

Growth, Standards and Accountability[†]

Damian W. Betebenner
The Center for Assessment
DBetebenner@nciea.org

April 6, 2009

Abstract

Over the last decade, large scale annual testing has provided states with unprecedented access to longitudinal student data. Current use of this data focuses primarily upon analyses of growth most often directed toward accountability decisions. Analyses using this longitudinal data source for other purposes have gone largely untapped. This paper introduces analysis techniques and results showing how student growth percentiles, a normative growth analysis technique, can be used to examine the relationship between standards based accountability systems and the performance standards on which they are based.

Introduction

The impact of NCLB upon research connecting large scale assessment outcomes with school quality has been profound leading to a proliferation of initiatives centered on standards based accountability. The rapid development of all aspects of standards based accountability systems including the establishment of performance standards, the development of status based accountability systems, and research on growth analysis techniques has produced vast amounts of specialized knowledge promoting a sense that these areas are largely distinct. To the contrary, this paper promotes the view that these disparate topics are actually closely interconnected, needing only a lens to assist stakeholders in viewing the interconnections between the sometimes disparate pieces. This paper supplies that lens.

[†]Special thanks to the Colorado Department of Education for their generous support for work on the Colorado Growth Model on which this work is based. All opinions are those of the author and are not necessarily those of the Colorado Department of Education. Permission is granted to copy, distribute and/or modify this document and its contents under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License. No commercial or for-profit use of this work is permitted without the express written consent of Damian W. Betebenner. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Current state accountability systems rely heavily upon performance standards to make judgments about the quality of education. Specifically, accountability systems constructed according to federal adequate yearly progress (AYP) requirements use annual “snap-shots” of student achievement relative to state performance standards to make judgments about education quality. Since their adoption, such standards based *status measures* have been the focus of persistent criticism (Linn, 2003; Linn, Baker, & Betebenner, 2002). Though appropriate for making judgments about the achievement level of students, they are inappropriate for judgments about educational *effectiveness*. In this regard, status measures are blind to the possibility of low achieving students attending effective schools. It is this possibility that has led some critics of No Child Left Behind (NCLB) to label its accountability provisions as unfair and misguided and to demand the use of growth analyses as a better means of auditing school quality and increasing the validity of accountability systems (Braun, 2008).

A fundamental premise associated with using student growth for school accountability is that “good” schools bring about student growth in excess of that found at “bad” schools. Students attending such schools—commonly referred to as highly effective/ineffective schools—tend to demonstrate extraordinary growth that is causally attributed to the school or teachers instructing the students. The inherent believability of this premise is at the heart of current enthusiasm to incorporate growth into accountability systems. It is not surprising that the November 2005 announcement by Secretary of Education Spellings for the Growth Model Pilot Program (GMPP) permitting states to use growth model results as a means for compliance with NCLB achievement mandates was met with great enthusiasm by states. (Spellings, 2005).

Consistent with current accountability systems that hold schools responsible for the assessment outcomes of their students, the primary thrust of growth analyses over the last decade has been to determine, using sophisticated statistical techniques, the amount of student progress/growth attributable to the school or teacher (Braun, 2005; Rubin, Stuart, & Zanutto, 2004; Ballou, Sanders, & Wright, 2004; Raudenbush, 2004). Such analyses, often called *value-added* analyses, purport to estimate the teacher/school contribution to student achievement. This contribution, called the *school or teacher effect*, quantifies the impact on achievement that this school or teacher would have, on average, upon similar students assigned to them for instruction. Clearly, such analyses lend themselves to accountability systems that hold schools or teachers responsible for student achievement.

A weakness of value-added analyses is the difficulty anchoring the normative results with the performance standard criteria on which the accountability system rests. In essence, universal proficiency does not necessarily follow from above average growth (No Child Left Behind *versus* No Child Left Behind on Average). The gap between value-added analyses and standards based accountability systems is simply a reflection of a deeper schism between underlying norms and standards. This paper puts establishes that reconciling this gap involves looking at *both* normative and standards based interpretations simultaneously within a unifying normative framework. Doing so illuminates the interconnections between growth, standards, and accountability:

Growth and Standards

- What rates of student growth are necessary to fulfill universal proficiency mandates

embedded in accountability systems?

- Are rates of student growth necessary to fulfill universal proficiency mandates realistic?

Standards and Accountability

- Given current performance standards, are universal proficiency mandates realistic?
- How should accountability balance the desire for ambitious yet reasonable performance targets for all students?

Growth and Accountability

- Given current rates of student progress is failure to meet universal proficiency mandates the fault of the education system?
- What levels of effectiveness are required of the education system (and the education system in general) to fulfill universal proficiency mandates?

This paper argues that the barriers currently present between growth, standards, and accountability are largely the result of incomplete information: normative data lacking criterion referenced standards or criterion referenced standards lacking a normative basis. This view is consistent with an aphorism attributed to Angoff (1974): “scratch a criterion and you’ll find a norm”. A complete understanding of phenomena associated with large scale assessment results requires *both* a normative and criterion referenced component. The normative framework proposed to unify growth, standards and accountability are *student growth percentiles*.

In contrast to the majority of longitudinal analysis techniques using student assessment data that seek to *explain* the variability of student scores *vis-à-vis* teacher or school effects, the primary concern with calculating student growth percentiles is to *describe* this variability and, following similar descriptions used in pediatrics, give stakeholders a sense of what the current range of student growth is.¹ With this normative vocabulary in place, informed and productive discussions of student growth regarding *what is?*, *what should be?*, and *what is reasonable?* can follow.

An intention of providing this normative backdrop to student progress is to promote a move toward *descriptive* (Linn, 2008) or *regulatory* (Edley, 2006) approaches to accountability and away from the current high stakes accountability systems geared toward assigning blame for success/failure (i.e., establishing the cause). In the introductory chapter to *The Future of Test Based Educational Accountability* Linn describes such a descriptive approach:

Accountability system results can have value without making causal inferences about school quality, solely from the results of student achievement measures and demographic characteristics. Treating the results as descriptive information

¹See http://www.nutropin.com/patient/3_5_4-growth-velocity.jsp for an online implementation of pediatric growth percentiles for height.

and for identification of schools that require more intensive investigation of organizational and instructional process characteristics are potentially of considerable value. Rather than using the results of the accountability system as the sole determiner of sanctions for schools, they could be used to flag schools that need more intensive investigation to reach sound conclusions about needed improvements or judgments about quality (Linn, 2008, p. 21).

Christopher Edley (2006), in his invited presidential address at the 2006 AERA conference expresses similar sentiments:

This is the difference between a retrospective question of identifying fault as opposed to a prospective strategy to engineer some corrective measure, almost independent of considering whether there was blame-worthiness. And to move away from the blame-worthiness paradigm toward something that is more regulatory in nature where one might seize upon disparities or circumstances that are for some reason deemed unacceptable and engineer the interventions needed to bring about the necessary change. ... It's the no-fault gap closing strategy in which the effort is to build a consensus about a vision of an improved society rather than figure out where's the person ... we want to pillory.

As Linn (2008) notes, such an accountability system would represent a profound change from current systems. An essential first step toward such a change is the creation of appropriate and compelling descriptive measures on which to base the system.

Student Growth Percentiles

It is a common misconception that to measure student growth in education, the subject matter and grades over which growth is examined must be on the same scale—referred to as a vertical scale. Not only is a vertical scale not necessary, but its existence obscures fundamental concepts necessary to understand growth. Growth, fundamentally, requires change to be examined for a single construct, like math achievement, over time—*growth in what?* A single scale for the construct is necessary to measure the *magnitude* of growth, but not growth in general (Betebenner, 2008; Yen, 2007).

Consider the familiar situation from pediatrics where the interest is on measuring the height and weight of children over time. The scales on which height and weight are measured possess properties that educational assessment scales aspire towards but can never meet.

An infant male toddler is measured at 2 and 3 years of age and is shown to have grown 4 inches. The magnitude of increase—4 inches—is a well understood quantity that any parent can grasp and calculate at home using a simple yardstick. However, parents leaving their pediatrician's office knowing only how much their child has grown would likely be wanting for more information: Parents are not interested in an absolute magnitude of growth, but instead in a normative criterion locating that 4 inch increase alongside the height increases of

similar children. Examining this height increase relative to the increases of similar children permits a diagnosis of how (ab)normal such an increase is.

Given this reality in the examination of change where scales of measurement are perfect, it is absurd to think that in education, where scales are, at best, quasi-interval, one can/should examine growth differently.

Supposing scales did exist in education similar to height/weight scales that permitted the calculation of absolute measures of annual academic growth for students, the response parents receive to questions such as, “How much did my child progress?”, would come as a number of scale score points—an answer likely to leave most parents bewildered wondering whether the number of points is good or bad. With regard to measures used in education, the search for a description regarding change in achievement over time (i.e., growth) is best conducted in two steps: First by considering a normative quantification of student growth—a student growth percentile. And second by establishing growth adequacy criteria—a percentile growth projection.

A student’s growth percentile describes how (ab)normal a student’s growth is by examining their current achievement relative to their *academic peers*—those students with identical prior achievement. That is, a student growth percentile examines the current achievement of a student relative to other students who have, in the past, “walked the same achievement path”. Heuristically, if the state assessment data set were extremely large (in fact, infinite) in size, one could examine the data set and select out those students with the exact same prior scores and compare how the selected student’s current year score compares to the current year scores of those students with the same prior year’s scores—their academic peers. If the student’s current year score exceeded the scores of most of their academic peers, in a normative sense they have done well. If the student’s current year score was less than the scores of their academic peers, in a normative sense they have not done well.

The four panels of Figure 1 depict what a student growth percentile represents in a situation considering students having only two consecutive achievement test scores.

Upper Left Panel Considering all pairs of scores for all students in the state yields a bivariate (two variable) distribution.

Upper Right Panel Conditioning upon prior achievement fixes a the value of the 2005 scale score (in this case at 600) and is represented by the red slice taken out of the bivariate distribution.

Lower Left Panel Conditioning upon prior achievement defines a *conditional distribution* representing the distribution of outcomes on the 2006 test assuming a 2005 score of 600. This distribution is indicating as the solid red curve.

Lower Right Panel The conditional distribution provides the context within which a student’s 2006 achievement can be understood normatively. Students with achievement in the upper tail of the conditional distribution have demonstrated high rates of growth relative to their academic peers whereas those students with achievement in the lower tail of the distribution have demonstrated low rates of growth. Students with current

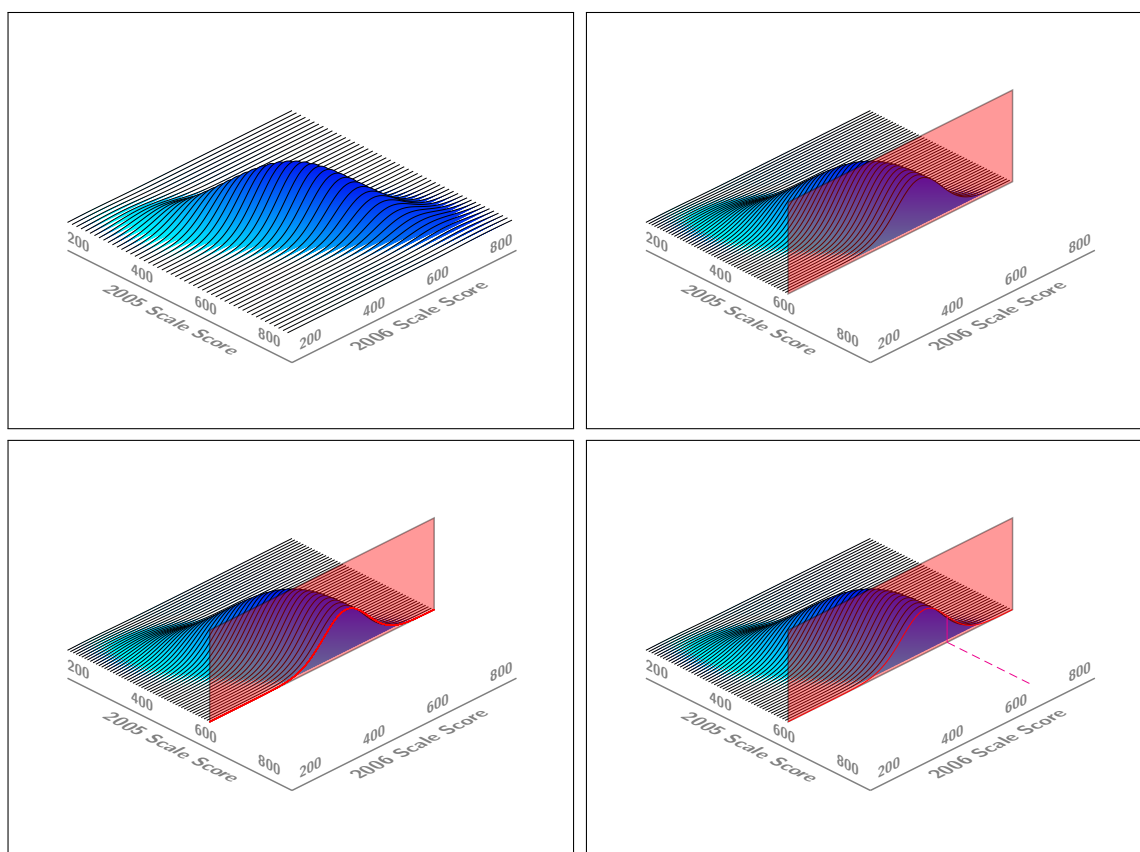


Figure 1: Figures depicting the distribution associated with 2005 and 2006 student scale scores together with the conditional distribution and associated growth percentile

achievement in the middle of the distribution could be described as demonstrating “average” or “typical” growth. In the figure provided the student scores approximately 650 on the 2006 test. Within the conditional distribution, the value of 650 lies at approximately the 70th percentile. Thus the student’s growth from 600 in 2005 to 650 in 2006 met or exceeded that of approximately 70 percent of students starting from the same place. This 50 point increase is above average. It is important to note that qualifying a student growth percentile as “adequate”, “good”, or “enough” is a standard setting procedure requiring stakeholders to examine a student’s growth relative to external criteria such as performance standards/levels.

Figure 1 also illustrates the relationship between a vertical scale and student growth percentiles. Using the vertical scale implied by Figure 1, the student grew 50 points (from 600 to 650) between 2005 and 2006. This 50 points represents the magnitude of change. Quantifying the magnitude of change is scale dependent. However, relative to other students, the achievement growth of the student has not changed—their growth percentile is invariant to scale transformations common in educational assessment. Student growth percentiles normatively situate achievement change bypassing questions associated with the magnitude of change, and directing attention toward relative standing which is likely to interest stakeholders most.

The percentile of a student's current score within their corresponding conditional distribution translates to a probability statement of a student obtaining that score taking account of prior achievement. That is:²

$$\text{Student Growth Percentile} \equiv \Pr(\text{Current Achievement} | \text{Past Achievement}) \cdot 100.$$

Whereas unconditional percentiles normatively quantify achievement, conditional percentiles normatively quantify growth. Because past scores are used solely for conditioning purposes, one of the major advantages of using growth percentiles to measure change is that estimation does not require a vertical scale.

Student Growth Percentile Calculation

Calculation of a student's growth percentile is based upon the estimation of the conditional density associated with a student's score at time t using the student's prior scores at times $1, 2, \dots, t - 1$ as the conditioning variables. Given the conditional density for the student's score at time t , the student's growth percentile is defined as the percentile of the score within the time t conditional density. By examining a student's current achievement with regard to the conditional density, the student's growth percentile normatively situates the student's outcome at time t taking account of past student performance. The percentile result reflects the likelihood of such an outcome given the student's prior achievement. In the sense that the student growth percentile translates to the probability of such an outcome occurring (i.e., rarity), it is possible to compare the progress of individuals not beginning at the same starting point. However, occurrences being equally rare does not necessarily imply that they are equally "good". Qualifying student growth percentiles as "(in)adequate", "good", or as satisfying "a year's growth" is a standard setting procedure requiring external criteria (e.g., growth relative to state performance standards) combined with the wisdom and judgments of stakeholders.

Estimation of the conditional density is performed using quantile regression (Koenker, 2005). Whereas linear regression models the conditional mean of a response variable Y , quantile regression is more generally concerned with the estimation of the family of conditional quantiles of Y . Quantile regression provides a more complete picture of the conditional distribution associated with the response variable(s). The techniques are well suited for estimation of the family of conditional quantile functions (i.e., reference percentile curves). Using quantile regression, the conditional density associated with each student's prior scores is derived and used to situate the student's most recent score. Position of the student's most recent score within this density can then be used to qualify deficient/sufficient/excellent growth. Though many state assessments possess a vertical scale, such a scale is not necessary to produce student growth percentiles.

In analogous fashion to the least squares regression line representing the solution to a minimization problem involving squared deviations, quantile regression functions represent

²Since $\Pr(\text{Current Achievement} | \text{Past Achievement}) \cdot 100$ is not always an integer between 1 and 100 the expression denotes a student growth quantile. To simplify, the result is rounded down and termed a percentile.

the solution to the optimization of a loss function (Koenker, 2005, p. 5). Formally, given a class of suitably smooth functions, \mathcal{G} , one wishes to solve

$$\arg \min_{g \in \mathcal{G}} \sum_{i=1}^n \rho_{\tau}(Y(t_i) - g(t_i)), \quad (1)$$

where t_i indexes time, Y are the time dependent measurements, and ρ_{τ} denotes the piecewise linear loss function defined by

$$\rho_{\tau}(u) = u \cdot (\tau - I(u < 0)) = \begin{cases} u \cdot \tau & u \geq 0 \\ u \cdot (\tau - 1) & u < 0. \end{cases}$$

The elegance of the quantile regression Expression 1 can be seen by considering the more familiar least squares estimators. For example, calculation of $\arg \min \sum_{i=1}^n (Y_i - \mu)^2$ over $\mu \in \mathbb{R}$ yields the sample mean. Similarly, if $\mu(x) = x'\beta$ is the conditional mean represented as a linear combination of the components of x , calculation of $\arg \min \sum_{i=1}^n (Y_i - x'_i\beta)^2$ over $\beta \in \mathbb{R}^p$ gives the familiar least squares regression line. Analogously, when the class of candidate functions \mathcal{G} consists solely of constant functions, the estimation of Expression 1 gives the τ th sample quantile associated with Y . By conditioning on a covariate x , the τ th conditional quantile function, $Q_y(\tau|x)$, is given by

$$Q_y(\tau|x) = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - x'_i\beta).$$

In particular, if $\tau = 0.5$, then the estimated conditional quantile line is the median regression line.³

Following Wei & He (2006), we parameterize the conditional quantile functions as a linear combination of B-spline cubic basis functions.. B-splines are employed to accommodate non-linearity, heteroscedasticity and skewness of the conditional densities associated with values of the independent variable(s). B-splines are attractive both theoretically and computationally in that they provide excellent data fit, seldom lead to estimation problems (Harrell, 2001, p. 20), and are simple to implement in available software.

Figure 2 gives a bivariate representation of linear and B-splines parameterization of decile conditional achievement curves. Note that the assumption of linearity imposes conditions upon the heteroscedasticity of the conditional densities. Close examination of the linear deciles indicates slightly greater variability for higher grade 5 scale scores than for lower scores. By contrast, the B-spline based decile functions better capture the greater variability at both ends of the scale score range together with a slight, non-linear trend to the data.

Calculation of student growth percentiles is performed using R (R Development Core Team, 2009), a software language and environment for statistical computing, with the SGP package (Betebenner, 2009). Other possible software (untested with regard to student growth percentiles) with quantile regression capability include SAS and Stata. Estimation of student growth percentiles is conducted using all available prior data, subject to certain suitability

³For a detailed treatment of the procedures involved in solving the optimization problem associated with Expression 1, see Koenker (2005), particularly Chapter 6.

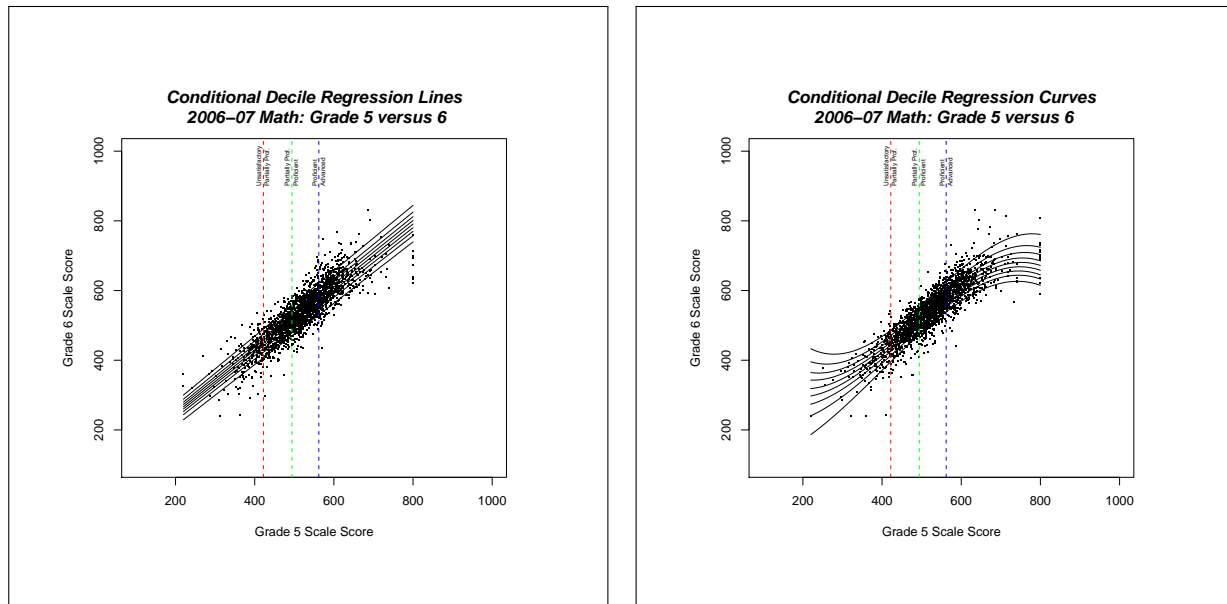


Figure 2: Linear and B-spline conditional deciles based upon bivariate math data, grades 5 and 6

conditions. Given assessment scores for t occasions, ($t \geq 2$), the τ -th conditional quantile for Y_t based upon $Y_{t-1}, Y_{t-2}, \dots, Y_1$ is given by

$$Q_{Y_t}(\tau|Y_{t-1}, \dots, Y_1) = \sum_{j=1}^{t-1} \sum_{i=1}^3 \phi_{ij}(Y_j) \beta_{ij}(\tau), \quad (2)$$

where $\phi_{i,j}$, $i = 1, 2, 3$ and $j = 1, \dots, t-1$ denote the B-spline basis functions. Currently, bases consisting of 7 cubic polynomials are used to “smooth” irregularities found in the multivariate assessment data. A bivariate rendering of this is found in Figure 2 where linear and B-spline conditional deciles are presented. The B-spline basis models the heteroscedasticity and non-linearity of the data to a greater extent than is possible using a linear parameterization.

Percentile Growth Projections/Trajectories

Operational work calculating student growth percentiles with state assessment data yields a large number of coefficient matrices derived from estimating Equation 2. These matrices, similar to a lookup table, “encode” the relationship between prior and current achievement scores for students in the norming group (usually an entire grade cohort of students for the state) across all percentiles and can be used both to qualify a student’s current level growth as well as predict, based upon current levels of student progress, what different rates of growth (quantified in the percentile metric) will yield for students statewide.

When rates of growth necessary to reach performance standards are investigated, such calculations are often referred to as “growth-to-standard”. These analyses serve a dual purpose in that they provide the growth rates necessary to reach these standards and also shed light on the standard setting procedure as it plays out across grades. To establish growth per-

centiles necessary to reach different performance/achievement levels, it is necessary to investigate what growth percentile is necessary to reach the desired performance level thresholds based upon the student's achievement history.

Establishing criterion referenced growth thresholds requires consideration of multiple future growth/achievement scenarios. Instead of inferring that prior student growth is indicative of future student growth (e.g., linearly projecting student achievement into the future based upon past rates of change), predictions of future student achievement are contingent upon initial student status (where the student starts) and subsequent rates of growth (the rate at which the student grows). This avoids fatalistic statements such as, "Student *X* is projected to be (not) proficient in three years" and instead promotes discussions about the different rates of growth necessary to reach future achievement targets: "In order that Student *X* reach/maintain proficiency within three years, she will have to demonstrate *nth* percentile growth consecutively for the next three years." The change in phraseology is minor but significant. Stakeholder conversations turn from "where will (s)he be" to "what will it take?"

Percentile Growth Projection/Trajectory Calculation

Parallel growth/achievement scenarios are more easily understood with a picture. Using the the results of a statewide assessment growth percentile analyses, Figures 3 and 4 depict future growth scenarios in math and reading, respectively, for a student starting in third grade and tracking that student's achievement timeline based upon different rates of annual growth expressed in the growth percentile metric. The figures depict the four state achievement levels across grades 3 to 10 in color together with the 2007 achievement percentiles (inner most vertical axis) superimposed in white. Beginning with the student's achievement starting point at grade 3 a grade 4 achievement projection is made based upon the the most recent growth percentile analyses derived using prior 3rd to 4th grade student progress. More specifically, using the coefficient matrices derived in the quantile regression of grade 4 on grade 3 (see Equation 2), predictions of what 10th, 25th, 40th, 50th, 60th, 75th, and 90th percentile growth lead to are calculated. Next, using these six projected 4th grade scores combined with the student actual 3rd grade score, 5th grade achievement projections are calculated using the most recent quantile regression of grade 5 on grades 3 and 4. Similarly, using these six projected 5th grade scores, the 6 projected 4th grade scores with the students actual third grade score, achievement projections to the 6th grade are calculated using the most recent quantile regression of grade 6 on grades 3, 4, and 5. The analysis extends recursively for grades 6 to 10 yielding the *percentile growth trajectories* in Figures 3 and 4. The figures allow stakeholders to consider what consecutive consecutive rates of growth, expressed in growth percentiles, yield for students starting at different points.

Figure 3 depicts percentile growth trajectories in mathematics for a student beginning at the unsatisfactory/partially proficient threshold. Based upon the *achievement* percentiles depicted (the white contour lines), approximately 7 percent of the population of 3rd graders rate as unsatisfactory. Moving toward grade 10, the percentage of unsatisfactory students increases dramatically to near 35 percent. The black lines in the figure represent six different growth scenarios for the student based upon consecutive growth at a given growth per-

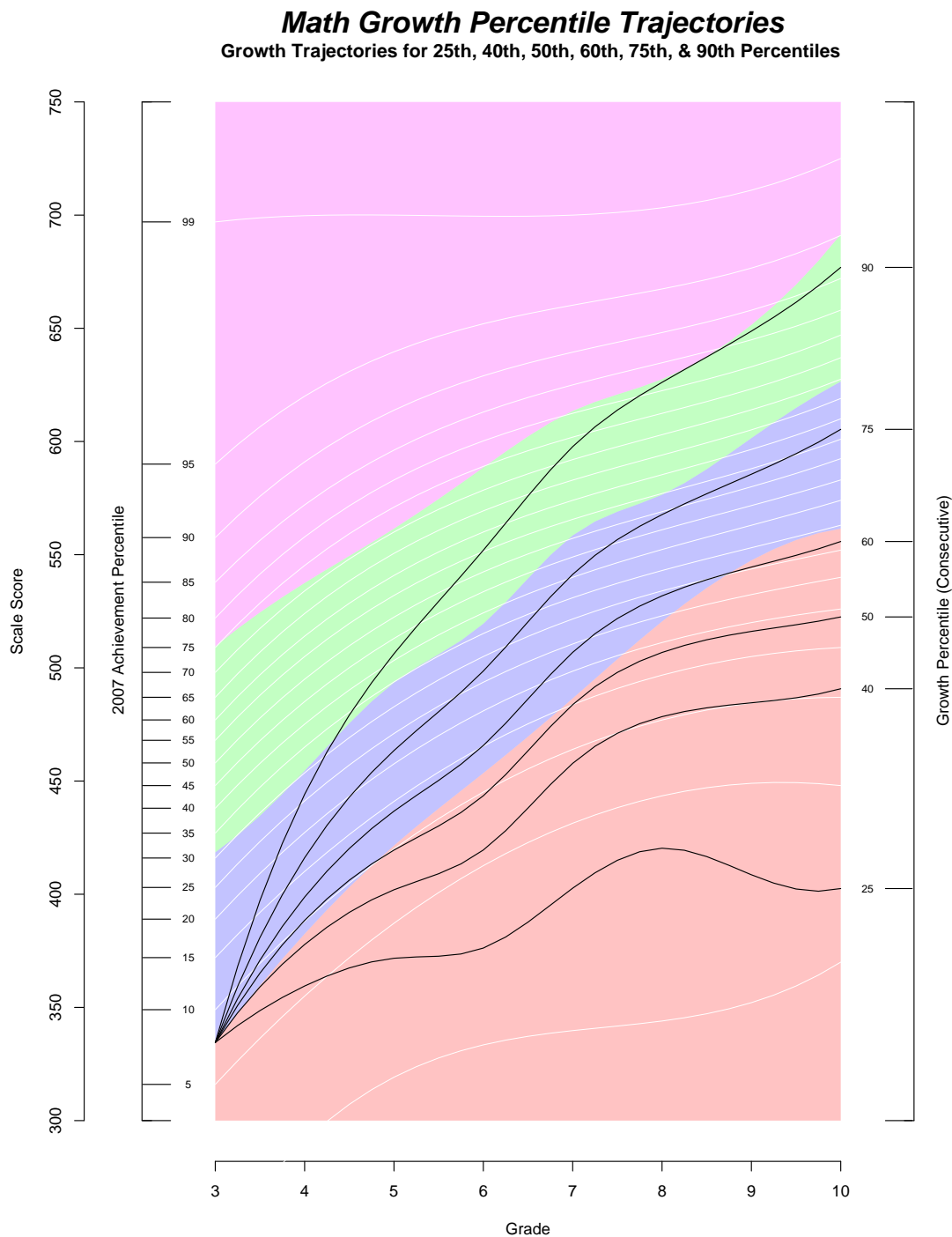


Figure 3: Growth chart depicting future mathematics achievement conditional upon consecutive 25th, 40th, 50th, 60th, 75th, and 90th percentile growth for a student beginning the third grade at the unsatisfactory/partially proficient cutpoint

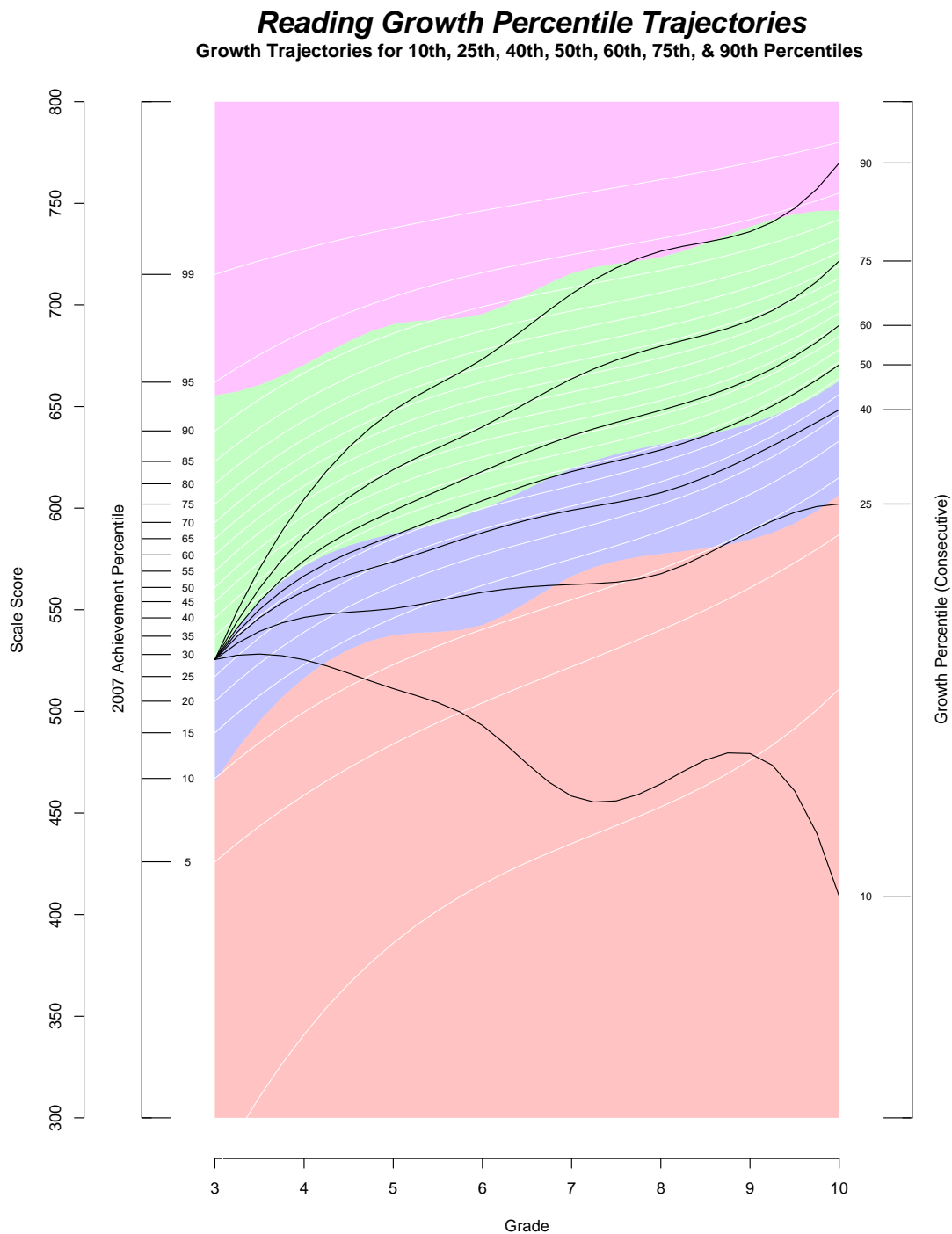


Figure 4: Growth chart depicting future reading achievement conditional upon consecutive 10th, 25th, 40th, 50th, 60th, 75th, and 90th percentile growth for a student beginning the third grade at the partially proficient/proficient cutpoint

centile, denoted by the right axis. At the lower end, for example, consecutive 25th percentile growth leaves the student, unsurprisingly, mired in the unsatisfactory category. Consecutive 40th, 50th and 60th percentile growth also leave the child in the unsatisfactory category. This demonstrates how difficult (probabilistically, based upon current rates of progress) it is for students to move up in performance level in math statewide. With the green region representing proficient, a student would need to demonstrate growth percentiles consecutively in excess of 75 to reach proficiency showing how unlikely such a event currently is. In light of NCLB universal proficiency mandates, the growth necessary for non-proficient students to reach proficiency, absent radical changes to growth rates of students statewide, is likely unattainable for a large percentage of non-proficient students.

Figure 4 depicts percentile growth trajectories in reading for a student beginning at the partially proficient/proficient threshold in grade 3. In a normative sense, the performance standards in reading are less demanding than those in mathematics (particularly in the higher grades) with approximately 30 percent of students below proficient in grades 3 to 10. The black lines in the figure represent seven growth scenarios for the hypothetical student based upon consecutive growth at a the given growth percentile. Compared with the growth required in mathematics, more modest growth is required to maintain proficiency. Typical growth (50th percentile growth) appears adequate for such a student to move up slightly into the proficiency category.

Discussion

Having established a normative framework for understanding student growth and anchored that framework to the performance standards used in current accountability systems, addressing issues/questions related to growth, accountability and standards is straightforward:

Growth and Standards

- What rates of student growth are necessary to fulfill universal proficiency mandates embedded in accountability systems?
- Are rates of student growth necessary to fulfill universal proficiency mandates realistic?

Clearly, the rates of student growth necessary to reach or maintain proficiency differ by student based upon the current level of achievement for the student. As the achievement level of the student decreases, the rate of growth necessary for them to “catch-up” increases. Figures 3 and 4 indicate the rates of growth necessary in mathematics and reading for third graders beginning 3rd grade at two different levels of achievement to reach various targets.. Similar figures are possible for every student in the education system. Along these lines, the state of Colorado currently produces student reports (see Figures 5, 6 and 7 on Pages 18 to 20) for parents and teachers showing individualized historical rates of growth as well as anticipated rates of growth necessary to reach future achievement goals.

One of the strengths of quantifying student growth normatively is that the growth percentile targets that are calculated of what it will take to reach a level of achievement quickly translate into the likelihood of such an event occurring. This dimension of student progress

as it relates to accountability is absent from most growth-to-standard discussions. Today, achievement mandates are stipulated based upon the moral imperative of high standards for all children with little concern regarding the likelihood the students reaching these goals. Given current progress of students, it is unlikely that the sustained levels of growth necessary to reach these standards will occur without substantial changes to both the education system and society as a whole.

Standards and Accountability

- Are universal proficiency mandates realistic in terms of current performance standards?
- How should accountability balance the desire for ambitious yet reasonable performance targets for all students?

The difference in growth required to maintain or progress in mathematics and reading in Figures 3 and 3 is not surprising once one examines the normative stringency of the performance standards across grades. In mathematics, for example, with decreasing percentages of students reaching partially proficient, proficient, and advanced in higher grades, it is a tautology that high normative rates of growth will be required to maintain or move up. Similar, with near constant percentages of students at or above proficient in reading, typical rates of growth will be sufficient to maintain. With increasing percentages of students reach performance standards in higher grades, below typical rates of growth would be needed for students to “catch-up”.

Anchoring growth to achievement goals normatively adds the prospect of making adequacy judgments to the normative measures that are used. If standards based accountability policy demands universal proficiency, then for each student it is possible to determine what growth percentile is necessary to reach proficiency in 1 or more years. Accountability policy mandates can be immediately scrutinized in terms of reasonableness. If the growth percentiles indicate an unreasonable expectation then realization of these goals will likely require either policy expectation or the standards on which they are based to change. The metric allows for clear discussions regarding the demands being made.

Growth and Accountability

- Given current student growth rates is failure to meet universal proficiency mandates the fault of the education system?
- What levels of effectiveness are required of the education system (and the education system in general) to fulfill universal proficiency mandates?

The question of responsibility for (lack of) student growth is extremely difficult. Viewed normatively, this is the question value-added models address. When anchored to standards, the question becomes as much philosophical as statistical. For thousands of students in state education systems, the consecutive growth percentiles necessary for these students to reach proficiency exceed 90. No school is systemically producing these rates of growth for its average student. Is it fair to place responsibility for failure to reach these impossible goals on schools and the educators working within them? A fundamental dictum of moral philosophy ascribed to Kant is that “ought implies can”: If someone *ought* to do something, they

can do it, in the sense that they have the possibility/capacity to do it. Growth percentiles bring Kant's dictum to the fore when considering performance standards, the likelihood of students reaching/maintaining these standards, and accountability systems holding stakeholders responsible for student failure to reach/maintain these standards.

The reality of the present needn't define a blueprint for the future. A primary goal of education reform is to turn today's exemplary growth into tomorrow's typical growth. Achieving that would bring us closer to the goal of universal proficiency engendered by NCLB. To that end, this paper introduces student growth percentiles as a quantification of student progress and a descriptive measure of *what is?*. Criterion referenced questions of *what should be?* coincide with decisions about whether growth is "enough" or "adequate" to reach or maintain desired levels of achievement. Student growth percentiles and percentile growth projections/trajectories serve to inform the standard setting procedure by communicating *what is reasonable?*. Only by considering, what is, what should be, and what is reasonable simultaneously, can accountability systems, built upon performance metrics like student growth percentiles, be equitable, just, and truly informed.

References

- Angoff, W. H. (1974). Criterion-referencing, norm-referencing and the sat. *College Board Review*, 92, 2–5.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment for teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37–65.
- Betebenner, D. W. (2008). Toward a normative understanding of student growth. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 155–170). New York: Taylor & Francis.
- Betebenner, D. W. (2009). SGP: Student growth percentile and percentile growth projection/-trajectory functions [Computer software manual]. (R package version 0.0-4)
- Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models* (Tech. Rep.). Princeton, New Jersey: Educational Testing Service.
- Braun, H. I. (2008). *Viccissitudes of the validators*. Presentation made at the 2008 Reidy Interactive Lecture Series, Portsmouth, NH, September, 2008.
- Edley, C. (2006, April 10). *Educational "Opportunity" is the highest civil rights priority. so what should researchers and lawyers do about it?* [Video of speech with slides posted on the World Wide Web]. (Retrieved June 22, 2006 from the World Wide Web: <http://www.softconference.com/MEDIA/WMP/260407/#43.010>)
- Harrell, F. E. (2001). *Regression modeling strategies*. New York: Springer.
- Koenker, R. (2005). *Quantile regression*. Cambridge: Cambridge University Press.
- Linn, R. L. (2003, July). *Accountability: Responsibility and reasonable expectations* (Tech. Rep.). Los Angeles, CA: Center for the Study of Evaluation, CRESST.
- Linn, R. L. (2008). Educational accountability systems. In *The future of test-based educational accountability* (pp. 3–24). New York: Taylor & Francis.
- 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.
- R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from <http://www.R-project.org> (3-900051-07-0)
- Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29(1), 121–129.
- Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, 29(1), 103–116.
- Spellings, M. (2005, Nov). *Secretary Spellings announces growth model pilot* [Press Release]. U.S. Department of Education. (Retrieved August 7, 2006 from <http://www.ed.gov/news/pressreleases/2005/11/1182005.html>)
- Wei, Y., & He, X. (2006). Conditional growth charts. *The Annals of Statistics*, 34(5), 2069–2097.
- Yen, W. M. (2007). Vertical scaling and No Child Left Behind. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 273–283). New York: Springer.

Individual Growth and Achievement Reports

The individual growth and achievement reports that follow are presented for reading, writing, and mathematics. The reports show an achievement timeline for the student depicting historical achievement and growth data together with projections of what low, typical and high growth will lead to. The reports use the state achievement levels as a backdrop to convey to stakeholders the absolute criteria against which student progress is judged. The figures are intended to assist in better understanding the range of student growth and what different growth rates lead to in terms of student achievement relative to state designated performance performance levels.

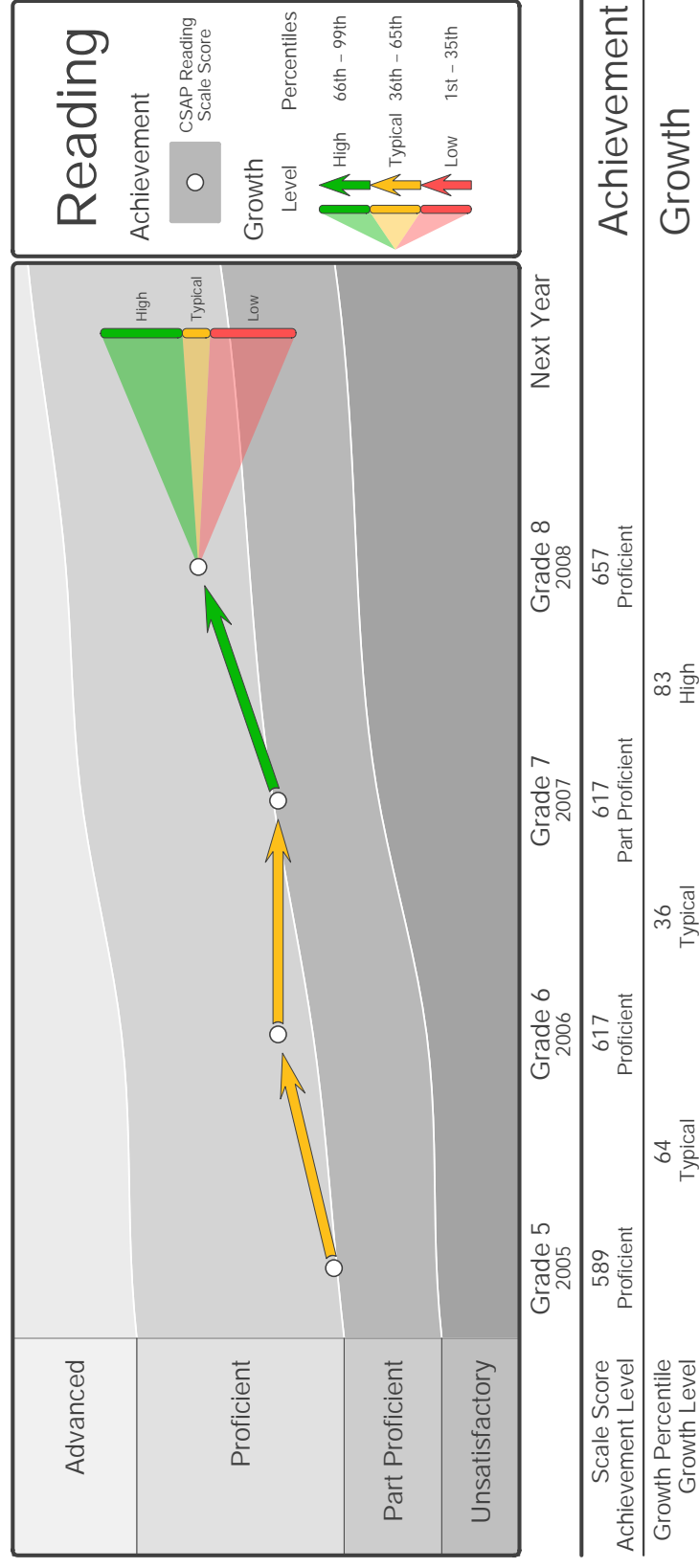


Figure 5: Individual growth and achievement report in reading depicting historical achievement and growth together with growth projections.

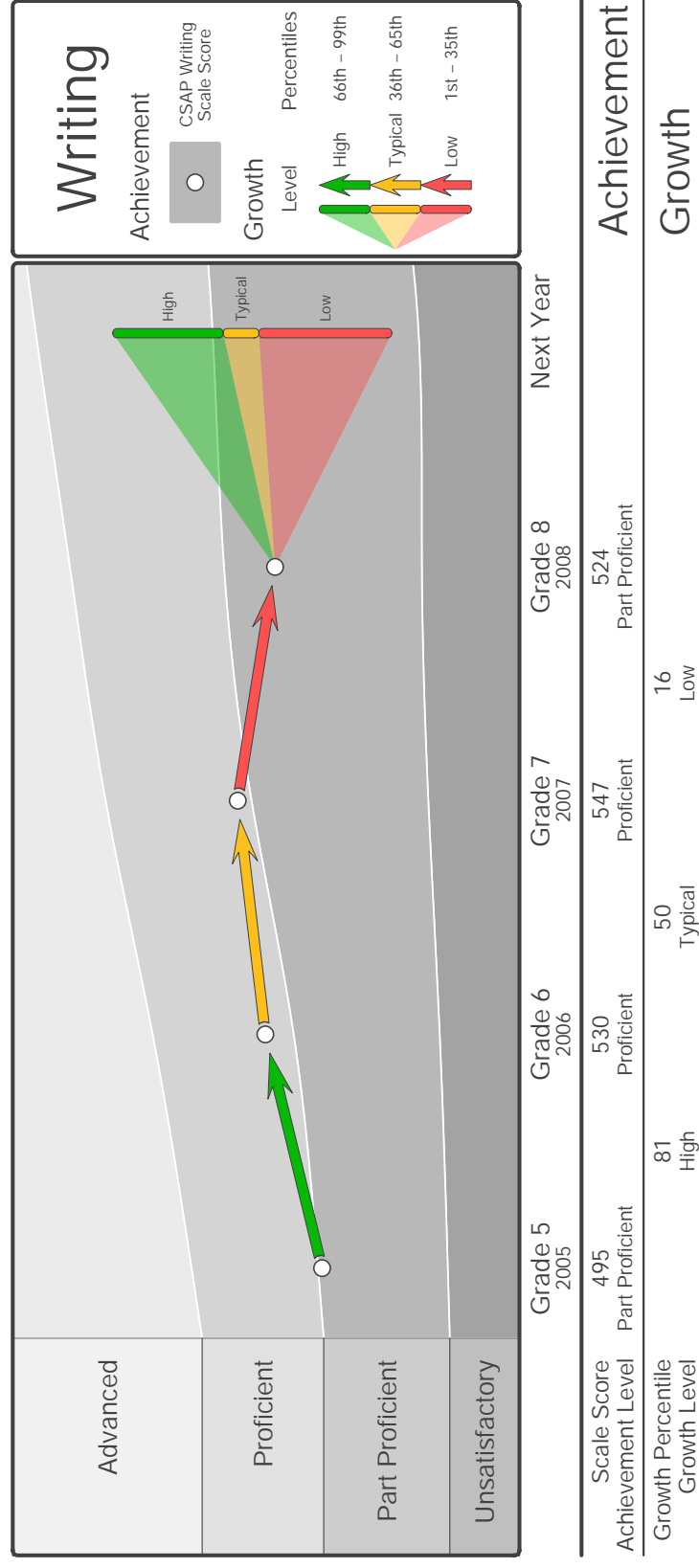


Figure 6: Individual growth and achievement report in writing depicting historical achievement and growth together with growth projections.

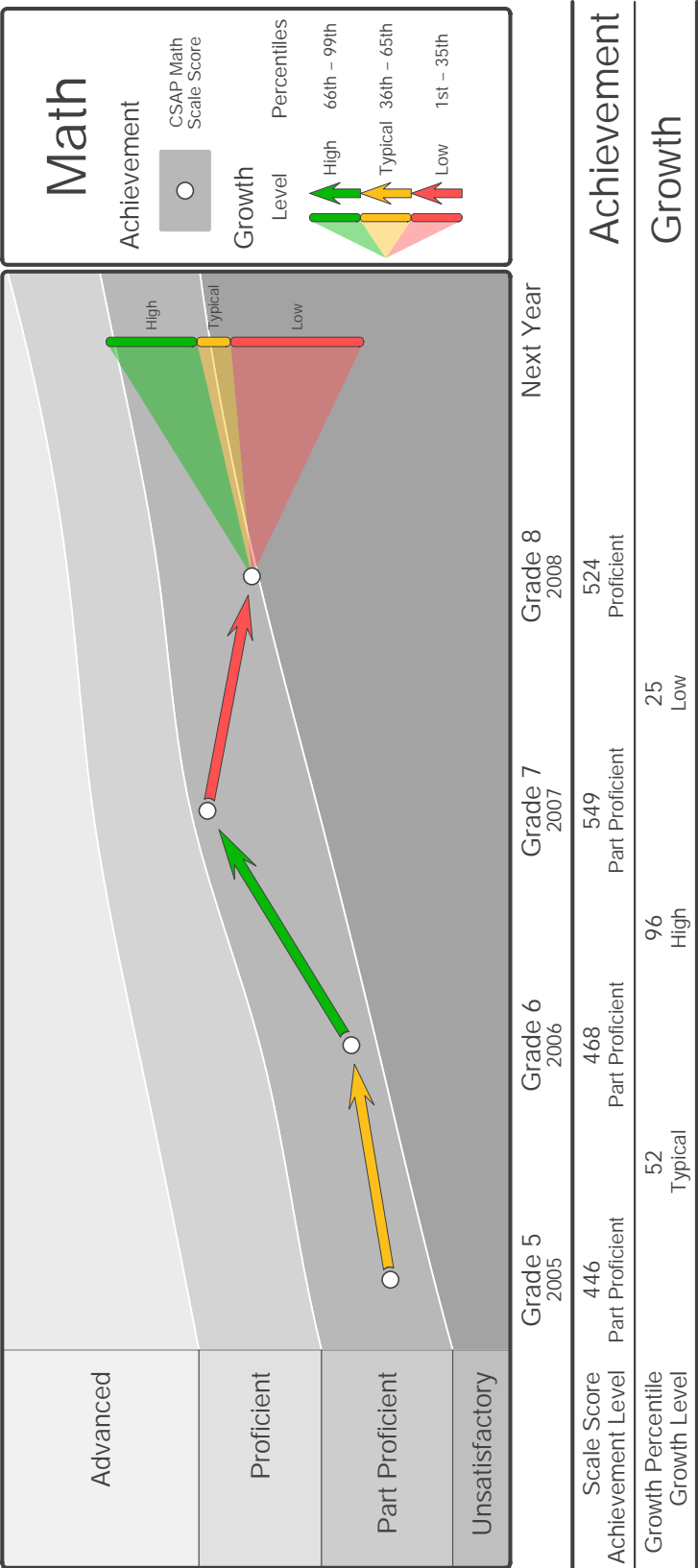


Figure 7: Individual growth and achievement report in mathematics depicting historical achievement and growth together with growth projections.