IDEA DATA CENTER
Building Capacity for High-Quality *IDEA* Data

Using Growth Models to Measure Child/ Student Outcomes for State Systemic Improvement Plans

A Guide for States

Jose Blackorby
Cornelia Taylor
Xin Wei

December 2016
Version 1.0

Westat®

The *IDEA* Data Center (IDC) developed this document under U.S. Department of Education, Office of Special Education Programs Grant No. H373Y130002. Richelle Davis and Meredith Miceli serve as the project officers.

The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, service, or enterprise mentioned in this publication is intended or should be inferred. This product is public domain. Authorization to reproduce it in whole or in part is granted.

For more information about the *IDEA* Data Center's work and its partners, see www.ideadata.org.

Suggested Citation:

Blackorby, J., Taylor, C., and Wei, X. (2016, December). *Using Growth Models to Measure Child/Student Outcomes for State Systemic Improvement Plans*, (Version 1.0). *IDEA* Data Center. Rockville, MD: Westat.

Introduction

In thinking about student performance for general and special education populations alike, it is popular to consider growth as a metric. After all, children develop as they grow older. They grow in height. They grow in weight. They grow in social and academic skills as well. It is natural for parents and caregivers to think about how their children are growing or progressing and how fast their growth is relative to their own past growth or in comparison to other similar children (Yen 2007). To look at student growth, states must first identify how they expect students to grow (e.g., 2 inches per year—linear growth, scores double every year—curvilinear growth). When states identify what they expect of students' growth, they must define a model to which they can compare student growth. Within educational accountability systems, status (current performance) models and growth models are both used to measure progress and outcomes (Raudenbush 2004) at the classroom, school, district, and state levels. This paper focuses on the use of growth models, broadly defined as the measurement of growth over time, not limited to measurement of growth over time in individual children.

Growth models have a number of advantages over status models. Growth models are sensitive to students' meaningful progress when performance remains within the same accountability performance level, predict whether students' progress is sufficient to meet a higher performance level at some point in the future, describe performance of groups of students in terms of proportions of students with increasing or decreasing rates of growth based on their prior performance, and estimate the effects of teachers or schools on improving student achievement (Castellano and Ho 2013; Council of Chief State School Officers [CCSSO] 2010). As such, growth models are being used for a variety of different accountability, instructional planning, and evaluation purposes in U.S. schools.

These advantages have prompted increased interest in using growth models for special education populations as alternatives to status models (Ahearn 2009; Buzick and Laitusis 2010a; Farley, Saven, Tindal, and Nese 2013; Thurlow, Lazarus, Quenemoen, and Moen 2010). This paper, with a focus on special education populations, is intended to provide state and local education personnel with an overview of issues to consider with growth models, models currently in use, and a description of common models and a scenario of their potential use within the State Systemic Improvement Plan (SSIP) process. This paper will describe the range of models and the advantages and disadvantages associated with their use. We do not expect to establish one recommended model but rather a range of options that may align with SSIPs.

This paper provides a series of growth models that states and schools are currently using for a variety of accountability and instructional decisionmaking practices. Most models use summative state accountability test scores over multiple school years, with the exception of progress monitoring, which focuses on within-year growth. The purpose of the paper is to provide readers with a brief overview of each growth model approach, its advantages and disadvantages, as well as a potential example of its use in the SSIP process. The growth models covered in this paper include the following:

- **Cross-Cohort Models** measure growth across cohorts of students over time. These models are commonly used in school accountability systems under *No Child Left Behind* (NCLB) (CCSSO, 2007).
- **Progress Monitoring Systems** focus on individual student growth over the course of the school year using multiple relatively short data collection events (Fuchs 2004; Gersten et al. 2009).
- **Simple Gain and Trajectory Models** usually depict growth across years by computing the difference between a student's status at two time points—prior and current performance.

- **Residual Gain Models** use statistical regression techniques to evaluate the degree to which a student's observed performance at a point in time differs from what was predicted based on student characteristics or prior performance.
- **Projection Models** use historical information about the longitudinal performance of cohorts of students to predict how a current cohort of students will perform in the future.
- **Value Tables** describe the performance of groups of students across 2 years in terms of their movement from one proficiency level to another.
- **Conditional Growth Percentile Models** develop a percentile rank for a student's current status, based on an expected status as determined by student characteristics and previous performance.
- **Value-Added Models** (VAM) are a family of models that measure the effects of teacher or school on student achievement.

Special Education Student Populations

Growth over time in individual children with disabilities is measured within Part B of IDEA through individualized education programs (IEPs) and, for 3- through 5-year-old children, the Part B 619 preschool outcomes indicator (B7). Through the IEP process, teams document current levels of performance and specify annual goals for students. Through the preschool outcomes indicator, preschool special education programs funded under Part B 619 of IDEA compare actual growth to expected growth for all children exiting the program within a federal fiscal year using the model described in Exhibit 1 (Early Childhood Technical Assistance Center 2014).

Exhibit 1. The five categories of progress used for accountability in Part B 619

Progress category	Explanation
a. Did not improve functioning	Children who acquired no new skills or regressed during their time in the program
b. Improved functioning, but not sufficient to move nearer to functioning comparable to same-aged peers	Children who acquired new skills but continued to grow at the same rate throughout their time in the program
c. Improved functioning to a level nearer to same-aged peers but did not reach it	Children who acquired new skills but accelerated their rate of growth during their time in the program. They were making progress toward catching up with their same-age peers but were still functioning below age expectations when they left the program.
d. Improved functioning to a level comparable to same-aged peers	Children who were functioning below age expectations when they entered the program but were functioning at age expectations when they left
e. Maintained functioning at a level comparable to same-aged peers	Children who were functioning at age expectations when they entered the program and were functioning at age expectations when they left

Students with disabilities in the K–12 system are included within the accountability framework adopted by their state to meet the requirements of the *Elementary and Secondary Education Act* (ESEA) (formerly NCLB), as they will be under the newly passed *Every Student Succeeds Act* (ESSA). These requirements specify that states, districts, and schools must establish Annual Measurable Objectives (AMOs) for third-grade to eighth-grade students in reading and mathematics and high school that define Adequate Yearly Progress or AYP. These targets must be met by the population as a whole as well as by student subgroups (e.g., students with disabilities). A status model is generally used to measure whether all students and

subgroups of students in a school meet AMO and AYP requirements. Status models label schools as *failing to meet* or *meeting* depending on whether the school or any of its subgroups fail to meet AMO in reading and math at one time point or on average over several time points. As a result of these determinations, schools may be designated as needing improvement, required to make staffing changes, forced to reconstitute, or closed. These status models used under ESEA accountability have been criticized in that they may provide misleading information about a school's improvement because they may reflect preexisting differences among schools but not necessarily the school's ability to educate students, can create unintended incentives for schools to focus resources on the students just below the proficiency threshold, may ignore the progress of students within a level and not yet ready to move to the next level, and so on (Pacific Research Institute 2004; Wyckoff 2005).

During the last decade, there has been increasing interest in using measurement of student growth for a range of educational applications (Braun 2005; McCaffrey, Lockwood, Koretz, Louis, and Hamilton 2004). This collection of approaches, frequently referred to as growth models, is currently being used in states for a variety of purposes and at a variety of system levels (Castellano and Ho 2013; CCSSO 2009). As shown in Exhibit 2, growth models are being used to measure individual student progress, make instructional decisions, and evaluate teachers. The most common application has been use within school and district accountability systems; therefore, much of the attention in this paper is paid to that application. Exhibit 2 further presents the purpose of each growth model.

Exhibit 2. Purposes for growth models across various system levels, by growth model

System level	Purpose of growth models	Cross-cohort	Progress monitoring	Trajectory models	Residual gain models	Projection models	Value tables	Conditional growth percentile	VAM
Child or student	Monitoring student progress; Informing families about their child's rate of growth		×	×	×	×		×	×
Classroom	Instructional decisionmaking		×	×	×	×			×
School	Teacher or provider evaluation	×		×	×	×	×	×	×
District/ Program	School accountability; Program/school improvement	×		×	×	×	×	×	
State	LEA accountability; Program/school improvement	×		×	×	×	×	×	
Federal	State accountability	×		×	×	×	×	×	

Staff and stakeholders across levels of the system are interested in using growth models because these models capture one of the key metrics of success, the degree to which program participation accelerates

acquisition of important knowledge and skills beyond what the student would have experienced without the program. Within the context of NCLB, the application of growth models was spurred by the U.S. Department of Education (ED) Growth Model Pilot Project (GMPP) in 2006, which approved Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee's implementation of growth models for accountability on a temporary basis (ED 2007). Under the GMPP, the federally approved NCLB growth models remained grounded in the mechanisms presented in the status models described above (Dunn and Taylor 2007). A large-scale evaluation of the GMPP showed results across states to be mixed, revealed some limitations and technical problems of the growth models, and did not result in markedly different accountability decisions for many schools than did status models (ED 2008). The continuing appeal of growth is evidenced by interest in and adoption of growth models of various kinds. By 2012, 35 states had already implemented or were designing some kind of growth model. In addition, there have been ongoing interest and publications from the CCSSO (Castellano and Ho 2013; CCSS, 2009, 2010), National Association of State Directors of Special Education (NASDSE) (Burdette, 2011), the federally funded testing consortia (Briggs 2011; FairTest 2010), the Race to the Top program (RTT) (ED 2009), and an active research community examining the design, technical characteristics, and applications of growth models to educational decisionmaking (Buzick and Laitusis 2010a; Ho 2008). It is also appropriate for states to consider whether growth models have a place in their plans for Office of Special Education Programs' (OSEP's) SSIP process.

Some Common Approaches to Modeling Growth

There are a number of different approaches to capturing, describing, and making decisions using growth data. The approaches described here all share the characteristic that they use the term *growth model* and have been used to varying degrees by states for a range of purposes. They differ, however, in their approach to defining growth, the type of data required, the availability of longitudinal data, the statistical analyses required, and, ultimately, the types of inferences that are possible. The descriptions that follow provide information about some commonly used growth models, their advantages and disadvantages, and their applicability to SSIP and state-identified measurable results (SIMR) goal setting and implementation.

Cross-Cohort Models

Although cross-cohort models are not technically growth models because they do not follow the progress of individual children and do not require statistical methods, cohort models have been commonly used in school accountability systems under NCLB (CCSSO 2007). The typical application of cohort models had schools document current performance, usually as a percentage of the population reaching the threshold of proficiency on an accountability test, and produce a plan for annual improvement such that percentage over time ultimately reaches 100 percent proficiency. This standard is applied to the total school population and subgroups such as students with disabilities. Schools that make progress toward annual goals are considered to be making AYP. Success or failure in any given year is dependent on the target set for that year. For example, 75 percent of sixth-grade students scoring as proficient in 2010 could be sufficient for the school to meet AYP goals in that year. In the same school 3 years later, 85 percent of sixth-grade students could be proficient—this is 10 percentage points more but the school might or might not make AYP based on the target for that year. Complex rules relating to subgroups, minimum *n* sizes, safe harbor, and similar factors all contribute to these determinations and vary by state. All states, and the vast majority

of schools in the country, have been operating on some variation of cohort model since the passage of NCLB in 2002.

Advantages. The advantages of this cross-cohort approach include its wide use in states and in schools, ambitious goals for all students, depiction of data-based goals for schools, identification of schools in need of improvement, and general understanding on the part of policymakers and the public. There are now data about the performance of states, districts, and schools readily available on websites, in online databases, and on school report cards. Cohort comparisons provide a way for looking at how school populations perform at different points in time. For example, sixth-graders in 2010 can be compared to a different group of sixth-graders in 2014.

Disadvantages. There are disadvantages as well. For example, an important issue is that the performance levels reported on successive years are not the same students. The focus on individual growth is assumed for most other growth models. In addition, in terms of measurement, such models assume that the characteristics of the student population are stable over time. There are instances in which this assumption may not hold up. For a variety of reasons, school enrollments may shift to include larger or smaller numbers of higher- or lower-performing students, which will influence the degree to which schools can achieve targets, resulting in potentially inaccurate inferences about school performance. Such models also are best when the test in use remains consistent or linking has been established. Shifts in tests have been common over the last decade. As 2014 approached, the year in which schools were to meet the 100 percent proficiency target, many schools and districts found it difficult to reach goals using the required approach. Nearly all states have applied for and have received waivers for ESEA flexibility in reaching these goals. Exhibit 3 provides an example of how modifications to the proficiency model might look.

Exhibit 3. Example of cohort model—Percent at or above proficient for third-grade students with learning disabilities from 2015 to 2020

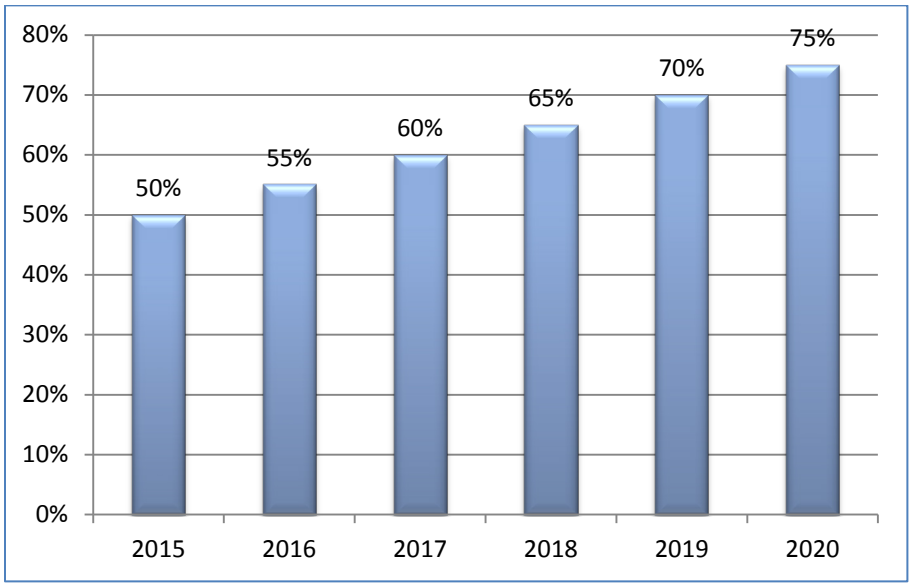


Illustration. Some states may select SIMRs that are similar to cohort models under NCLB. For example, a state may select performance on a state test on reading among third-grade students with learning disabilities as a SIMR. If the 2015 baseline showed 50 percent of these students to be at level 3 or 4, and the state planned to implement interventions to improve reading performance by 5 percentage points each

year until 2020, the percentage of 2020 third-grade students would be 75 percent, or 25 percent more than their peers in 2015.

Measuring Growth With Progress Monitoring

Another common way of measuring growth in academic skills is with a set of tools collectively referred to with the term *progress monitoring* (PM) (Safer and Fleischman 2005). PM systems differ substantially from the other growth models presented here in several ways. They typically focus on student growth as it occurs within the school year, rather than across multiple years as is common in the other examples in this paper. PM systems require multiple relatively short data collection events over the course of a year, with some districts recommending that measurements occur weekly (Fuchs 2004; Gersten et al. 2009). This approach allows data to be charted over time, and student progress to be expressed as growth by the slope of progress over time, particularly in light of an end-of-year goal. PM systems are generally not used in accountability systems.

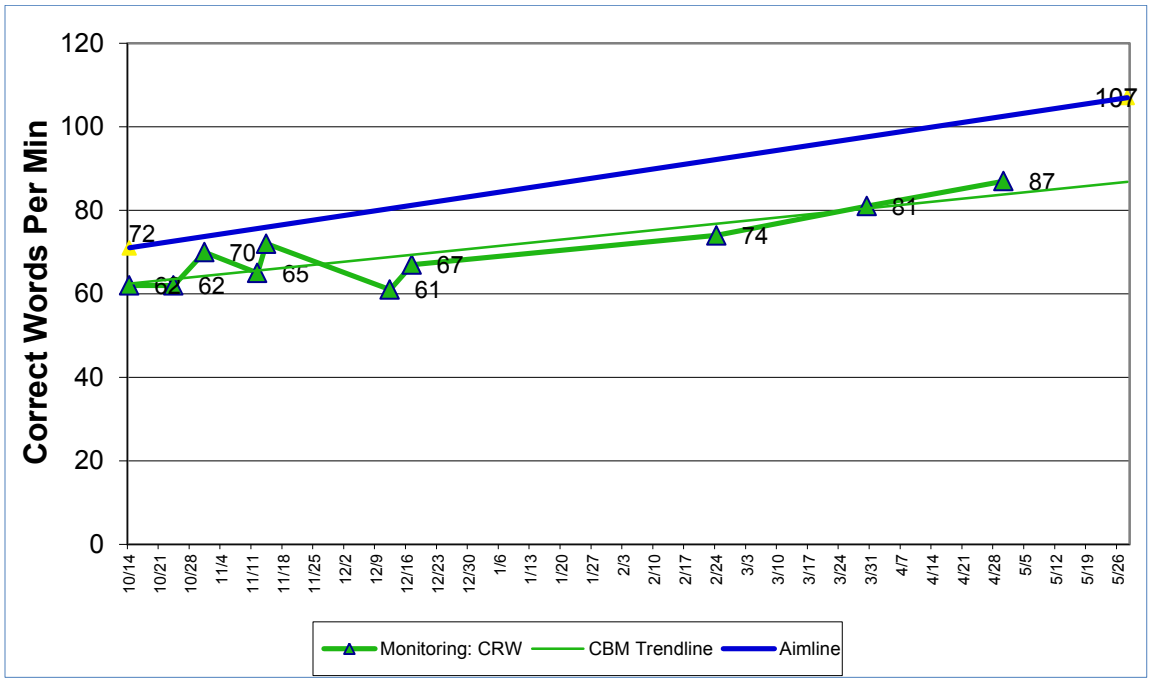
PM systems are available from a range of commercial publishers, such as DIBELS, AIMSWeb, and EasyCBM. Their use is quite common in U.S. schools today (National Center on response to Intervention 2012). DIBELS Next, for example, is used by more than 20,000 schools. They are frequently used in Response to Intervention (RtI) systems both as universal screening tools to identify students in need of more intensive service (usually at the beginning of the school year), as well as to evaluate the effects of additional services provided in Tier 2 or Tier 3 and to inform intervention decisions as well as the shifts in tier placement decisions (National Center on Response to Intervention 2012). Historically, the majority of technical work in PM measurements has focused on elementary-aged students when there is rapid growth in academic skills and reading because of the central importance of those skills. There has been increasing attention and availability of PM tools for older students as well as for other content areas such as mathematics (Fuchs et al. 2007). The measures are commonly based on fluency, such as words read correctly per minute, nonsense words read correctly per minute, correct maze items, or digits correct in mathematics (Shapiro, Zigmond, Wallace, and Marston 2011).

Advantages. The principle advantage of PM systems is that they are generally more useful for the purposes of informing instruction than other growth measurements because they are administered at multiple points throughout the year. Therefore, they can be used to identify students who need additional help and to evaluate progress relative to the end-of-year objectives (Jenkins, Schiller, Blackorby, Thayer, and Tilly 2013). In addition, PM measures are generally considered sensitive to effects derived from instructional interventions, allowing teachers to see whether interventions are working in general or working fast enough to make the desired changes (National Center on Response to Intervention 2010). Most current providers of PM systems support computer-based or online data collection and charting, which makes implementation much easier for teachers.

Disadvantages. Although easier to implement than in the past, PM systems do impose a burden on teachers and students to collect the data and to chart and interpret them. Even if weekly PM data are collected for only 15 percent of students in a class, the data collection and analysis time does add up. Much of the technical work around PM measures has focused on early elementary and reading (McCardle, Scarborough, and Catts 2001). The method is less fully developed in other content areas (Shapiro et al. 2011). To limit the data collection burden, many proponents recommend frequent measurement only for students who need additional service. This approach means that there is variability in the amount of data that are available for entire classes or schools.

Illustration. Exhibit 4 shows a fairly representative example of PM data. It shows an individual third-grade student's progress in oral reading fluency in words correct per minute measured at multiple intervals throughout a school year. The blue line represents the fall to spring growth and oral reading fluency for students at the 50th percentile at each time point, which increases from 72 to 107 words per minute over the course of the year (Hasbrouck and Tindal 2006). Using multiple administrations of the PM probes, the student's slope or improvement can be estimated through a line of best fit represented by the green line[1]. In this case, it shows that the student started the year somewhat behind grade-level peers and improved to 87 words per minute by the final data collection point; however, the student did not close the gap with general education peers. Commercial systems allow for charts such as this one to be generated at the individual, class, and even school level. Within an RtI system, screening early in the year would determine whether a student would be eligible for additional services. If the benchmark were 65 words per minute, then the student might be eligible and might receive small group or more intense instruction in Tier 2. Based on data points later in the fall, the student may have reached criterion to no longer require Tier 2 services and continue to receive Tier 1 only. The subsequent drop in performance in December could trigger additional service again. In the context of a SIMR, a state might be seeking to implement RtI systems at low-performing schools, and the goal might be to increase the number of students receiving intensive services who closed the gap with age norms over time such that 90 percent of students achieve grade-level goals and fluency and require only Tier 1 services. Alternatively, a state might set goals for reducing the number of students who are nonresponsive to intervention or who need Tier 2 services by third grade.

Exhibit 4. Progress monitoring data in third-grade reading



[1] Correctly read words (CRW); Curriculum based measurement (CBM)

Simple Gain and Trajectory Models

Simple gain and trajectory models take a similar approach to depicting growth (usually across years). A simple gain model computes the difference between two time points, prior and current performance, and the result represents the amount of change that occurred over the specified period of time. If a student received a score of 350 on a state accountability test in 2015, and scored 375 in 2016, the 25-percentage point difference is the simple gain. The progress categories defined for Part B 619 child outcomes reporting use the gain, both in absolute skills and relative to same-age peers, to describe and categorize growth between entry and exit. Trajectory models predict this difference into future years, assuming that the rate of growth will continue at the same rate. Consequently, the student would have a score of 400 in 2017, 425 in 2018, and so forth. This approach can be applied to individual children or groups, such as students with disabilities. The amount of growth for an individual or group can be qualitatively evaluated as positive, negative, or neutral in terms of the slope over time. Individual or group observed or predicted trajectories can be compared to proficiency levels, to local or state norms, or to reference groups (e.g., the general population). The on-track trajectory is determined by the annual gain needed to reach the goal. The difference between predicted trajectory and on-track trajectory tells us whether a student or a group is on track to meet the goal. In NCLB terms, this trajectory has been referred to as the *on track to be proficient* category, and some states allowed students in this category to count toward AYP.

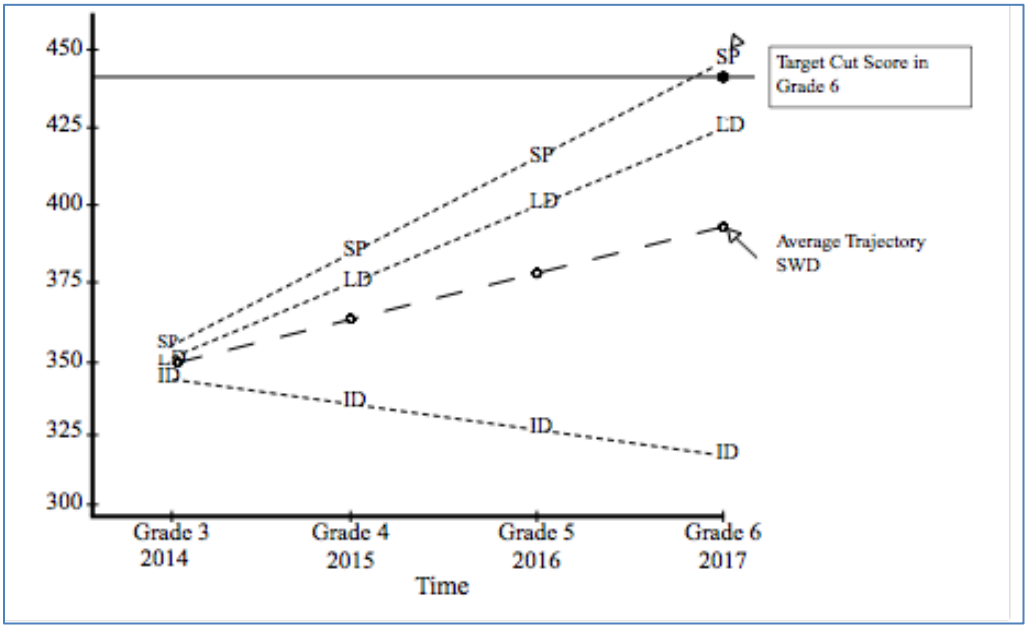
Advantages. Simple gain and trajectory model strengths include intuitive face validity with conceptions of growth and comparative ease of communication to and comprehension by a range of audiences, including policymakers, educators, and parents. These models allow students who are making progress toward proficiency to be positively counted in accountability systems and provide an important control lacking in cohort measurement by comparing students to themselves.

Disadvantages. One disadvantages of simple gain and trajectory models is the need for vertically scaled measures so that inferences drawn about the amount of measured gain have comparable meaning across years. Another disadvantage is the assumption that the rate of measured growth will be a valid representation of the trajectory of student achievement. Differences in tested grade-level content across multiple years make these assumptions questionable: Growth over time may not be linear, and absolute changes may be different for both high- and low-performing students. Unlike the cohort approach, these measures can only be taken for students who participate in both pretest and posttest. The results of these approaches can be sensitive to both the selection of the relevant time threshold and to cut points for proficiency, which involve value judgments by policymakers. An individual student or group's gain could be positive compared to other students with disabilities but negative compared to age peers. Finally, simple gain and trajectory models do not take into account relevant information, such as student demographic characteristics.

Illustration. It is possible to think of trajectory models in the context of a SIMR. Following the example presented in Exhibit 4, a state has selected third-grade reading achievement for its SIMR and might focus on intervening early in students' educational careers in the hope of improving outcomes as students age. Baseline data could indicate that students with disabilities perform, on average, below proficiency, and the goal would be to have them reach proficiency by sixth grade. Exhibit 5 presents a simplified district with only three students, each in one of three disability categories: speech–language impairment (SP), specific learning disability (SLD), and intellectual disability (ID). Their average is presented. All three students had similar performance in 2014 when they were in third grade. They had different scores in fourth grade. The SP and SLD students improved by 25 and 19 points, respectively. The student with ID had lower scores in fourth grade by 11 points. Grades 5 and 6 simply extend the prior performance by the same amount. The three students show very different trajectories over time. The SP student could be in the proficient

category by sixth grade, but the other two would not, nor would the average of the three. A state could set a SIMR to have a certain percentage of students with disabilities on track to reach proficiency by sixth grade. The straight lines on the chart highlight a limitation. It is not likely that the performance—positive or negative—would continue at the same rate.

Exhibit 5. Individual and group trajectory



Source: Adapted from Castellano, D.E., and Ho, A. (2013). *A Practitioner's Guide to Growth Models.* Washington, DC: Council of Chief State School Officers.

Residual Gain Models

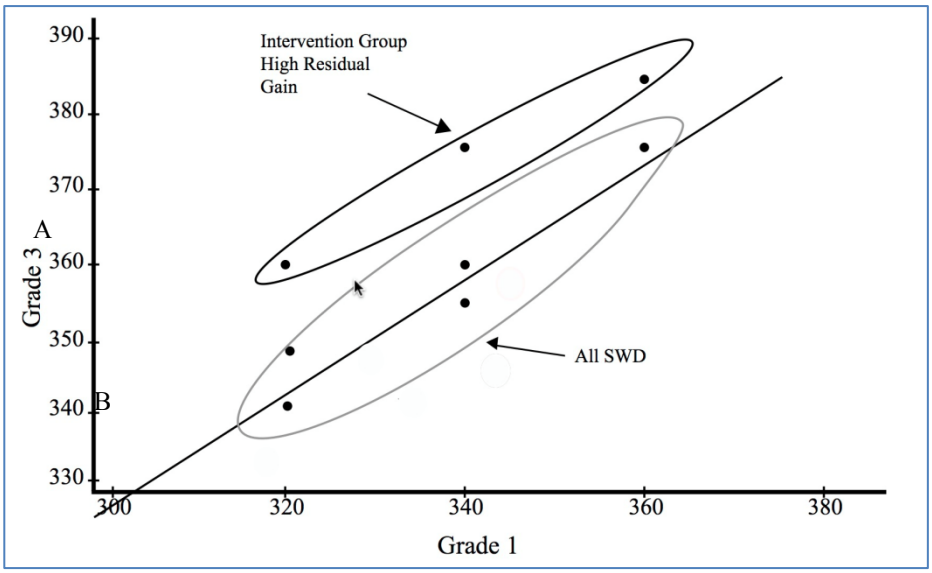
Trajectory models focus on the absolute amount of gain from one testing occasion to the next. A different way of thinking about growth is to think of the observed performance as not a gain at all, but as *relative* to what was expected. Residual gain models use statistical regression techniques to evaluate the degree to which an individual's observed performance in a given year differs from what was predicted based on the student's prior performance and/or other characteristics. The difference between students' observed score and their predicted score is referred to as the *residual*. The direction and the magnitude of the residual value allows a person to answer questions of whether a student is performing as well as predicted, above expectations, or below them. This residual is often interpreted as the student's growth relative to other students in the same group (i.e., the group used to develop the statistical model). This interpretation assumes that the predicted performance, which is based on the average performance of all students in the group, reflects the result of the normal growth of similar students. Thus, a student whose actual performance is equal to predicted performance is often characterized as having made normal growth. It is also possible to compute the average residuals for programs or groups for comparison purposes. The magnitude of the residual gain does not have specific meaning. It can be interpreted in relation to the expectations based on a proficiency threshold (e.g., distance to proficiency), to other groups (e.g., general population), or the distribution of residuals themselves (e.g., identifying the top 25 percent as exceeding expectations). Covariate adjustment models, the extension of residual gain models, are essentially the simplest VAM, which are discussed later.

Advantages. Residual gain models use regression techniques to predict performance, which allows the predictions to be more precise than trajectory models, assuming that the underlying assumptions of regression are not violated. Rather than thinking about growth in absolute terms, residual gain models support thinking about growth in *relative* terms, relative to the previous performance, which means that they are based on judgments to some degree. For example, a relatively large absolute growth might be below what is expected, or a small one might be more than was expected. Unlike simple gain or trajectory models, residual gain models use regression techniques to control for the prior test scores and sometimes student or program characteristics. The comparison of growth is more valid because it allows different growth trajectories depending on prior performance. For students with disabilities, growth comparison allows for thinking about their performance based on what was predicted, rather than in relation to a threshold. For example, a student might be well below proficiency thresholds but have score gains that are higher than predicted with the same prior test score.

Disadvantages. There are several disadvantages associated with residual gain models. One potential issue is confusion of relative gain with actual gain. Another possible issue stems from the fact that residual gain models assume that variation in residuals is comparable across the scale. Residual gain models assume the range and variability in residuals for lower-performing students, like many students with disabilities, and higher-performing students are comparable. This assumption may not always be tenable. Such models assume that growth will be linear, which may not hold uniformly across age groups or disability categories. In addition, it can be confusing to users that the average of residuals is always zero and lands on the regression line, meaning it is statistically not possible for all students in a given dataset to exceed expectations. Thus, this model is not useful for assessing how much the entire group has gained.

Illustration. Exhibit 6 uses the following SIMR example. A state is interested in improving early reading instruction such that, over a 3-year period, more students will be closer to grade level by third grade. To accomplish this goal, the state plans to implement a pilot program in 10 districts around the state. Exhibit 6 shows the linear regression analysis of first-grade scores and third-grade scores for the state as a whole. The slope line in the chart represents the best fit among all of the observed third- and fourth-grade scores and represents what performance would be on average. The dots on the chart represent student performance in first and third grades. All students in the state who had 320 in first grade are predicted to have a score of 340 in third grade, and all students who had 340 in first grade are predicted to have a score of 355 in third grade. Student A scored 320 in first grade and 360 in third grade. In the example, Student A's residual gain is 360 minus 340 or 20. The student performed above expectations. Student B also was predicted to have a score of 340, but his or her score was 337, so the residual would be minus 3 points. So, student A could be considered to have performed above prediction, and Student B performed below prediction.

Exhibit 6. Residual Gain Models



Source: Adapted from Castellano, D.E., and Ho, A. (2013). *A Practitioner's Guide to Growth Models.* Washington, DC: Council of Chief State School Officers.

Student A could be considered particularly successful if the gain of 20 was especially high and/or if it put that student closer to an accountability threshold. In the SIMR context, if all the scores were above expectations, it would indicate growth that is greater than expected in the intervention group. One might state that this intervention was worthy of wider implementation because 75 percent of intervention recipients scored better than expected based on population expectations.

Projection Models

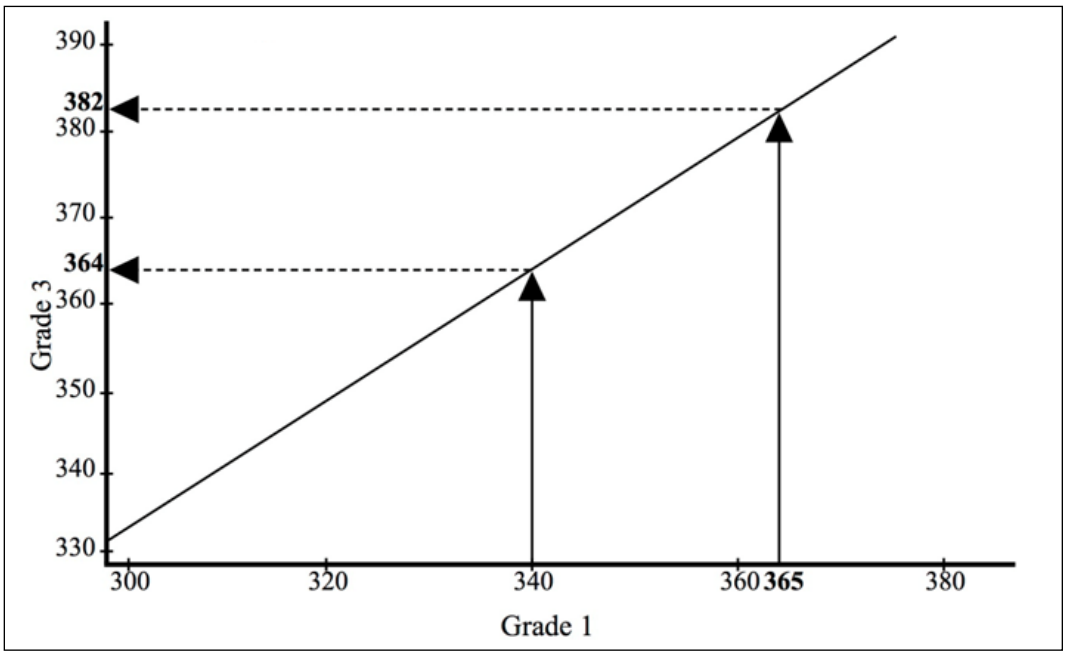
Projection models predict or project the performance of a current cohort of students using the regression line that was estimated from longitudinal data of a past cohort of students' performance (Castellano and Ho 2013). These models use linear regression methods to estimate future performance of a past cohort of students using their performance data from at least 1 year. Additional years of performance data as well as additional information about the past cohort of students (e.g., race or ethnicity, disability category) can improve the reliability of slope of the regression line. Projection models assume that the successive cohorts of students have comparable schooling experiences over time. Unlike trajectory models, which extrapolate an individual student's own past performance into the future, projection models estimate an individual student's performance to a past cohort of students' predicted performance. For example, a regression line will be estimated using the performance of a past cohort's third- and fourth-grade scores. To calculate the projected fourth-grade score, the observed performance score of a third grader in the current cohort is plugged into the estimated regression equation. The projected fourth-grade score will be compared to the accountability standards to determine whether students would be considered proficient on accountability at some point in the future. Students could be considered *on track to proficiency* if their projected scores fall at or above the proficiency level for future grades on the state assessment (most commonly 3 years under NCLB). In some states, a student who is on track to be proficient can be counted toward AYP calculation. Four states—Arkansas, Iowa, Ohio, and Pennsylvania—use variations of the projection model. As with other models, projections can be made for groups of students or schools.

Advantages. Both trajectory models and projection models offer predicted scores for a target grade level using linear regression models (Castellano and Ho 2013). The use of regression as a statistical procedure allows for the inclusion of more information (e.g., multiple prior years of performance data, student characteristics) in predicting future growth, and regression is relatively straightforward to interpret. Projection models can also be used to identify students in greatest need of assistance for meeting proficiency and to reallocate resources toward them (Goldschmidt et al. 2005). It is possible to estimate such models separately for students with disabilities, various disability categories, and other subgroups.

Disadvantages. To enhance accuracy, projection models commonly require multiple years of data; they require extra effort in data collection and can leave out more students due to missing data. Projection models can be sensitive to both data quality and completeness because a large number of variables (multiple years of test score as well as background characteristics variables) can be included in the regression model. Projection models also require that considerable amounts of longitudinal data are available, which increases the likelihood of missing data in the dataset. If the amount of missing data is substantial or highly variable across subgroups, imputation may be required, and the projections may no longer be as accurate. Projection models also assume that the characteristics of cohorts remain comparable over time. If there are numerous population shifts over time, the accuracy of the projections may be reduced. If the content that is tested over time is variable, the projections also may be less accurate. For example, fourth-grade performance is likely to be a better predictor for fifth-grade content mastery than it is for eighth-grade content mastery. If there are large numbers of students with disabilities with low performance, their projected performance could be low also. Projection models assume that growth will be linear, which may not hold uniformly across age groups or disability categories.

Illustration. Exhibit 7 returns to the prior example of a state choosing a SIMR to improve performance in third grade. The state plans to implement an intervention at pilot schools and wants to know what first-grade performance levels are required for students to be proficient by third grade. The exhibit shows that any student's third-grade score can be predicted from his or her first-grade score. A student with a first-grade score of 340 would be predicted to have a third-grade score of 364. A student with a first-grade score of 365 would be predicted to have a third-grade score of 382. If the third-grade threshold for determining proficiency was set to be 375, then one student would be predicted to be proficient and the other would not. A SIMR using this approach might state that 75 percent of first-graders would reach 355, the level at which they would be predicted to be at the third-grade cut score in third grade.

Exhibit 7. Projection Models



Source: Adapted from Castellano, D.E., and Ho, A. (2013). *A Practitioner's Guide to Growth models.* Washington, DC: Council of Chief State School Officers.

Value Tables

Value tables provide another approach to document students' movement from one proficiency level to another proficiency level over successive years. Value tables use a tabular format to describe the performance of groups of students across 2 years. Columns and rows of the table are typically organized by performance levels (e.g., *not proficient*, *proficient*). The table can show the number of students who were *not proficient* in one year and remained *not proficient* in the next one, as well as those who were not proficient but moved up a category the subsequent year (Castellano and Ho 2013). These models are typically used for school performance assessment and accountability, where the expectation is that students will move toward higher proficiency levels. Note that this model is not intended for tracking individual students' growth, but for tracking the effectiveness of programs or education organizations. In using value tables, states can apply different weights to represent values ascribed to different types of transitions. Policymakers assign scores to different combinations of students' achievement levels in 2 years. Value tables usually assign positive scores to students who move from lower performance levels to higher ones (Hill et al. 2006). For example, a state could assign positive scores for students moving up, negative ones for those moving down, and zero for those who stayed the same. Several states use value table approaches (Arkansas, Delaware, Iowa, Illinois, Minnesota, Virginia). Delaware uses subcategories to evaluate progress within a proficiency level. A similar approach, a transition matrix, has been demonstrated appropriate for students with significant cognitive disabilities in the alternate assessment program in Oregon (Farley et al. 2013).

An equal value table is an application of value tables that assigns the same score for students who stay at the same proficiency level, equal but higher scores for every movement from a level to one level higher, equal but second higher scores for every movement from a level to two levels higher, equal but third higher

scores for every movement from a level to three levels higher, and so on (Hill et al. 2006, p. 264). For example, students who remain in the same proficiency level across years receive a value of 100; students who move up one level receive 150. The weighting formula applied across the cells of the table results in a summary. Exhibit 8 presents a hypothetical distribution of student proficiency level change from one year to the next, with the number of students in the diagonal[2] indicating the no change in proficiency level in 2 years. A score for a school is a weighted mean value score for students in the school. See the footnote below Exhibit 8 for an example of how to calculate the weighted score.

Advantages. Value tables present change in student performance levels over time in an easy-to-follow table. They do not have as stringent requirements for vertical scales of tests, that the same test is used over time, nor do they require large sample sizes or sophisticated statistical modeling. They can be used for students who take alternate assessments, and these students' performances can be combined with peers who take the general assessment. Students taking different tests on different scales can be included in the same table.

Disadvantages. There are several disadvantages with value table approaches. First, they do not address student change within a proficiency level (unless subcategories are defined), nor account for the amount of change. For example, a student who is very close to the cut point and another many points way from it are treated the same if they both improve to a higher level. In addition, value tables only allow for transitions over a 1-year period and do not predict estimates for future years. Also, in practice, some students do go to lower proficiency levels either through regression to the mean or poorer performance. Finally, the value table approach relies on human judgment on two counts: the establishment of the cut scores in the first place and weighting of different student transitions (Buzick and Laitusis 2010b).

Illustration. Exhibit 8 considers a state that is choosing a SIMR to improve performance in adjacent years and that wants to focus its intervention on those students who were below proficient (levels I and II in Exhibit 8) in the first year. The state plans to implement an evidence-based intervention for students in those performance levels statewide and is interested in moving students into higher proficiency levels. Based on research on the intervention, the state believes it is reasonable to expect 35 percent of students to make sufficient growth to make it into a higher level and sets its SIMR accordingly. In the case of Exhibit 8, 38 percent of year 1, performance level 1 students moved up one or two levels, so they would meet the target, yet 33 percent of year 1, performance level 2 students moved up, and therefore would not meet the target. Note that this example ignores both high-performing students and students who lose ground.

[2] The term *diagonal* refers to the diagonal scores in the table in Exhibit 8 representing students whose scores remained at the same performance level—40,30,60 and 100,100,100.

Exhibit 8. Distribution of students in a school

Counts of students	Year 2 performance level		
Year 1 performance level	I	II	III
I	40	20	5
II	10	30	20
III	5	15	60
Value score	**Year 2 performance level**		
Year 1 performance level	**I**	**II**	**III**
I	100	150	200
II	50	100	150
III	0	50	100

Note. Total number of students is 205. Score= weighted value scores = [100*40+150*20+200*5+50*20+100*30+150*20+0*5* +50*15+100*60]/205=103.66

Conditional Growth Percentile Model

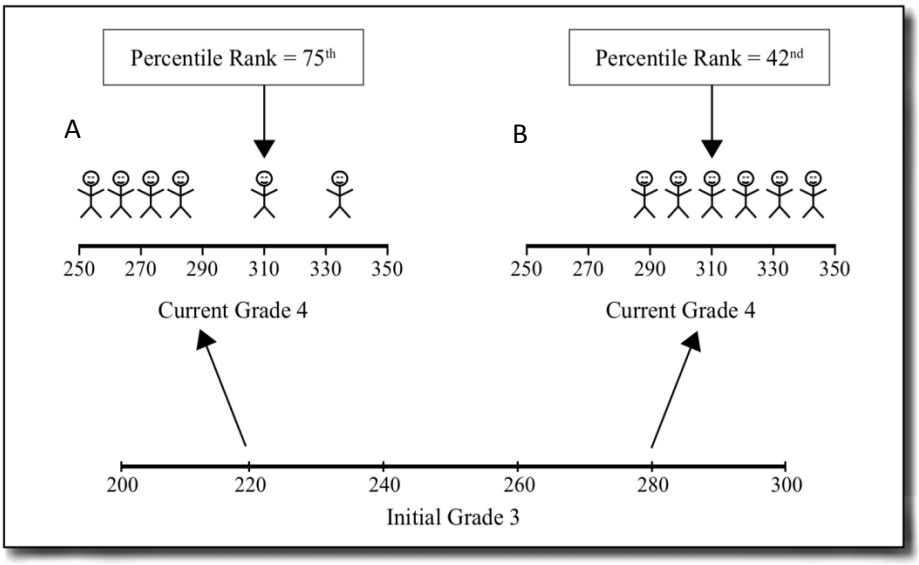
The conditional growth percentile model (CGP) draws on the techniques used in pediatric conditional reference growth charts that doctors use to report to parents what their child's height or weight is in percentile terms in comparison to other children of the same age. For example, they may report that a child's height puts him or her at the 85th percentile of children the same age. CGP takes a similar approach and uses quantile regression models to estimate performance quantiles given previous test scores in percentile ranks. CGP describes a student's status in percentile ranks with values ranging from 0 to 99 (Betebenner 2011; Betebenner and Shang 2007). Different from linear-regression-based growth models for which future achievement is predicted from one regression line, CGP fits 99 lines and one for each conditional percentile. By conditioning on prior measurements on percentile rank, conditional reference growth charts produce longitudinal growth percentiles to screen aberrant movement of individual students or groups of students. Unlike other growth models, CGPs compare student growth to others with comparable scores, rather than the entire distribution. In Exhibit 9, provided by Castellano and Ho (2013), two students (A and B) have the same fourth-grade score of 310 but have different percentile ranks because of the differences in their grade 3 scores. Student A's expected percentile rank is 75th percentile when compared to his peers who perform at the same level at grade 3. Although achieving the same score in grade 4, student B's expected growth percentile rank is only 42nd percentile because he did not improve at the same rate as his peers who scored at the same level at grade 3. Similar to trajectory models and projection models, CGP supports growth prediction; therefore, the predicted student CGP can be compared to the standard to determine whether a child or a group is on track (Castellano and Ho 2013).

Twelve states currently use CGP to describe growth at the state, district, school, and subpopulation levels (Betebenner 2010). Colorado was the first state that adopted CGP. The Colorado Department of Education (CDE) categorizes individual growth into three groups: low growth, typical growth, and high growth, which means a student's CGP falls within 0 to 34th percentile range, 35th-65th percentile, 66th to 99th percentile, respectively.

Based on individual CGP, the Colorado CGP models aggregate growth at the student level to the group level, which can inform state policymakers, educators, and parents about the growth of different groups of students at the state, district, and school levels. At the level of student groups, median growth percentile determines the low growth, typical growth, and high growth status of a student group. Typical growth for a

group of students requires that group to have an aggregated CGP median of 50. For schools to achieve high growth status, their students need to achieve high growth. The CDE model captures both status and growth as schools are placed into four groups: high growth and high achievement, high growth and low achievement, low growth and low achievement, and low growth and high achievement (Exhibit 9, http://www.schoolview.org/GMFAQ.asp#Q29).

Exhibit 9. Conditional Growth Percentiles



Source: Adapted from Castellano, D.E., and Ho, A. (2013). *A Practitioner's Guide to Growth Models.* Washington, DC: Council of Chief State School Officers.

To understand whether there is a gap in the growth rates among subgroups of students and how the gap changes over time, CDE identifies three additional categories of students by percentage: students catching up,[3] keeping up,[4] or moving up.[5] These categories are calculated for all students in a school, as well as major subgroups of students. For example, the school-year growth summary report presents median growth percentile, percent catching up, percent keeping up, and percent moving up for current school year and two previous school years for each tested subject for all students, grade-level subgroups, majority vs. minority, free or reduced-price lunch (FRL) vs. non-FRL, English language learner vs. non-English language learner, and girls vs. boys subgroups.

Students with disabilities make up one of the subgroups reported on in theColorado Department of Education SchoolView® website. Exhibit 10 shows the growth summary for the students with disabilities subgroup in a school in Colorado. This figure reflects that the students with disabilities subgroup showed low achievement and low growth. For schools with fewer than 20 students with disabilities, the median growth is not reported.

Advantages. Similar to the height and weight growth charts, CGP is easier to understand and explain than regression-based growth models. CGP uses quantile regression, which is complicated but also more flexible

[3] Previously scoring at the Unsatisfactory or Partially Proficient achievement level and demonstrated enough growth in the past year to reach Proficient or Advanced within 3 years or by 10th grade.

[4] Previously scoring at the Proficient or Advanced achievement level and demonstrated enough growth in the past year to maintain proficiency over 3 years or until 10th grade.

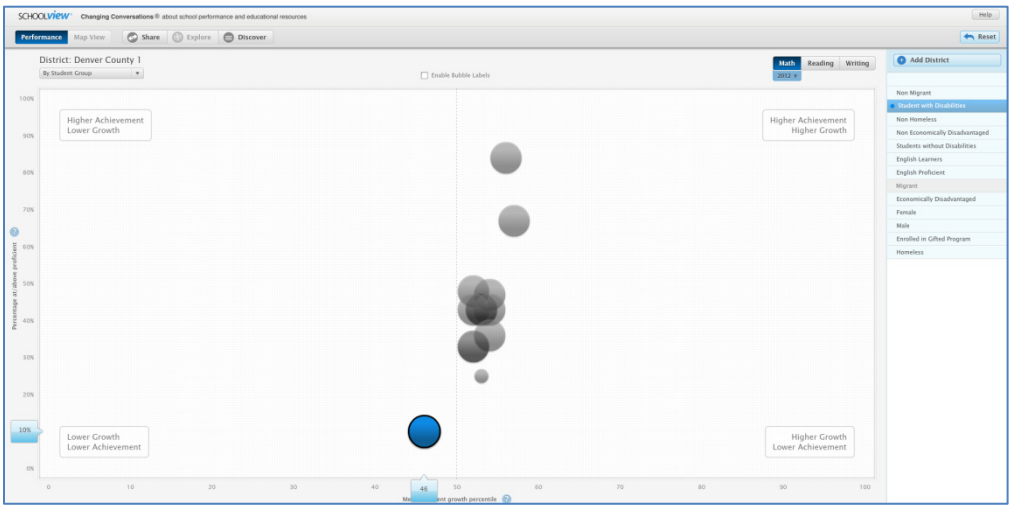
[5] Previously scoring at the Proficient achievement level and demonstrated enough growth in the past year to reach the level of Advanced within 3 years or by 10th grade.

than other regression-based approaches. For example, it does not require a vertically aligned scale, does not require linear growth, and is robust to outliers (Castellano and Ho 2013). The CGP approach compares student performance to other students at a similar percentile rank in the distribution. It can display both status and growth in a straightforward way.

Disadvantages. Grady, Lewis, and Gao (2010) and Castellano and Ho (2013) indicate that CGP requires a very large sample size (at least 5,000 students) to achieve accurate estimates. CGP lacks good statistical estimates for precision. CGP may not perform as well as ordinary least square regression of current status on past scores as percentile ranks when data follow a multivariate normal distribution. The CGP does not control for student background characteristics that are included in regression-based models and Value-Added Models (VAMs) used by states and school districts.

Illustration. It is possible to think of the CGP framework in a SIMR context. For example, Exhibit 10 shows that students with disabilities are in the lower-left quadrant, indicating low performance and slow growth compared to others at the same prior-test percentile. The state could use this information to identify schools in need of intervention. As a SIMR target, states might use the metrics of catching up,[6] keeping up,[7] and moving up[8] and set goals for the percentages of students expected to catch up or move up.

Exhibit 10. An example of growth summary of students with disabilities in a school in Colorado



Source: Colorado Department of Education.

Value-Added Models

Value-added models are among the most widely implemented, methodologically and politically debated, and technically studied applications of growth models. The term value-added is originally from economics, and it refers to the additional value created by factors of production, such as labor or land. The application

[6] Previously scoring at the Unsatisfactory or Partially Proficient achievement level and demonstrated enough growth in the past year to reach Proficient or Advanced within 3 years or by 10th grade.

[7] Previously scoring at the Proficient or Advanced achievement level and demonstrated enough growth in the past year to maintain proficiency over 3 years or until 10th grade.

[8] Previously scoring at the Proficient achievement level and demonstrated enough growth in the past year to reach the level of Advanced within 3 years or by 10th grade.

of VAMs to school accountability or teacher evaluation attempts to isolate the school or teacher effects on student achievement from non-school-related factors, such as family, peers, and student prior achievement (McCaffrey, Lockwood, Koretz, Louis, and Hamilton 2004).

The residual gain model is the simplest value-added estimate for schooling effect on student academic progress. It is calculated by taking the difference between students' actual growth and their expected growth (Goldschmidt et al. 2005). Value-added teacher or school effects are estimated from more complicated models than the residual gain models because VAMs control for prior student achievement on multiple measures and other student background characteristics.

VAMs usually measure teacher or school effects, yet other growth models also may measure individual student growth. VAMs differ in how they control for student-, classroom-, and school-background characteristics and whether they compare teachers within schools or across schools (district or state). Consequently, they may produce different teacher value-added scores, resulting in quite different teacher quality rankings (Goldhaber and Theobald 2012). The correlation of teacher value-added scores across different VAMs can be as high as 0.90 (Goldhaber and Theobald 2012). Although the correlation can be high across different VAMs, teacher effectiveness ranking can still be different. For example, even when the correlation is 0.97 between two VAM models, 6 percent of teachers at the top quintile in one model were placed at the bottom quintile by the other model.

As a distinction from other growth models, VAMs are interested in attributing some portion of observed performance to a school or a teacher, not directly to student proficiency on accountability tests. Indeed, depending on the context of a school or teacher, comparatively low performance could still be considered *significant* and rewarded in an accountability system. Variations of VAMs are used in a number of states, large districts, and the Gates Foundation Measurement of Effective Teaching Study (METS).

Estimating value-added scores for special education teachers is an under-studied area. Teachers of disadvantaged students benefit from models that control for student background factors, suggesting that VAMs controlling for student disability status and severity may give a higher ranking for special education teachers than models that do not control for student background factors (Goldhaber and Theobald 2012). McCaffrey and Buzick (2014) stated some possible reasons for misclassification when using a VAM to estimate special education teacher effects on students with disabilities include low scores, inconsistent use of test accommodations, and low participation rates in the standardized tests among students with disabilities.

Advantages. VAMs are designed for an education accountability system that evaluates the effect of a school or a teacher on student achievement. To estimate school or teacher effect on student achievement, some VAMs take into account student achievement from the previous years as well as demographic characteristics. By contrast, the gain score model or the cohort model only take into account one previous year's test scores. VAMs also have the advantage of handling various patterns of missing data [e.g., SAS Institute's Educational Value-Added Assessment System (EVAAS) models].

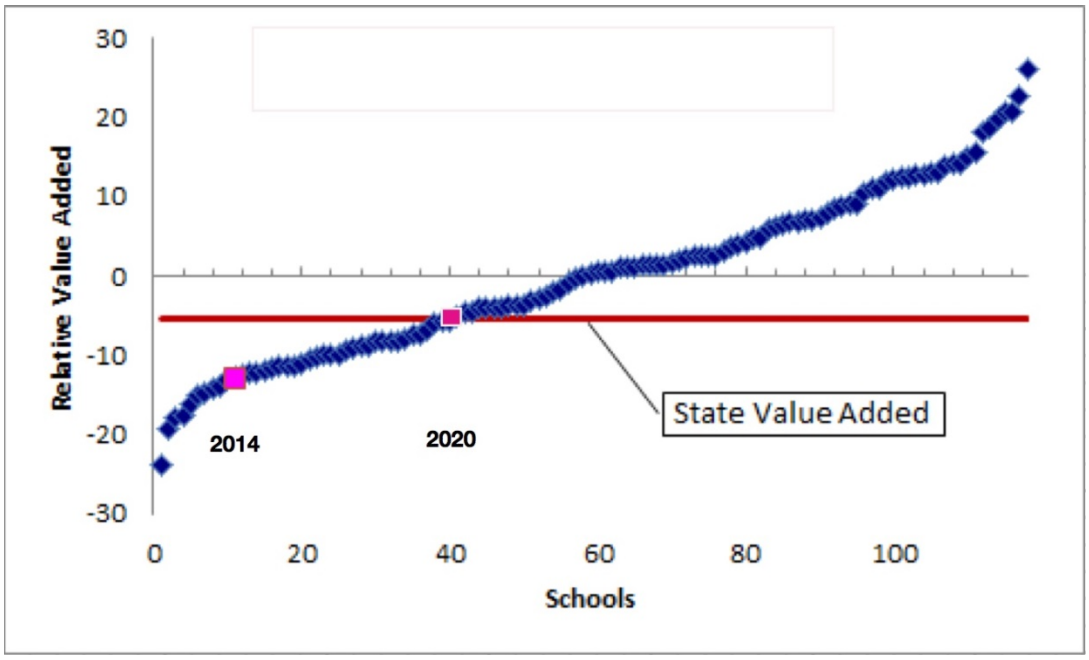
Disadvantages. There are active methodological and policy debates around the technical adequacy of VAM methods, as well as their use in evaluating schools and teachers. For example, there are questions about the reliability of VAMs for measuring of teacher quality. It is important to examine the confidence intervals[9] around the teacher and school estimates, as well as VAM estimates themselves, to compare teachers or schools. There are both false positives and false negatives in VAM systems, and some experts believe that this uncertainty makes VAMs inappropriate for high-stakes decisions (Raudenbush and Jean 2012). In

[9] This could be a strength of the VAM over CGP because VAM provides how reliable and precise the teacher effect point estimate is.

addition, some types of teachers, such as science and social studies teachers, may not have data on which VAM analyses can be conducted, so they are either outside of the accountability system or measured differently. Also, the relatively smaller number of special education teachers and students make the estimation of models and policy applications problematic. Finally, some argue that VAM evaluations create misplaced incentives. Implementing VAMs is not straightforward, and some statistical assumptions are debated. The results derived from VAMs are not always easy to understand for nontechnical audiences (Koretz 2008).

Illustration. In practice, through VAM models, schools or teachers can receive VAM scores that can be compared to one another for decisionmaking purposes. Value-added measures differ from state to state. Teacher value-added scores are usually calculated by averaging the differences between actual and predicted scores of all students in a teacher's classes. One of the popular VAM models, EVAAS, claims that it takes a complicated formula to account for student, classroom, and school characteristics to be fair; however, the exact model and data-processing steps used to calculate the scores are not available. Exhibit 11 shows hypothetical school-level VAM scores for students with disabilities in third-grade reading, accounting for poverty, size, and school-level resources. Schools are ranked from low to high and show substantial variability in comparison to one another as well as to the state average. In an SSIP context, a state might want to target schools at the low end of scale to implement new intensive intervention programs. A possible SIMR, then, could be to move the schools to the state average over a 5-year period of time.

Exhibit 11. An example of school-level VAM scores in reading achievement



Discussion

Advantages of Growth Models

This paper has provided an overview for state offices considering of common applications of growth models for education evaluation purposes as part of their SSIPs and SIMR. The concept of growth for thinking about educational progress has a number of advantages. It is intuitive because it connects well to commonly understood principles of child development and is readily understood by a wide range of constituents; therefore, the concepts of typical growth, the rate of growth, and closing the gap between subgroups of students are all ideas constituents can generally comprehend, communicate, and use in decisionmaking (Auty et al. 2008). Particularly in comparison to simple status models, growth models can be a promising alternative (Ahearn 2009). For example, it is plausible that a fifth-grader with a learning disability might have performance on an accountability test that falls in a category of not *proficient* and that over the course of a school year, the student could make meaningful progress but remain in the same performance category at the time of the next assessment. Growth models can be used to establish whether students' progress puts them on target to become proficient at some point in the future (either within year or across years), or is greater than would have been expected compared to other similar students or to their own prior performance (Castellano and Ho 2013; CCSSO 2010). Similarly, growth can be aggregated to describe performance of groups of students or schools in terms of proportions of students in specific performance categories or in terms of growing more or less quickly based on their prior performance.

Challenges

There are a number of challenges in using growth models for educational purposes (Buzick and Laitusis 2010b). For example, some growth models require assessments that have been developed, designed, and evaluated specifically for that purpose, and longitudinal data at the school or individual level must be available. When comparing growth over time at aggregate levels (e.g., school or district), population characteristics should be relatively stable over time. In addition, the more sophisticated growth models, such as VAM or CGP, require use of complex statistical techniques that can prove difficult to explain to stakeholders, particularly if high stakes are involved (State Longitudinal Data System [SLDS] 2012). No model is perfect for all contexts or populations. Each one has strengths and limitations. It is important that inferences made from the results, and decisions to be made on the basis of particular growth models, be consistent with data and methodological assumptions. Educators are learning more about the technical aspects of growth models through an active and growing field of research that is evolving over time as new methods are developed and the strengths and weaknesses of existing ones become clearer (Betebenner 2010a; Braun 2005; CCSSO 2010; Wei, Lenz, and Blackorby 2012). Finally, decisions to use growth models need to be mindful of the context into which they are introduced. For example, the proliferation of different models (even within states) for different purposes has created a potential for confusion among constituents (SLDS 2012).

Applications to Students With Disabilities

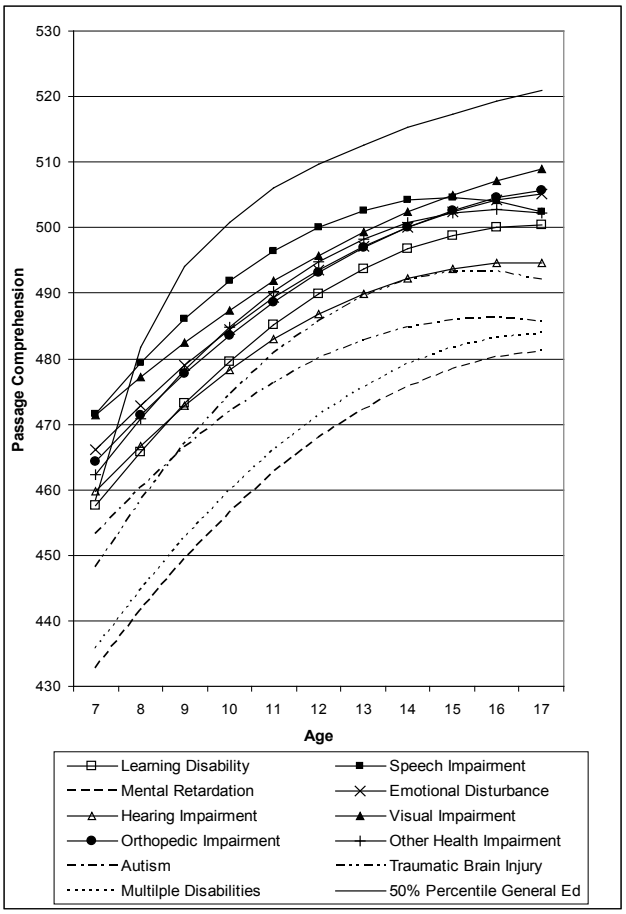
Most of the development work on growth models described in this paper has been conducted for the general student population. There is a range of additional issues that must be considered when developing growth models for students with disabilities (Buzick and Laitusis 2010a). Students with disabilities differ from their peers in the general population on a number of dimensions. They also differ from each other in

substantial and meaningful ways. States need to consider which students will be included. Policymakers need to decide if all students with disabilities should be the focus of the growth model or if they should be disaggregated by disability category (Ysseldyke et al. 1998). If disaggregation occurs, both the level and trajectory of expected growth could and does vary significantly (Wei et al. 2012). There are a number of disability categories that are low incidence, and for which numbers at the district and or school level could be too small to conduct the types of analyses required by some of the models (Thurlow et al. 2010). States must consider which assessments will be used for examining growth. Most students with disabilities will participate in the general education assessments either with or without accommodations. A small number of students with significant cognitive disabilities will be taking an alternate assessment, which may require a different growth model to describe achievement over time than the growth model that is used for general education assessments. Additional issues include use of and changes in testing accommodations and modifications, the large percentage of students performing substantially below grade level, tracking and linking of student scores across testing programs, and psychometric properties of alternate assessments (Buzick and Laitusis 2010a). Although much has been learned from the contributions by individual researchers as well as an IES-funded National Center on Assessment and Accountability for Special Education (NCAASE, http://www.ncaase.com) study about the nature of naturally occurring growth and statistical issues for modeling it, there are still many unanswered questions about how these models can be applied appropriately to students with disabilities.

Growth Trajectories for Students With Disabilities

Most of the models described in this paper calculate growth by considering the difference between two adjacent time points and then extrapolating that difference into the future, essentially assuming the rate will continue. That is, they assume that growth will proceed in a linear fashion. Even the application of some relatively sophisticated growth models (which have the potential to model nonlinear growth) assumes that growth is constant and linear. Particularly when looking at growth over a longer period of time, there is evidence that growth is actually not linear. Growth in academic skills during elementary school can be described as approximately linear, progressing at a comparable rate across grades. Studies of both general and special education students, however, suggest growth slows down in middle school and continues to flatten as students progress through high school (see Exhibit 12). The methodological upshot of these studies is that caution is required when comparing growth in elementary schools with that in upper grades. It is also the case that student performance throughout a school year tends to follow a similar pattern (Tindal 2014). There is greater growth in the fall and winter semesters followed by slower growth in the spring. States selecting SIMRs that span multiple years should consider this issue.

Exhibit 12. Growth in passage comprehension from ages 7 to 17, by disability category



Source: Wei, X., Blackorby, J., and Schiller, E. (2011). Growth in Reading Achievement in a National Sample Of Students With Disabilities Ages 7 to 17. *Exceptional Children*, *78*(1), 89–106.

Challenge of Population Shifts

Some of the models described above look at how different cohorts of students perform over time (e.g., 2015 sixth-graders compared to 2010 sixth-graders). One of the challenges in interpreting data that are collected over a period of years is that changes in the student population over time may have an effect on the observed results. For example, over a 10-year period, communities may grow or shrink, and the population characteristics relative to socioeconomic status, race/ethnicity, or language use may change as well. Therefore, observed differences in school performance, either positive or negative, could be due to population differences, rather than actual changes in performance (Buzick and Laitusis 2010). Similarly, there may be shifts in the administration policies within states that could lead to students being assessed in the general assessment in one year but on an alternate assessment at some subsequent point. While this may be a small problem for the population as a whole, this is a larger challenge for special education because there are substantial numbers of students who move in and out of testing programs. Finally, just on the basis of shifts in the population and the eligibility requirements under IDEA, the population of special education students served across grade levels differs substantially (Ysseldyke and Bielinski 2002).

For example, speech and language impairments is among the most common identified disability category in early elementary school, but much less common as students approach middle and high school. Relative to other students with disabilities, students with speech and language impairments, on average, perform better on summative assessments than peers in other disability categories. A comparison between third-grade and eighth-grade performance at the group level may reflect the reduced number of students with speech and language impairments, rather than real differences in the populations. These factors need to be considered in the design of state SSIP plans as well as in interpreting baseline and outcome data.

Summary

The growth models described in this paper represent some potentially useful tools that state policymakers can consider as they continue their efforts to implement SSIPs and select appropriate SIMR measures. These growth models share an advantage over status models because they explicitly consider change over time. The models differ in the data and methods they require and the types of inferences they support. The choice of a specific model is important because each makes different assumptions; poses different requirements on the state; and supports different types of inferences about the progress of individuals, schools, or the state. States should carefully weigh the advantages and disadvantages of specific models to evaluate their suitability for SSIP efforts.

References and Bibliography

Ahearn, E. (2009). Growth Models and Students With Disabilities: Report of State Interviews. Alexandria, VA: National Association of State Directors of Special Education. Retrieved Dec. 7, 2016, from http://nasdse.org/DesktopModules/DNNspot-Store/ProductFiles/89_301cac3e-e6aa-437d-875e-049abd89b21d.pdf.

Anderson, D., Rowley, B., Alonzo, J., and Tindal, G. (2014). Criterion Validity Evidence for the EasyCBM CCSS Math Measures: Grades 6-8 (Technical Report No. 1402). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Aspen Institute. (2006, August). Growth Models: An Examination Within the Context of NCLB: Commission Staff Research Report. Retrieved April 8, 2009, from https://assets.aspeninstitute.org/content/uploads/files/content/docs/nclb/Growth_Models_and_NCLB_Report.pdf.

Auty, W. (2008, January). *Implementer's Guide to Growth Models: A Paper Commissioned by the CCSSO Accountability Systems and Reporting State Collaborative Project*. Washington, DC: The Council of Chief State School Officers. Retrieved Nov. 11, 2016, from http://ccsso.org/Documents/2008/Implementers_Guide_to_Growth_2008.pdf.

Auty, W., Bielawski, P., Deeter, T., Hirata, G., Hovanetz-Lassila, C., Rheim, J., Goldschmidt, P., O'Malley, K., Blank, R., and Williams, A. (2008). Implementer's Guide to Growth Models. Washington, DC: Council of Chief State School Officers.

Betebenner, D.W. (2009). Norm- and Criterion-Referenced Student Growth. Educational Measurement: Issues and Practice, 28(4): 42–51.

Betebenner, D.W. (2010). New Directions for Student Growth Models. Retrieved Nov. 11, 2016, from http://www.nnstoy.org/download/evalution/LinkClick%20Betebenner.pdf.

Betebenner, D.W. (2010b). SGP: Student Growth Percentile and Percentile Growth Projection/ Trajectory Functions. (R package version 0.0-6).

Betebenner, D.W. (2011). New Directions in Student Growth: The Colorado Growth Model. Paper presented at the National Conference on Student Assessment, Orlando, FL. Retrieved March 29, 2012, from http://ccsso.confex.com/ccsso/2011/webprogram/Session2199.html.

Betebenner, D., and Shang, Y. (2007) Norm and Criterion Referenced Student Growth. Washington, DC: Council of Chief State School Officers.

Blackorby, J., and Wagner, M. (2005). Students with Disabilities in Elementary and Middle School: Progress Among Challenges. In J. Blackorby, M. Wagner, R. Cameto, E. Davies, P. Levine, L. Newman, C., Marder, and C. Sumi. C. *SEELS Engagement, Academics, Social Adjustment, and Independence: The Achievements of Elementary and Middle School Students with Disabilities.* Retrieved Dec. 7, 2016, from http://www.seels.net/designdocs/engagement/All_SEELS_outcomes_10-04-05.pdf.

Braun, H. (2005). *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*. Princeton, NJ: Educational Testing Service, Policy Information Center.

Briggs, D.C., Weeks, J.P., and Wiley, E. (2008). *Vertical Scaling in Value-Added Models for Student Learning*. National Conference on Value-Added, Madison, Wisconsin.

Briggs, D.C. (2011). Making Inferences About Growth and Value-Added: Design Issues for the PARCC Consortium. PARCC Consortium.

Burdette, P. (2011). Special Education Value-Added Performance Evaluation Systems: A State-Level Focus. Alexandria, VA: NASDSE.

Buzick, H.M., and Laitusis, C.C. (2010a). A Summary of Models and Standards-Based Applications for Grade-to-Grade Growth on Statewide Assessments and Implications for Students with Disabilities (Educational Testing Service TS RR-10-14). Princeton, NJ: ETS. Retrieved March 29, 2012, from http://www.ets.org/Media/Research/pdf/RR-10-14.pdf.

Buzick, H.M., and Laitusis, C.C. (2010b). Using Growth for Accountability: Measurement Challenges for Students with Disabilities and Recommendations for Research. *Educational Researcher, 39*(7), 537–544.

Cameto, R., Knokey, A.M., Nagle, K., Sanford, C., Blackorby, J., Sinclair, B., and Riley D. (2009). National Profile on Alternate Assessments Based on Alternate Achievement Standards. A Report from the National Study on Alternate Assessments (NCSER 2009-3014). Menlo Park, CA: SRI International.

Castellano, D.E., and Ho, A. (2013). *A Practitioner's Guide to Growth Models.* Washington, DC: Council of Chief State School Officers.

Center for Public Education. (2007). *Measuring Student Growth: A Guide for Informed Decision Making.* Retrieved April 8, 2009, from http://www.centerforpubliceducation.org/site/c.kjJXJ5MPIwE/b.3570269/k.B44C/Measuring_student_growth_A_guide_for_informed_decision_making.htm.

Colorado Growth Model: *Charting the Course to Postsecondary Readiness (2009).* Available via the LDS Share at http://nces.ed.gov/programs/slds/LDSShare/SLDS.aspx.

Council of Chief State School Officers (CCSSO) (2007). Summary of State-Proposed Growth Models.

Council of Chief State School Officers (2008). *Implementer's Guide to Growth Models.* Available at http://ccsso.org/Documents/2008/Implementers_Guide_to_Growth_2008.pdf.

Council of Chief State School Officers (2009, March). *Guide to United States Department of Education Growth Model Pilot Program 2005-2008. A Paper Commissioned by the CCSSO Accountability Systems and Reporting State Collaborative*. Retrieved May 8, 2009, from http://www.ccsso.org/Resources/Publications/Guide_to_United_States_Department_of_Education_Growth_Model_Pilot_Program_2005-2008.html.

Council of Chief State School Officers. (2010). *Key elements for educational accountability models in transition: A guide for policymakers.* Retrieved Nov. 16, 2016, from http://ccsso.org/Documents/2010/Key_Elements_for_Accountability_2010.pdf.

Doran, H., and Izumi, L.T. (2004). *Putting Education to the Test: A Value-Added Model for California*. San Francisco, CA: The Heartland Institute.

DuBois Area School District (2007). *Practical Application of Value-Added Growth Model.* Retrieved Nov. 11, 2016 from http://nces.ed.gov/programs/slds/LDSShare/SLDS.aspx.

Dunn, J.L., and Taylor, M.A. (2007, April). Modelling Growth: Purposes and Consequences of Using Various Models. Paper Presented at the Annual Meeting of the Council of Measurement in Education, Chicago, IL.

Early Childhood Technical Assistance Center. (2014). *Federal Reporting Categories for Child Outcomes Data.* Retrieved Nov. 11, 2016, from http://ectacenter.org/~pdfs/eco/Federal_reporting_categories.pdf.

FairTest Examiner. (2010, May). *Testing Consortium Guidelines Unlikely to Improve Assessments.* Retrieved Nov. 11, 2016, from http://fairtest.org/testing-consortium-guidelines-unlikely-improve-ass.

Farley, D., Saven, J.L., Tindal, G., and Nese, J.F. (2013) *Analysis of Growth on State Tests for Students with Significant Cognitive Disabilities*. Eugene, OR: BRT, University of Oregon.

Fuchs, L.S. (2004). The Past, Present, and Future of Curriculum-Based Measurement Research*. School Psychology Review*, *33*: 188–192.

Fuchs, L.S., Fuchs, D., Compton, D.L., Bryant, J.D., Hamlett, C.L., and Seethaler, P.M. (2007). Mathematics Screening and Progress Monitoring at First Grade: Implications for Responsiveness to Intervention. *Exceptional Children, 7:* 311–330.

Gersten, R., Compton, D., O'Connor, R., Dimino, J., Santoro L., Linan-Thompson, S., and Tilly, D. (2009). *Assisting Students Struggling with Reading: Response to Intervention and Multitier Intervention in the Primary Grades*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Gibson, N. (2011). *Primer on Growth Models and Their Application.* Bethesda, MD: National Forum on Education Statistics.

Goldhaber, D. and Theobald, R. (2012). *Do Different Value-Added Models Tell Us the Same Things?* Carnegie Foundation for the Advancement of Teaching. Retrieved Nov. 11, 2016, from http://www.carnegieknowledgenetwork.org/wp-content/uploads/2012/10/CKN_2012-10_Goldhaber_Nov2013-Update.pdf.

Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S., Blank, R., and Williams, A. (2005). *Policymakers Guide to Growth Models for School Accountability: How do Accountability Models Differ?* Washington, DC: Council of Chief State School Officers. Retrieved April 9, 2009, from http://www.ccsso.org/Documents/2005/Policymakers_Guide_To_Growth_2005.pdf.

Goldschmidt, P., Choi, K., and Beaudoin, J.P. (2012). *Growth Model Comparison Study: Practical Implications of Alternative Models for Evaluating School Performance*. Washington, DC: Council of Chief State School Officers.

Gong, B. (2007). *Learning Progressions: Sources and Implications for Assessment*. Retrieved May 8, 2009, from http://www.nciea.org/publications/LearningProgressions_BG07.pdf.

Grady, M., Lewis, D., and Gao, F. (2010*). The Effect of Sample Size on Student Growth Percentiles*. In annual meeting of the National Council on Measurement in Education. Denver, CO.

Hasbrouck, J., and Tindal, G. A. (2006). Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. *The Reading Teacher, 59*(7): 636–644.

Hess, K. (2008). *Developing and Using Learning Progressions as a Schema for Measuring Progress*. Retrieved August 8, 2009, from http://www.nciea.org/publications/CCSSO2_KH08.pdf.

Hill, R. (2001). *The Impact of Including Special Education Students in Accountability Systems.* Retrieved April 8, 2009, from http://www.nciea.org/publications/CCSSOSpecialEd_Hill01.pdf.

Hill, R., Gong, B, Marion, S, DePascale, C., Dunn, J., and Simpson, M.A. (2006). Using Value Tables to Explicitly Value Growth. In R. Lissitz (Ed.), *Longitudinal and Value Added Models of Student Performance* (pp. 255–290). Maple Grove, MN: JAM Press.

Ho, A.D. (2008). The Problem with "Proficiency": Limitations of Statistics and Policy Under No Child Left Behin*d*. *Educational Researcher, 37*(6): 351–360.

Hoffer, T.B., Hedberg, E.C., Brown, K.L., Halverson, M.L., Reid-Brossard, P., Ho, A.D., and Furgol, K. (2011). *Final Report on the Evaluation of the Growth Model Pilot Project*. Washington, DC: U.S. Department of Education. Retrieved Nov. 11, 2016, from http://www2.ed.gov/rschstat/eval/disadv/growth-model-pilot/index.html.

Jenkins, J.R., Schiller, E.S., Blackorby, J., Thayer, S., and Tilly D. (2013). Responsiveness to Intervention in Reading Architecture and Practices. *Learning Disability Quarterly, 36*(1): 36–46.

Kearns, J., Towles-Reeves, E., Kleinert, H., Kleinert, J.O., and Kleine-Kracht, M. (2011). Characteristics of and Implications for Students Participating in Alternate Assessments Based on Alternate Academic Achievement Standards. *The Journal of Special Education, 45*: 3–14.

Koretz, D. (2008). A Measured Approach: Value-Added Models are a Promising Improvement, But No One Measure an Evaluate Teacher Performance. *American Educator*, *32*(2): 18–39.

Martineau, J.A. (2006). Distorting Value Added: The Use of Longitudinal, Vertically Scaled Student Achievement Data for Growth-Based, Value-Added Accountability. *Journal of Educational and Behavioral Statistics, 31*(1): 35–62.

McCardle, P., Scarborough, H., and Catts, H. (2001). Predicting, Explaining, and Preventing Children's Reading Difficulties. *Learning Disabilities Research and Practice, 16*(4): 230–239.

McCaffrey, D.F. and Buzick, H. (2014). Is Value-Added Accurate for Teachers of Students with Disabilities? Stanford, CA: Carnegie Knowledge Network. Retrieved Nov. 11, 2016, from http://www.carnegieknowledgenetwork.org/briefs/teacher_disabilities/.

McCaffrey, D.F., Lockwood, J.R., Koretz, D., Louis, T.A., and Hamilton, L. (2004). Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics, 29*(1): 67–101.

McCardle, P., Scarborough, H., and Catts, H. (2001). Predicting, Explaining, and Preventing Children's Reading Difficulties. *Learning Disabilities Research and Practice, 16*(4): 230–239.

National Center on Response to Intervention. (2010). *Essential Components of RtI – A Closer Look at Response to Intervention.* Washington, DC: Department of Education, Office of Special Education Programs, National Center on Response to Intervention.

National Center on Response to Intervention. (2012). *RTI Implementer Series: Module 2: Progress Monitoring – Training Manual.* Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention.

O'Malley, K., Blank, R., and Williams, A. (2008). *Implementer's Guide to Growth Models.* Professional Development Presentation. Retrieved Nov. 11, 2016, from http://nces.ed.gov/forum/ppt/Growth_Models.ppt.

Doran, H.C. and Izumi, L.T. (2004, June). *Putting Education to the Test: A Value-Added Model For California.* San Francisco, CA: Pacific Research Institute. Retrieved December 28, 2016, from http://www.pacificresearch.org/fileadmin/news_import/200702021_Value_Added_01.pdf.

Raudenbush, S.W. (2004). Schooling, Statistics and Poverty: Can We Measure School Improvement? (Angoff lecture). Princeton, NJ: Educational Testing Service.

Raudenbush, S.W., and Jean, M. (2012, October 15). *How Should Educators Interpret Value-Added Scores?* Carnegie Knowledge Network. Retrieved Nov. 11, 2016, from http://www.carnegieknowledgenetwork.org/briefs/value-added/interpreting-value-added/.

Riddle, W.C. (Updated 2008). *Adequate Yearly Progress (AYP): Growth Models Under the No Child Left Behind Act.* Congressional Research Service #RL33032. Retrieved April 8, 2009, from http://congressionalresearch.com/RL33032/document.php?study=Adequate+Yearly+Progress+AYP+Growth+Models+Under+the+No+Child+Left+Behind+Act.

Safer, N., and Fleischman, S. (2005). How Student Progress Monitoring Improves Instruction. *Educational Leadership, 6*: 81–83.

Shapiro, E.S., Zigmond, N., Wallace, T., and Marston, D. (2011). *Models for Implementing Response to Intervention: Tools, outcomes, and implications.* New York: Guilford Press.

State Longitudinal Data Systems. (2012). *Growth Models: Issues and Advice from the States.* Washington, DC: SLDS.

The DaSy Center. (September, 2014). *DaSy Data Systems Framework*. Menlo Park, CA: SRI International.

Thurlow, M.L., Quenemoen, R., Altman, J.R., and Cuthbert, M. (2008). *Trends in the Participation and Performance of Students with Disabilities* (Tech. Rep. 50). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M., Lazarus, S., Quenemoen, R., and Moen, R. (2010). *Using Growth for Accountability: Considerations for Students with Disabilities* (Policy Directions 21). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/Policy21.

U.S. Department of Education. (2006a, October 11). Letter to Chief State School Officers Reminding of the Opportunity to Participate in the Department's Growth Model Pilot for the 2006-07 School Year (October 11, 2006). Retrieved August 28, 2008, from http://www.ed.gov/policy/elsec/guid/secletter/061011.html.

U.S. Department of Education. (2006, January 25). *Peer Review Guidance for the NCLB Growth Model Pilot Applications*. Retrieved January 15, 2008, from https://www2.ed.gov/policy/elsec/guid/growthmodelguidance.doc.

U.S. Department of Education. (2006c, May 17). *Summary by the Peer Review Team of April 2006 Review of Growth Model Proposals.* Retrieved January 15, 2008, from http://www.ed.gov/admins/lead/account/growthmodel/cc.doc.

U.S. Department of Education. (2008, June 10). *U.S. Secretary of Education Margaret Spellings Approves Additional Growth Model Pilots for 2007–2008 school year*. Retrieved August 28, 2008, from http://www.ed.gov/news/pressreleases/2008/06/06102008.html.

U.S. Department of Education. (2009, November 4). *Race to the Top Program Executive Summary.* Retrieved Dec. 7, 2016, from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf.

U.S. Department of Education (2009). *Evaluation of the 2005–06 Growth Model Pilot Program.* Retrieved April 8, 2009, from http://www.ed.gov/admins/lead/account/growthmodel/gmeval0109.doc.

U.S. Department of Education (2011). Final Report on the Evaluation of the Growth Model Pilot Project. Washington, D.C.: Author.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. (2011). *Final Report on the Evaluation of the Growth Model Pilot Project.* Washington, DC: Author.

U.S. Department of Education, Office of Special Education and Rehabilitative Services. (2014). *36th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act.* Washington, DC: Author.

Wei, X., Blackorby, J., and Schiller, E. (2011). Growth in Reading Achievement in a National Sample of Students with Disabilities ages 7 to 17. *Exceptional Children, 78*: 89–106.

Wei, X., Lenz, K.B., and Blackorby, J. (2012). Math Growth Trajectories of Students with Disabilities: Disability Category, Gender, Racial, and Socioeconomic Status Differences from ages 7 to 17. *Remedial and Special Education*. Retrieved July 16, 2012, from http://rse.sagepub.com/content/early/2012/07/12/0741932512448253.

Wright, S.P., Sanders, W.L., and Rivers, J.C. (2006). Measurement of Academic Growth of Individual Students Toward Variable and Meaningful Academic Standards. In R.W. Lissitz (Ed.). *Longitudinal and value added modeling of student performance*. Maple Grove: JAM Press.

Wyckoff, J. (2007) Measuring Proficiency With Status and Growth Models: How Best to Maximize the Achievement of All Students? Paper Presented at the Aspen Congressional Conference, Aspen, CO.

Yen, W.M. (2007). Vertical Scaling and No Child Left Behind. In N.J. Dorans and P.W. Holland (Eds.), *Linking and Aligning Scores and Scales* (pp. 273–283). New York: Springer.

Ysseldyke, J.E., Thurlow, M.L., Langenfeld, K.L., Nelson, J.R., Teelucksingh, E., and Seyfarth, A. (1998). *Educational Results for Students with Disabilities: What do the Data Tell Us?* Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Ysseldyke, J.E., and Bielinski, J. (2002). Effect of Different Methods of Reporting and Reclassification on Trends in Test Scores for Students With Disabilities. *Exceptional Children*, *68*(2): 189–200.