universidade de São Paulo

Daniel Schugurensky

Ontario Institute for Studies in
Education, Canada

Daniel Suarez

Laboratorio de Políticas
Publicas-Universidad de
Buenos Aires, Argentina

Jurjo Torres Santomé

Universidad de la Coruña,
España