An Overview of Student Growth Percentiles

Damian W. Betebenner
National Center for the Improvement of Educational Assessment
Dover, New Hampshire
DBetebenner@nciea.org

August 12, 2011

Why Student Growth?

Following the 2001 reauthorization of the *Elementary and Secondary Education Act of 1965*, commonly known as the *No Child Left Behind Act* (NCLB), states were required to implement large scale testing of all students to an extent never before seen in the United States (No Child Left Behind Act of 2001 (NCLB), 2002). Starting with the 2005-2006 school year, NCLB required states to test students in reading and mathematics from grades 3 through 8 and at least once in grades 10 through 12. In addition, beginning with the 2007-2008 school year, states must assess students in science at least once in elementary, middle and high school. As a consequence, states find themselves buried in assessment data with mandates from state and federal policy makers to utilize the data to improve the quality of education.

Accountability systems constructed according to federal adequate yearly progress (AYP) requirements currently rely upon annual measurement of student *achievement* to make judgments about school quality. Since their adoption, such *status measures* have been the focus of persistent criticism (Linn, 2003; Linn, Baker, & Betebenner, 2002). Status measures, though appropriate for making judgments about the achievement level of students at a school for a given year, 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 NCLB to label its achievement mandates as unfair and misguided and to demand the use of growth analyses as a better means of auditing the quality of schools.

More recently, the Race to the Top program has expanded the role of large scale testing even further in accountability by mandating the use of student outcomes in the evaluation of teachers and principals. A fundamental premise associated with the use of student growth for accountability is that "good" schools, teachers, or principals, bring about student growth in excess of that found with "bad" schools, teachers, or principals. 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 personnel within the school. The inherent believability of this premise is at the heart of current enthusiasm to incorporate growth models into state accountability systems. It is not surprising that the November 2005 announcement by Secretary of Education Spellings for the Growth Model Pilot Program permitting states to use growth model results as a means for compliance with NCLB achievement mandates and the Race to the Top competitive grants program were met with great enthusiasm by states. (Spellings, 2005).

Following these use cases, the primary thrust of growth analyses over the last decade has been to determine, using sophisticated statistical techniques, the amount of student progress/growth that can be justifiably attributed to the school or teacher—that is, to disentangle current *aggregate* level achievement from effectiveness (Braun, 2005; Rubin, Stuart, & Zanutto, 2004; Ballou, Sanders, & Wright, 2004; Raudenbush, 2004). Such analyses, often called *value-added* analyses, attempt to estimate the teacher/school contribution to student achievement. This contribution to student achievement, called the *school* or *teacher effect*, purports to quantify the impact on achievement that this school or teacher would have 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. Despite this utility, however, such analyses fail to address one of the fundamental questions concerning the growth of students: How much growth did a student make?

This primer returns to the task of quantifying change in student achievement. To do so, we introduce a normative quantification of individual student growth which we call a *student growth percentile* (Betebenner, 2008, 2009). These quantities, derived using quantile regression techniques, are easily interpretable descriptive statistics that permit growth comparisons between all students regardless of the scale students are measured on.[1]. In addition, growth percentiles can be criterion referenced *vis-á-vis* current growth-to-standard methodologies in order to establish qualifications of what is *enough* growth. The purpose of this primer is to introduce student growth percentiles and to demonstrate, using state assessment data, how current discussions of student growth lack a normative dimension necessary to set challenging yet attainable individual achievement goals. We assert that the establishment of a normative basis for student growth eliminates a number of the problems of incorporating growth into accountability systems by providing needed insight to various stakeholders by answering the simple question of how much a student progressed.

Student Growth Percentiles

It is a common misconception that to quantify student progress 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 concepts necessary to fully understand student growth. Growth, fundamentally, requires change to be examined for a single construct like math achievement across time—*growth in what?*.

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.[2]

[1]More specifically, the technique does not require a vertical scale in order to calculate student growth percentiles

[2]Height and weight scales are interval (actually, ratio scales) where a unit increase reflects an equivalent increase in the underlying quality being measured no matter where on the scale the increase occurs.

> 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. In this situation, parents are not so much interested in a magnitude of growth, but instead in a norm-referenced understanding of that magnitude that locates that 4 inch increase alongside the height increases of similar children. Examining this height increase relative to the increases of similar children permits one to diagnose how (ab)normal such an increase is.

Given this reality in the examination of change where scales of measurement are perfect, we argue that it is absurd to think that in education, where scales are quasi-interval, one can/should examine growth differently.[3]

Going further, suppose that 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 to a parent's question such as, "How much did my child progress?", would be a number of scale score points—an answer that would leave most parents confused wondering whether the number of points is good or bad. As in pediatrics, the search for a description regarding changes in achievement over time (i.e., growth) is best served by considering a norm-referenced quantification of student growth—*a student growth percentile*.

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 beginning at the same place. 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 open the infinite 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 2010 and 2011 scores for all students in the state yields a bivariate (two variable) distribution. The higher the distribution, the more frequent the pair of scores.

Upper Right Panel Taking account of prior achievement (i.e., conditioning upon prior achievement) fixes a the value of the 2010 scale score (in this case at 600) and is represented by the red slice taken out of the bivariate distribution.

[3] The scales on which students are measured are often assumed to possess properties similar to height and weight but they don't. Specifically, scales are assumed to be interval where it is assumed that a difference of 100 points at the lower end of the scale refers to the same difference in ability/achievement as 100 points at the upper end of the scale. This assumption, however, fails to hold.

Lower Left Panel Conditioning upon prior achievement defines a *conditional distribution* which represents the distribution of outcomes on the 2011 test assuming a 2010 score of 600. This distribution is indicating as the solid red slice of the distribution.

Lower Right Panel The conditional distribution provides the context against which a student's 2011 achievement can be examined and provides the basis for a norm-referenced comparison. 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 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 2011 test. Within the conditional distribution, the value of 650 lies at approximately the 75th percentile. Thus the student's growth from 600 in 2010 to 650 in 2011 met or exceeded that of 75 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 that requires stakeholders to examine a student's growth *vis-à-vis* external criteria such as performance standards/levels.

Figure 1 also serves to illustrate 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 2010 and 2011. This 50 points represents the absolute magnitude of change. Quantifying the magnitude of change is scale dependent. For example, different vertical achievement scales in 2010 and 2011 would yield different annual scale score increases: A scale score increase of 50 could be changed to a scale score increase of 10 using a simple transformation of the vertical scale on which all the students are measured. However, relative to other students, their growth 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, we would assert, is what stakeholders are most interested in.

Group Level Results

An advantage of quantifying growth at the student level is that it is generally an easy task to combine the individual level growth results to retrieve a group level aggregates. For example, after growth percentiles are calculated for each of 500 students at a school, the distribution of growth percentiles for those 500 students represents how much the students at that school grew in the previous year. Summarizing this distribution's "average" would supply a single number describing the growth of a "typical" student at that school. Because it is innapropriate to calculate a typical average of a set of percentiles, the median is used as the single number which best describes where the middle of the distribution of student growth percentiles lies. It is important to note, however, that the median is a *summary* measure of the growth of *many* students at the school. In reality at almost all schools one can find students who grow slowly—students with low growth percentiles—as well as students who grow quickly—students with high growth percentiles. There is room for improvement in all

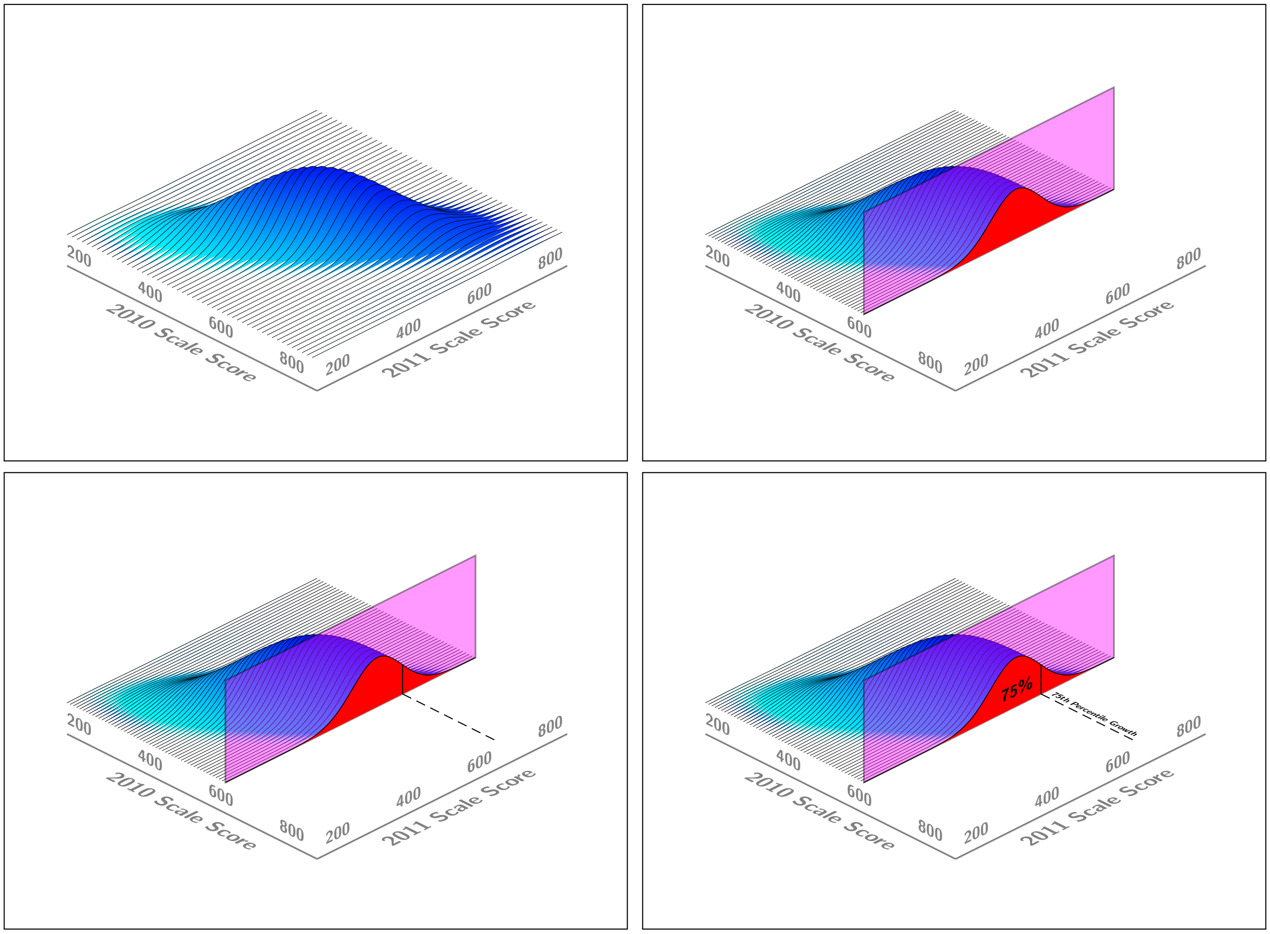

Figure 1: Figures depicting the distribution associated with 2010 and 2011 student scale scores together with the conditional distribution and associated growth percentile

schools.

In the previous example we considered a group being a school but nothing prevents us from considering groups of students in a single classroom, a demographic subgroup, or students receiving the same curricular program. If students were randomly assigned to schools, then the median growth percentile for a school is expected to be 50. Schools with median student growth percentiles above 50 have students demonstrating, on average, greater than typical (i.e., 50th percentile) growth. And schools with median student growth percentiles below 50 have students demonstrating, on average, less than typical growth. In this way, student growth percentiles can be used to identify schools where student growth is extraordinarily good and bad. And like with student achievement, it would be incorrect to assert that the school is solely responsible for the growth of its students.

Measurement of student growth and assignment of responsibility for that growth involves answering two distinct but related questions:

- How much the the students at this school grow/progress?
- How much did this school contribute to student growth?

The median student growth percentile is descriptive and makes no assertion about the cause of student growth. This differs from current value-added models where the purpose is to specify the contribution to student achievement provided by a given school or teacher. It is likely, and society would certainly like to believe, that schools and teachers have a significant impact upon student learning: That their efforts are reflected in the academic progress of students. That is certainly why we summarize student growth for the groups we think are most relevant to the academic outcomes of students. The median student growth percentile is *one of many* indicators that stakeholders can use to judge the quality of the education students receive. It is hoped the as growth percentiles become more widely available, stakeholders will carefully use this piece of data in combination with other data assess both the progress of the student as well as the factors contributing to their progress.

References

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). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42–51.

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models* (Tech. Rep.). Princeton, New Jersey: Educational Testing Service.

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., 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.

No Child Left Behind Act of 2001, Public Law No. 107-110, 115 Stat. 1425.

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)