Overview of Student Growth Models

White Paper

Kimberly J. O'Malley, Ph.D.
Stephen Murphy, Ph.D.
Katie Larsen McClarty, Ph.D.
Daniel Murphy, Ph.D.
Yuanyuan McBride, Ph.D.

September 2011

Abstract

The term “student growth model” has been used to refer to a variety of methods used to connect student scores over time. This paper first provides an overview of student growth modeling, including benefits and limitations. Then it describes how states use student growth models in the federal accountability system, including the growth model pilot program. Although fifteen states were approved for different growth models, the models can be classified into three general types: growth to proficiency, value/transition, and projections. Each type of model is briefly explained followed by a description of the impact that each model may have on the larger practical, psychometric, and policy decisions required when implementing student growth models.

Keywords: growth measures, student growth, growth models, progress measures

Overview of Student Growth Models

This paper provides an overview of student growth modeling, including a definition, its benefits, and its limitations. Then it describes how states use student growth models in the federal accountability system, including the growth model pilot program. Although fifteen states were approved for different growth models, the models can be classified into three general types: growth to proficiency, value/transition, and projections. Each type of model is briefly explained followed by a description of the impact that each model may have on the larger practical, psychometric, and policy decisions required when implementing student growth models.

What is a Student Growth Model?

The term "student growth model" has been used to refer to a variety of methods used to connect student scores over time. Growth models have been used to refer to changes in the proportions of students meeting proficiency from one year to the next, to score changes made by students over two or more occasions, and to measures that predict a student's future scores. However, to most accurately use the term "student growth models" requires scores that:

- can be mathematically compared from one occasion to another;
- can be connected for the same students over two or more occasions;
- show changes that indicate trait changes.

A growth model captures a student's score changes over more than one occasion and focuses on the change itself.

Although the term "growth model" has been widely used when scores over multiple occasions are combined in some way, some of the models called growth models do not meet the requirements. For example, models that compare the scores of different cohorts of students (e.g., one cohort of students in a particular grade and the subsequent cohort of students in the same grade) have been called growth models. These models often evaluate changes, or growth, in the proportion of students meeting proficiency; however, they do not compare score changes for the same students over time and are therefore not correctly labeled as student growth models. Another type of model often mislabeled as a growth model is a projection model. In most projection models, a student's current scores or scores over multiple occasions are used to project that student's future score. Though scores from more than one occasion are connected for the same students, the focus of the model is not the score change itself, but a projected score. Projection models obtain no real measure of student score change.

Why are Growth Models Important?

Growth models are important for three reasons. <u>First, they conceptually align well with one of the fundamental goals of education—student learning</u>. They do this better than status measures, or measures of the proportion of students meeting proficiency year over year on a campus. Student learning implies change over time. This goal of learning or progress applies to all students, not just those who have not reached proficiency. Our educational systems expect continuous learning from all students; therefore, measures of student growth will help evaluate those expectations. Student growth models can be valuable in combining student scores over time to improve understanding of student learning.

<u>Second, growth models can provide richer information on student learning than a single score at one point in time because they connect scores from multiple assessments</u>. In contrast, status measures use a single year's assessment results to indicate school performance and attach decision rules to those results. When a student scores at least

proficient in a single year, the implication is that the student is on track in his or her educational development, but that measure provides no specific information about the extent to which a student is on track.

Third, growth models focus on the educational development of individual students. This is the focus of most parents and students. Status measures that calculate the proportion of students meeting proficiency in a school, on the other hand, focus on the performance of the school as a whole at any given time. The student population counted for status calculations might vary from year to year. Therefore, status measures cannot reflect the educational development of individual students, especially those who do not meet the proficiency level in a single year but do show improvement based on their previous academic performance.

Limitations of Growth Models

Though growth models are important for the reasons above, there are limitations to implementing a growth model and obtaining reliable information from it. It is challenging enough to accurately measure student learning at any one point, let alone to accurately measure score changes. Students often have missing data and the ability of growth models to estimate student growth when data are missing is varied.

One particular challenge in measuring student growth is the changing nature of constructs over time. Researchers have shown that when constructs shift across grades, such as when mathematics assessments move from testing arithmetic skills in third grade to testing pre-algebra and geometry skills in later grades, the growth model results may lead to imprecise longitudinal interpretations (Reckase 2004; Martineau 2006). Another challenge relates to accurately capturing the form of students' growth. For example, some students may demonstrate linear growth, whereas others may demonstrate nonlinear growth. Most growth models assume the same functional form of growth for all participants.

Therefore the accuracy of these models will not be the same for students who show academic growth in different patterns. Those models that do capture different growth patterns tend to be the most complex and least understood.

In addition, growth models—when compared with status measures—can be very difficult for non-statisticians to understand. States implementing growth models have needed to invest much time and energy in developing and disseminating information about those models.

National Focus on Measuring Growth

A November 2005 announcement by then-Education Secretary Margaret Spellings encouraged states to propose pilot programs for ***growth-based accountability models*** in the 2005–2006 and 2006–2007 school years. Spellings listed seven requirements for the pilot programs that included growth measures in the accountability models. The first three were alignment elements and the last four were foundational elements. These requirements included:

1. The accountability model must ensure that all students are proficient by 2013–14 and set annual goals to ensure that the achievement gap is closing for all groups of students.

2. The accountability model must not set expectations for annual achievement based upon student background and school characteristics.

3. The accountability model must hold schools accountable for student achievement in reading/language arts and mathematics.

4. The accountability model must ensure that all students in the tested grades are included in the assessment and accountability system. Schools and districts must

be held accountable for the performance of student subgroups. The accountability model includes all schools and districts.

5. The state's assessment system, the basis for the accountability model, must receive approval through the NCLB (No Child Left Behind) peer review process for the 2005–2006 school year. In addition, the full NCLB assessment system in grades 3–8 and in high school in reading/language arts and math must have been in place for two testing cycles.

6. The accountability model and related state data system must track student progress.

7. The accountability model must include student participation rates in the state assessment system and student achievement on an additional academic indicator.

Furthermore, after reviewing the first round of pilot projects, the U.S. Department of Education (USDE) proposal review team published a document summarizing cross-cutting issues that influenced their decisions to approve or disapprove states' proposals (USDE 2006). In particular, it reported that states should:

1. incorporate available years of existing achievement data, instead of relying on only two years of data;

2. align growth timeframes with school grade configuration and district enrollment;

3. make growth projections for all students, not just those below proficient;

4. hold schools accountable for the same subgroups as under the status model;

5. not use wide confidence intervals;

6. not reset growth targets each year; and

7. not average scores between proficient and non-proficient students.

Though these issues were noted as influential in the peer review group's decisions, not all proposals approved through the growth model pilot peer review process met all of these conditions (CCSSO Accountability Systems and Reporting Working Group, 2009). For example, neither North Carolina nor Delaware incorporated all available student data, and yet they were approved for the federal pilot study.

Since the announcement of the USDE growth model pilot program in 2005, fifteen states have been approved to use student growth in their calculations of Adequate Yearly Progress (AYP). This program was initially designed to allow up to 10 states to include a growth measure in their federal accountability model, but two years later, USDE announced that all states meeting eligibility requirements could apply to use growth in their accountability systems (USDE, 2009). This paper describes how the models can be classified into three general types: growth to proficiency, value/transition, and projections. Details of the growth models for the federally-approved states can be found in other documents (O'Malley, 2008; O'Malley & McBride, 2011).

Growth to Proficiency

These models take initial student performance and provide a yearly growth target for students such that students will reach proficiency in a set number of years (e.g., three or four years). Each year, states compare actual student performance to target performance to determine whether students have academically progressed or regressed over the school year. Calculations for these models are fully transparent. (**AL, AZ, AR, FL, MO, NC**)

Value / Transition

According to these models, student growth is evaluated based on the changes in performance categories or performance subcategories typically over two years. For a value table approach, states determine values for transitions across performance subcategories. The subcategories are determined by subdividing the main performance categories. The specific values awarded to students are typically set by an advisory panel using a process of ranking transitions, discussing ranks, and averaging ranks over multiple rounds, much like a standard-setting activity. For a transition table approach, states also subdivide performance categories. Students are expected to make a set number of transitions across subcategories each year so that students reach proficiency by a defined number of years, typically three or four. (**DE, IA, MN, MI, TX for their 1% assessment**)

Projection

These models predict student performance in a future year (up to three years in the future). Projections are based on past and current student performance and the performance of prior cohorts in the target grades. By comparing projected values to the proficiency standard for the target grade, these models offer an indicator of the extent to which students are on track to meet proficiency. The student growth percentile method (Betebenner, 2009), used by Colorado, includes both a measure of student growth (through individual growth percentile calculations) and a projection value indicating whether students are on track to meet proficiency for use in AYP calculations. Although projection models are statistically complex and most lack transparency, the model approved and implemented in Texas is transparent and expected to be replicated at the local level. The models implemented in Ohio, Pennsylvania, and Tennessee are unable to be replicated at the district level. (**CO, OH, PA, TN, TX**)

Nationally, growth models are gaining momentum. Current Secretary of Education Arne Duncan has mentioned student growth in many recent speeches, and encourages data systems that can track student growth. He states that higher standards alone will not make a difference unless teachers and principals get the information they need to ensure that students are learning. He notes that robust data systems that allow districts to better track the growth of individual students and growth models that provide educators with useful information are key to this process.

Growth Models and Index Systems

The USDE has given little formal guidance on how states can implement a growth model in addition to an index system in their accountability systems. In Secretary Spellings's November 21, 2005 initial letter to chief state school officers announcing the growth model pilot program, USDE encouraged states to consider implementing indexes. The letter suggested that an index could be implemented while a state developed the necessary elements for a growth model. One of the attachments to the letter states, "An index model provides a way for states that cannot track individual student progress to implement a form of accountability that captures subgroup growth, thus providing an alternative to the status model of AYP decision-making that only gives a school credit for the percentage of students at the proficient or above level."

A few states have successfully obtained approvals for both a growth model and an index system. Michigan, for example, was approved to implement a transition table growth model in addition to a proficiency index for AYP calculations. The proficiency index that Michigan implemented combined grade-level proficiency information and targets to make AYP decisions across grades. Pennsylvania was approved to implement a projection growth model in addition to using the Pennsylvania Performance Index in 2009 AYP calculations. Other states have had more difficulty obtaining USDE approval to use a growth measure in addition to a performance index. The June 10, 2008 USDE letter to Minnesota announcing that Minnesota's initial proposal was not approved stated, "The heart of the peer's concerns relates to the interaction of Minnesota's existing performance index with the value table. The Department has concerns about the appropriateness of allowing a state to include both a performance index and a growth model in its accountability system." (Briggs, 2008).

When Minnesota resubmitted its proposal in fall 2008, the state documented how the growth model and performance index interacted using 2008 AYP data and successfully illustrated that the addition of the growth model will not dilute accountability for the state.

Because indexes and growth measures both allow states flexibility in AYP calculations, states submitting proposals in the future will need to provide evidence showing that the implementation of both measures does not violate NCLB core principles. In addition, evidence suggests that USDE has been gathering input from experts on this issue. The October 2008 USDE Federal Register publication states that, "…the National TAC [Technical Advisory Committee] will consider complex technical issues that affect all States, and on which the Department would benefit from discussions with experts in the fields of educational standards, assessments, accountability, statistics, and psychometrics (e.g., the appropriate use of confidence intervals and performance indexes)." (p. 64468, USDE, 2008). More formal guidance on this issue might be published by USDE once it obtains recommendations from the National TAC.

Considerations for Measuring Growth

As states have implemented growth models, they have had to address practical, psychometric, and policy considerations. These considerations include whether the different growth models can be applied to the state assessment systems, whether the state data system meets the requirements of certain growth models, and how growth model results can be included in AYP calculations. Table 1 summarizes several important considerations for each of the three types of growth models. Note that though issues are summarized for the different types of growth models, there is variation among models within each of the three types.

Table 1. Considerations for Implementing Different Types of Growth Models

Issue	Growth to Proficiency Models	Value/Transition Tables	Projection Models
Description of Model Type	➢ Models inform about whether students are on track to meet the proficiency standard at some specified point in the future. ➢ The number of years states specify varies, typically 3 or 4 years. ➢ Models focus on score changes over past years.	➢ Models evaluate student transitions across performance levels (i.e., Unsatisfactory, Limited Knowledge, Proficient, Advanced) or subdivisions of performance levels (e.g., Low Proficient, Medium Proficient, High Proficient). ➢ Models focus on student changes over 2 years. ➢ States using these models subdivide performance levels and expect students to progress across performance sub-levels in such a way that students reach proficiency in a set number of years (typically 3 or 4).	➢ These models project student performance in the future. ➢ These models use sophisticated regression formulas to make projections. ➢ Using these models, states project separately for reading and mathematics. ➢ These models use student scores in the projection content area as well as in other content areas.
States Implementing Model Type	Alaska, Arizona, Arkansas, Florida, Missouri, North Carolina	Delaware, Iowa, Michigan, Minnesota	Colorado, Ohio, Pennsylvania, Tennessee, Texas

Issue	Growth to Proficiency Models	Value / Transition Tables	Projection Models
Inclusion of Students	➢ In general, this method will allow growth to be reported on most students.	➢ This model will provide growth information for all students with scores in two consecutive years on the same assessment.	➢ Due to the complexity of these models, they are typically not used with students taking alternate assessments. The one exception is Texas, which implements its projection measure for students taking the 2% assessment.
Replication and Verification at the Local Level	➢ These models are generally transparent to all. ➢ Districts and campuses would be able to calculate growth for their students. ➢ The steps for calculating growth with one of these models would be published and examples would be shown to enable districts to fully understand the model.	➢ These models are generally transparent to all. ➢ Districts and campuses would be able to calculate growth for their students. ➢ The steps for calculating growth with one of these models would be published and examples would be shown to enable districts to fully understand the model.	➢ The statistical sophistication and proprietary complex data manipulation techniques needed for several of these models to make exact replication is not possible in most states. ➢ The one exception is Texas, which implements a model that is fully transparent and replicable locally.
Reliability	➢ Moderate reliability in decisions about meeting growth targets.	➢ Moderate reliability in decisions about meeting growth targets.	➢ Moderate to high reliability in projecting student performance in the future.

Issue	Growth to Proficiency Models	Value / Transition Tables	Projection Models
Data Requirements	➢ States need to match student scores across grades in order to calculate and report growth with these types of models.	➢ States need to match student scores across pairs of consecutive grades in order to calculate and report growth with these types of models.	➢ States need to match student scores across grades in order to develop projection equations. ➢ For most of these models, the state would need to match student scores across grades to report growth. ➢ It is possible to report projections using only current year scores.
Growth Attributed to Educators	➢ The numbers of students in classes, schools, and districts that meet growth targets in one year and over several years might be used as one indicator of how well teachers, schools, and districts are performing.	➢ The numbers of students in classes, schools, and districts that meet growth targets in one year and over several years might be used as one indicator of how well teachers, schools, and districts are performing.	➢ Most of these models can be adapted to estimate the effect of grades, campuses, districts, and possibly teachers on student growth, though the accuracy of these value-added estimates has been questioned.
Application to End-of-Course (EOC) Exams	➢ Extending these models to EOC assessments will pose some challenges.	➢ These types of models can be extended to EOC assessments.	➢ These types of models can be extended to EOC assessments.

Issue	Growth to Proficiency Models	Value / Transition Tables	Projection Models
Timing of Reports	➢ Growth information from these types of models should be able to be included using current timelines.	➢ Growth information from these types of models should be able to be included using current timelines.	➢ Reporting growth for many of these models is done after the regular reporting of student scores due to the need to conduct analyses with statewide data to obtain projection measures. ➢ One exception is that Texas reports their projections in the current timeline by taking advantage of previous cohorts of students.
Link to College Readiness	➢ These types of models are not typically used to measure growth to college readiness.	➢ These types of models can be used to measure growth to college readiness.	➢ These types of models are often used to make projections to college readiness.

References

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4), 42-51.

Briggs, K. (2008, June 10). [Letter to Minnesota Department of Education]. Retrieved from USDE Web site: http://www2.ed.gov/admins/lead/account/growthmodel/mn/mnresponse.pdf

CCSSO ASR Working Group (2009). *Guide to United States Department of Education Growth Model Pilot Program 2005-2008.* Retrieved from the Council of Chief State School Officers Web site: http://www.ccsso.org/Resources/Publications/Guide_to_United_States_Department_of_Education_Growth_Model_Pilot_Program_2005-2008.html

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.

O'Malley, K. (2008, June). *Review of student growth models used by states.* Retrieved from Pearson Web site: http://www.pearsonassessments.com/NR/rdonlyres/E2F72E40-5CBD-48B9-9033-2830295A8DA8/0/Student_Growth_Models.pdf?WT.mcn_id=TMRS_Review_of_Student_Growth_Models

O'Malley, K., & McBride, Y. (2011, in preparation). *Summary of federally-approved student growth models.*

Reckase, M. D. (2004). The real world is more complicated than we would like. *Journal of Educational and Behavioral Statistics*, *29*, 117-120.

Spellings, M. (2005, November 21). [Letter to Chief State School Officers]. Retrieved from USDE Web site: http://www2.ed.gov/policy/elsec/guid/secletter/051121.html

United States Department of Education (2006, May 17). *Summary by the peer review team of April 2006 review of growth model proposals.* Retrieved from: http://www2.ed.gov/admins/lead/account/growthmodel/cc.doc

United States Department of Education (2008, October 29). Section 200.22 National Technical Advisory Committee (National TAC). *Federal Register, 73*(210), 64467-64471. Retrieved from: http://www.gpo.gov/fdsys/pkg/FR-2008-10-29/pdf/E8-25270.pdf

United States Department of Education (2009, January 8). *Secretary Spellings approves additional growth model pilots for 2008-2009 school year.* Retrieved from: http://www2.ed.gov/news/pressreleases/2009/01/01082009a.html