SCHOOLview.org

cde Improving Academic Achievement
Colorado Department of Education

The Colorado Growth Model

Using Norm- and Criterion-Referenced Growth Calculations to Ensure that All Students are Held to High Academic Standards

William J. Bonk, Ph.D.

Unit of Accountability and Improvement
Colorado Dept. of Education
201 E. Colfax Ave.
Denver, CO 80203
303.866.6763 voice

www.cde.state.co.us
www.schoolview.org

The ideas and analyses presented in this paper are in not attributable entirely to the author. The work of Damian Betebenner, Rich Wenning, Marie Huchton, and others fundamentally underpin the contents of this manuscript, although those sources are not specifically cited.

It is not enough to have a growth model, or an accountability system, or a set of consequences, or standard reporting of performance data, or a common tool for improvement planning, if they only exist in isolation. The integration of these components forms a *systems approach* to accountability and improvement, which is precisely what Colorado has attempted to implement as its overall accountability strategy. The following describes how growth is used as an integral part of Colorado's systems approach to accountability and improvement, specifically assessing if individual students are on track to meeting proficiency standards and are therefore making adequate progress toward college and career readiness.

Colorado's school and district accountability system is based in large part on growth because achievement status calculations alone are not enough to gauge school and district effectiveness. It is essential to know about the amount of learning taking place among all students, as well as whether that amount of learning is likely to lead to acceptable levels of achievement before it becomes too late for action. Federal accountability is not enough – states need to be responsible for the performance of students in all schools, Title I and non-Title I alike. Colorado has designed and adopted a comprehensive accountability system because simply gathering and reporting performance data is not enough – those elements must fit together in a coherent manner, with community-supported weightings to tell the complete story in a consistent way for every school in the state, and to lead to the desired actions to bring about improvement.

Also fundamental to Colorado's approach is the recognition that, in order to close persistent achievement gaps, observed growth needs to be significantly higher for historically disadvantaged groups. Achievement gaps can be the end result of multiple years of lower achievement and growth for impacted students, so growth will be a leading indicator of a situation in which gaps are closing for particular student groups, even if it they are not projected to reach proficiency until they are in another school in a higher grade. Colorado's accountability system looks specifically at the growth of these disaggregated subgroups to assess whether or not that growth was sufficient to get these students to college and career readiness in time.

The Colorado Growth Model produces both norm- and criterion-referenced data that are best understood together. The norm-referenced information is invaluable because it provides a consistent context in which to understand performance, along with achievement status. However, in addition to the normative output, by anchoring Colorado Growth Model calculations to the assessments' proficiency cut scores (which do not change from year to year), the model can also be used to assess whether the growth students are making is enough to get them to the destination in time: growth to a standard, with a consistent criterion for all students. This criterion-referenced use of the growth data is essential to a state's assessment of whether or not students are on track to attain and maintain proficiency. It reminds us not to accept anything less than college and career readiness for all students. How is this done?

As the growth model is constructed with multiple years of the state's data, it accumulates a general understanding of the likelihood of patterns of performance. This translates into an ability to consider hypothetical scenarios, such as "A student scoring x, y and z in grades 3, 4 and 5 in reading would like to reach the level of Advanced by grade 8 in 2014. How much growth would she need to achieve for this to happen? Answer: the *n*th percentile, sustained over each of the next three years." It should be emphasized that these are not predictions per se; they are rather calculations that flow from positing one piece of the scenario, and requesting model output for the other. In Colorado this aspirational level of individual student growth is referred to as Adequate Growth Percentiles (AGPs).

AGP values are reported to Colorado schools and districts along with the rest of the growth data for their students. Districts interacting most deeply with the data have found them to be useful in conversations around the setting of individual goals for students, especially those far behind in terms of their proficiency. In other words, the effort shifts from "Let's look at the students of _____ race/ethnicity" to "Let's devise a comprehensive set of interventions for students that are far behind grade level." Using this growth-to-a-standard serves as somewhat of a reality check in terms of how much effort will be required to get a student to proficiency within three years or by exit. If exceptional levels of growth are required, then an exceptional intervention is necessary. When this fact becomes widely understood by stakeholders, an opportunity is created to marshal a consensus for change.

A simple test of the validity of AGPs is to determine whether calculating them offers any advantage over not doing so. Colorado conducted this test using two cohorts of historical data for each content area. The simple model predicts that students already scoring at the proficient level in a given content area will continue to do so through the final year of the data, and that those scoring below proficient will not attain proficiency within the timeframe. Those predictions were then checked against what actually happened with those students to get a sense of the accuracy of the base rate prediction—the percentage of the predicted outcomes that actually came true several years later. The AGP-based prediction, on the other hand, uses the statistical power of the Colorado Growth Model to look at score history and the most recent growth for each student in order to estimate whether or not a student is on track to catch up (starting out below proficient) or keep up (staying proficient). These AGP-based predictions are also compared against actual data (what really happened with those students) to arrive at a percentage of correct predictions. A summary of results is included in the table below.

Correct Predictions of Proficiency Level using Prior Achievement versus Adequate Growth Calculations

		Percentage of correct predictions (using only prior proficiency level)	Percentage of correct predictions (using AGPs)	Improvement in percentage of correct predictions
Math	Below proficient	77.7	88.6	10.9
	Proficient	58.2	75.5	17.2
Reading	Below proficient	55.8	76.2	20.5
	Proficient	78.1	82.6	4.4
Writing	Below proficient	56.4	78.8	22.4
	Proficient	68.7	78.7	9.9

Merely using the simple prior proficiency model yields moderately good predictions in several cases – predicting that a below-proficient student will remain below proficient in math is accurate 77.7 percent of the time. However, AGP-based predictions incorporating the most recently observed growth for each student are better in all cases. The improvement in the percentage of correct predictions is quite impressive, providing evidence of the validity and usefulness of the AGPs. Of particular interest is the result suggesting that the AGPs are most useful at discerning whether students are beating the odds and "catching up" (on track to scoring

Proficient at a later date), because the improvements in correct predictions are highest for the "Below Proficient" rows; this is directly attributable to the power of the Colorado Growth Model and its extension to AGPs.

The predictions are not perfect. Such percentages of correct predictions are unlikely to approach 100 even under the best of circumstances due to the great variability of what happens in a student's life and schooling in the years subsequent to the growth calculations performed by the state. Indeed, the great improvements in prediction are quite remarkable by themselves, showing how useful growth data can be.

These growth-to-standard calculations are essentially a hybrid statistic, with both growth and proficiency components represented. Schools with large numbers of students scoring below proficient by definition have a more arduous task ahead because they will need their students to grow more in order to reach proficiency. No matter how high their normative growth is, the reality of what growth would be necessary for their students to achieve proficiency is calculated and reported. These AGPs are calculated at the individual level, but can be aggregated in the same way as student growth percentiles, by the creation of a central tendency in the form of the median. These median AGPs tell us what level of growth was needed for all students within a system/districts/school so that, on average, they would be reaching or maintaining proficiency within a reasonable timeframe. In the hands of a strategic building leader, these data are of immense value.

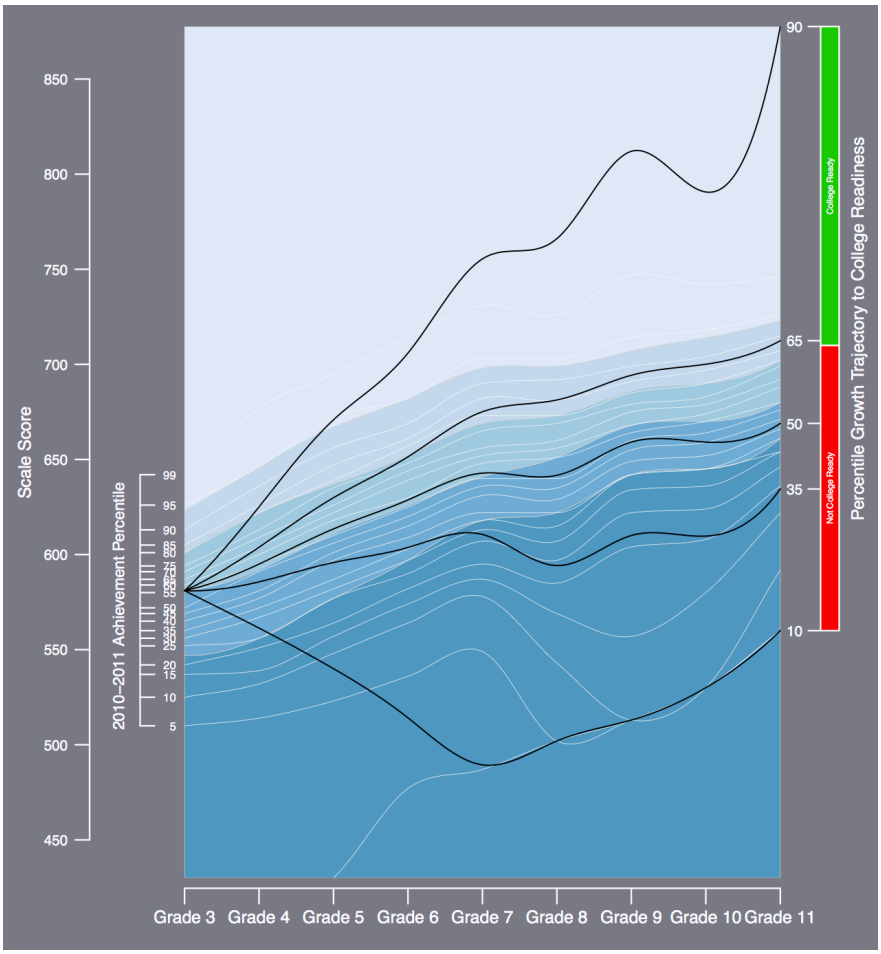
Achievement status still makes up a portion of a school or district's overall ratings on the performance frameworks that Colorado has put into place for its accountability system. Greater weights are given to the contributions from normative level of growth for all students, and from normative growth for historically disadvantaged subgroups of students, as well as from college and career readiness indicators for high schools. In the current performance frameworks, normative growth is always rated in the context of a reported AGP – in other words, growth is evaluated in light of whether it was sufficient to lead to the desired proficiency outcome for students.

The Colorado Growth Model method has distinctive characteristics that make the data particularly useful to stakeholders. Calculations are performed without regard to student demographics, so acceptable levels of performance are in no way defined by a norm for that race/ethnicity, income, or gender – only by what is enough for all students to reach a common destination of college and career readiness. Student growth scores expressed as percentiles have a straightforward interpretation: compared to one's academic peers, how much academic progress was made in the past year? Finally, in adopting a model that is descriptive rather than inferential, there is a reduced tendency to make a leap to causal conclusions and blame where performance is low. The data can therefore be understood in their own right and not seen as a threat, creating the needed space for a regulatory rather than punitive approach to accountability.

Ultimately the test of a growth model (and consequently of a growth-based accountability system) is whether or not it shows evidence of consequential validity – whether it causes the right conversations to happen, resulting in effective improvement efforts. Colorado has collected a great deal of data over the past few years suggesting that this is indeed the case. Evidence strongly suggests that Colorado's education community has embraced the state-produced growth calculations, believes in the data's underlying integrity and meaning, and shares a widespread sense of ownership and use of the data, particularly in conducting root cause analysis at scale as part of a systematic and public improvement planning process and in the accreditation of schools. Rather than arguing over what the evidence shows, Colorado's education stakeholders are instead

focusing their energies on understanding the data, attempting to identify the root causes of problems, and developing focused improvement strategies for ensuring that all students graduate college and career ready.

The overall effectiveness of a state system in promoting college and career readiness can be directly explored when norm- and criterion-referenced growth calculations are brought to bear on the issue. For example, the following figure shows what growth data tell us about students who score slightly above average on another state's 3rd grade mathematics assessment over time. Their starting place is just barely at the Proficient level, and the five dark lines illustrate where their scores would most likely end up in subsequent grades were they to grow at the 10th, 35th, 50th, 65th and 90th percentile level every year. Average growth (50th percentile) would probably keep them at the Proficient level until 8th grade, after which their scores would begin to fall below Proficient, remaining at that level through 11th grade. In other words, average growth is not sufficient to keep such students proficient until exit, even when they start out with a score falling within "proficient" in 3rd grade. Note that these students' test scores are even going up fairly substantially every year – but the reality of the situation is that these students are in fact losing ground over time, and thus are not likely to be college and career ready when they exit high school. Growth data such as these provide an invaluable source of information about the general functioning of the system, and about where urgent attention might be most needed.



cde Improving Academic Achievement