Applications of Growth Models within the Framework of No Child Left Behind

Sharon Cadman Slater

Carolyn Wentzel

Linda Chard

ETS, Princeton, NJ

Paper presented at the annual meeting of the National Council on Measurement in Education (NCME)

April 14-16, 2009, San Diego, CA.
Abstract

Fifteen states currently have approval to use growth models to determine whether school districts meet their No Child Left Behind (NCLB; 2001) targets. This paper describes the first nine state models approved in the growth model pilot. Beginning with the 2007-2008 academic year, the program is now eligible to all states. To be approved, states have to meet certain “Core Principles.” Interestingly, two of these principles actually limit the complexity of the statistical models that can be applied to measure growth. One requires that all students become proficient by 2013-2014; another excludes the use of demographic variables as covariates. In this way, the models are not very different from the original NCLB status model. A newer pilot program, Differentiated Accountability, has been introduced that can utilize value-added models to determine growth. Demographics still cannot be used in calculating AYP, but results from value-added models can affect how schools who fail AYP are treated. This paper examines how a number of states are applying the flexibility of growth modeling for NCLB.
Introduction

In November 2005, the United States Department of Education (USDE), under the direction of Margaret Spellings, announced that up to 10 states would be allowed the opportunity to use a growth model, in addition to using the NCLB status model, to calculate student academic performance and Academic Yearly Progress (AYP). To this end, the growth model pilot program (GM) was established.

In April 2007, Spellings described five policy priorities for reauthorization of the No Child Left Behind Act (NCLB); one of which was to “give States flexibility to better measure individual student progress” by incorporating growth models for making determinations of AYP. In fact, this change to the legislation may be one of the few things upon which the administration and congressional lawmakers can agree. The desire is to measure performance of individual students over time rather than comparing the scores of students in a certain grade to students in that grade the previous year. The Miller-McKeon NCLB reauthorization draft (Section 1123) called for states to institute longitudinal data systems within 4 years of the reauthorization date in order to collect the necessary information to make growth modeling possible. NCLB was up for reauthorization in 2007, but has yet to be acted on by Congress.

As of March 2009, as part of the GM, fifteen states have received approval from the USDE to use growth models as part of the determination as to whether school districts meet AYP targets. The states participating in the GM are Alaska, Arizona, Arkansas, Colorado, Delaware, Florida, Iowa, Michigan, Minnesota, Missouri, North Carolina,
Ohio, Pennsylvania, Tennessee, and Texas. This paper will describe the first nine state models included in the GM.

In order to be considered for the GM, states first had to meet certain criteria demonstrating that they were following the *bright line* goals of NCLB. These goals include annual assessment of students in grades 3-8 and one grade in high school, disaggregation of data by student subgroups, improving teacher quality, and informing parents of their educational options. Of highest importance was the goal of all students achieving at least grade level proficiency in mathematics and reading by the year 2014. Proposals from states that met those criteria advanced to the peer review stage. In January 2008, the growth model pilot was expanded to all eligible states for the 2007-2008 school year. These submissions followed the same peer review process as the initial applications.

Seven Core Principles are outlined in the peer review guidelines. A state must demonstrate that their proposed growth model meets these principles in order to be given approval to participate in the GM. The Core Principles are:

**CORE PRINCIPLE 1** -- The proposed accountability model ensures that all students are proficient by 2013-14 and sets annual goals to ensure that the achievement gap is closing for all students.

**CORE PRINCIPLE 2** -- The proposed accountability model establishes high expectations for low-achieving students, while not setting expectations for annual achievement based upon student demographic characteristics or school characteristics.

**CORE PRINCIPLE 3** -- The proposed accountability model produces separate accountability decisions about student achievement in reading/language arts and in mathematics.
CORE PRINCIPLE 4 -- The proposed accountability model ensures that all students in the tested grades are included in the assessment and accountability system. Schools and districts will be held accountable for the performance of student subgroups. The accountability model, applied statewide, will include all schools and districts.

CORE PRINCIPLE 5 -- Annual assessments in reading/language arts and math in each of grades 3-8 and high school must have been administered for more than one year, must produce comparable results from year to year and grade to grade, and must be approved through the peer review process for the 2005-06 school year.

CORE PRINCIPLE 6 -- The accountability model and state data system must track student progress.

CORE PRINCIPLE 7 -- The accountability system must include student participation rates in the state's assessment system and student achievement on an additional academic indicator.

Interestingly, the first two Core Principles actually limit the complexity of the statistical models that can be applied to measure growth. Core Principle 1 specifically requires that all students must become proficient by 2013-2014. This task is easier or more difficult depending upon students' initial achievement levels. If a student is at a very low achievement level, large gains must be made in order to be on track to become proficient. If a student is at a high achievement level, relatively small gains are needed in order to be on track to become proficient. In this way, the models are not very different from the original NCLB status model.

Core Principle 2 has resulted in some states, which have more statistically rigorous models accounting for student and/or school demographic characteristics, not being included in the GM; this includes the value-added models being applied in Pennsylvania and Tennessee. In order to meet the GM requirements, Tennessee proposed a projection model, rather than their value-added model. The goals of these two models are very
different in that value-added models seek to compare schools to each other or to a growth standard in a more equitable way by taking into account demographic characteristics. The goal of projection models is to determine whether students are “on track to proficiency.” Value-added models do not necessarily require all students to ever reach proficiency, where the projection models hold different schools to different standards, depending on how much progress must be made by 2013-2014.

In addition to the 7 Core Principles, during review of the first set of proposals, the peer review group outlined additional elements they would consider in evaluating the proposals. These additional considerations are based on the “Crosswalks” paper (retrieved from http://www.ed.gov/admins/lead/account/growthmodel/index on May 30, 2008):

The peer review group determined that states should not:
1. use wide confidence intervals,
2. reset growth targets each year, and
3. average scores between proficient and non-proficient students

They also added that states should:
1. incorporate all available years of existing achievement data, instead of relying on only two years of data,
2. align growth timeframe with school grade configuration and district enrollment,
3. make growth projections for all students, not just those below proficient, and
4. hold schools accountable for the same subgroups as required by status model.

To summarize the implementation of the GM from June 2005 – May 2008, Table 1 presents a brief timeline of events related to the incorporation of growth models into school accountability policy in the nine states.
Table 1. Timeline of the U.S. Department of Education Growth Model Pilot Program*

<table>
<thead>
<tr>
<th>Date</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>November 21, 2005</td>
<td>USDE announces the eligibility criteria for States to apply for the growth model pilot program</td>
</tr>
<tr>
<td>February 2006</td>
<td>Twenty states apply for the GM: Alaska, Arizona, Arkansas, Colorado, Delaware, Florida, Hawaii, Indiana, Iowa, Maryland, Nevada, New Hampshire, North Carolina, Ohio, Oregon, Pennsylvania, South Carolina, South Dakota, Tennessee, and Utah</td>
</tr>
<tr>
<td>March 2006</td>
<td>Proposals from 8 states forwarded to peer review for consideration in the growth model pilot: Alaska, Arizona, Arkansas, Delaware, Florida, North Carolina, Oregon, and Tennessee (7 states applies for the 2006-2007 school year, so were deferred – 5 states were rejected, including Colorado, Indiana, Iowa, South Carolina, and Utah)</td>
</tr>
<tr>
<td>May 17, 2006</td>
<td>USDE announces approval for Tennessee and North Carolina for the 2005-2006 school year</td>
</tr>
<tr>
<td>September 15, 2006</td>
<td>Deadline for the 5 states that were previously peer-reviewed to submit revised proposals to the USDE for consideration for the 2006-2007 school year, Alaska, Arkansas, Delaware, Florida, and Oregon.</td>
</tr>
<tr>
<td>November 1, 2006</td>
<td>Deadline for states to submit new growth model proposals to the USDE for the 2006-2007 school year. Limit of approved states remains at 10; 9 states applied for the remaining slots – Alaska, Arizona, Iowa, Hawaii, Ohio, Nevada, New Hampshire, Pennsylvania, and Utah</td>
</tr>
<tr>
<td>November 9, 2006</td>
<td>USDE announces Delaware and Arkansas for approval, and Florida for conditional approval for the 2006-2007 school year</td>
</tr>
<tr>
<td>May 1, 2007</td>
<td>Deadline for the 5 states that were asked to submit revised proposals: Alaska, Arizona, Iowa, New Hampshire, and Ohio</td>
</tr>
<tr>
<td>May 24, 2007</td>
<td>USDE announces Iowa for approval and Ohio for conditional approval for the 2006-2007 school year</td>
</tr>
<tr>
<td>June 26, 2007</td>
<td>USDE announces Florida for full approval in the GM for the 2006-2007 school year</td>
</tr>
<tr>
<td>August 15, 2007</td>
<td>USDE announces Ohio for full approval in the GM for 2006-2007</td>
</tr>
<tr>
<td>December 7, 2007</td>
<td>Expansion of the GM to all eligible states for the 2007-2008 school year; no longer a cap on the number of states that can participate</td>
</tr>
<tr>
<td>April 2008</td>
<td>Peer review of growth models submitted from the District of Columbia, Michigan, Minnesota, Missouri, New Mexico, and Pennsylvania.</td>
</tr>
<tr>
<td>June 2008**</td>
<td>Michigan and Missouri were approved to participate in the GM.</td>
</tr>
<tr>
<td>January 2009**</td>
<td>Colorado, Minnesota, Pennsylvania, and Texas were approved to participate in the GM.</td>
</tr>
</tbody>
</table>

* States appear in boldface when they were approved for the GM.  
** Descriptions of these state models are not included in this paper.
II. Summary of State Specific Growth Models

As mentioned earlier, fifteen states are currently implementing various types of growth models to measure school accountability. Nine of these state models will be briefly described in this section. When determining AYP, in general, states first apply status and safe harbor to determine if a school has met AYP. If not, the state then applies the growth model. If the school meets its growth target, then it is considered to have met AYP for the school year. In applying growth models for AYP purposes, the approved model is applied to determine how test performance compares across years (e.g., regression, value tables, projection models, etc.). The performance is then compared to performance goals established to move all students to proficient by 2014 to determine if a school has met AYP or not. An overview of the state growth models is presented in Table 2.

Table 2. First Nine States Participating in the Growth Model Pilot Program

<table>
<thead>
<tr>
<th>State (1st year)</th>
<th>Type of Growth Model</th>
<th>Years to reach proficiency</th>
<th>Vertically Aligned?</th>
<th>Vertically Scaled?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska (2007-08)</td>
<td>Growth to Proficiency – student must decrease the gap between initial score and the proficient score by 1/x, where x is the number of grades remaining</td>
<td>Within 4 years or by 10th grade, whichever comes first</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Arizona (2006-07)</td>
<td>Growth to Proficiency – targets set by subtracting the previous year scale score from the proficient scale score 3 years later and dividing by the number of grades remaining</td>
<td>Within 3 years or by 8th grade, whichever comes first</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State (1st year)</td>
<td>Type of Growth Model</td>
<td>Years to reach proficiency</td>
<td>Vertically Aligned?</td>
<td>Vertically Scaled?</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Arkansas (2006-07)</td>
<td>Growth to Proficiency – targets expect larger gains the first year after the student scores below proficient than in following years</td>
<td>Within 4 years or by 8th grade, whichever comes first</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Delaware (2006-07)</td>
<td>Performance Index, Value Table – assigns points based on combination of student’s performance level in 2 consecutive years. Individual points for each student, then averaged by subgroup. Calculates growth for all groups.</td>
<td>All students proficient by 2013-14</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Florida (2006-07)</td>
<td>Growth to Proficiency – sets a slightly steeper improvement slope, requiring students to meet state standards in three years for schools to achieve AYP</td>
<td>Within 3 years</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Iowa (2007-08)</td>
<td>Growth to Proficiency</td>
<td>Within 3 years</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>North Carolina (2005-06)</td>
<td>Standard Score Trajectory Model – Based on Z-scores</td>
<td>Within 4 years</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ohio (2007-08)</td>
<td>Projection Model – growth trajectories based on a student’s past performance on all previous state tests (reading, math, science, and social studies) but separate trajectories calculated for math and reading</td>
<td>Within 4 years or by the grade beyond the top level of the school where the student was initially tested, whichever comes first</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Tennessee (2005-06)</td>
<td>Projection Model – has set aside “value-added” model in favor of system based on projections of student growth three years into the future. Students currently proficient who are predicted to fall below proficient 3 years in the future do not count toward AYP.</td>
<td>Within 3 years</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Summary of the Alaska Growth Model

Alaska uses a Growth to Proficiency Model, which is applied only to students who are below Proficient. If students can be shown to be “on track to be proficient,” their numbers are added to those who are already at proficient or above. On track to be proficient is defined as closing the gap between the student’s base scale score (estimated true score) and the scale score for proficiency (300). The amount of the gap that the student has to close is dependent on the student’s score the previous year, the student’s grade, and the number of years the student has been in the school district.

Over a three year period, 2004-2007, Alaska adopted grade level expectations that measure similar, yet more sophisticated, content across grade levels. Thus, the standards on which the assessments are built are vertically aligned; however, the scores are not on a vertical scale.

A feature of the Alaska Proposal is that it requires substantial growth in year one and does not compensate for less growth in a subsequent year (non-compensatory model).

One criticism of the Alaska Growth Model is that it provides no incentive to improve performance of those already at proficient or above. Additional concerns that are specific to the state of Alaska are:

- Alaska does not have an alternative assessment for students with disabilities
- There is no alternate assessment for limited English proficient students.
- Twenty-three percent of Alaska schools have three or fewer teachers.
- A majority of schools in the State serve very small student populations.
- Many of the districts are rural and reachable only by plane.
From the state-wide assessment results, Alaska is making progress in improving student academic performance for all groups and specifically for minority groups. The following show examples of that growth:

- Of the schools that were identified for not meeting adequate yearly progress in 2007 (38.4%), 14.7% did not meet AYP partly as a result of the performance of Alaska Native students. This is reduced from 17.5% two years previous.
- Of the same schools, 14.5% did not meet AYP this year partly as the result of the performance of economically disadvantaged students. This also was reduced from 19.3% two years previous.

**Summary of the Arizona Growth Model**

Arizona also uses a **Growth to Proficiency Model** that incorporates growth targets and predicted scores. A 97.5% confidence interval is built around the predicted score. The confidence interval takes measurement error into account by using the standard error of prediction. If the lower bound of the confidence interval is greater than or equal to the student’s growth target, the student is said to have met their growth target for that particular year.

To determine whether a subgroup of students meets annual measurable objectives (AMOs) for a particular grade and subject, a percentage is calculated using the number of students meeting their growth target (regardless of proficiency) divided by the total number of students in the subgroup. If this percentage is greater than or equal to the AMO specific to that grade and subject, the group is considered to have made adequate yearly progress (AYP).
Students in grades three and four will have three consecutive years to attain proficient status; students in grades five through seven are expected to achieve proficiency by grade eight. Arizona’s vertical scale ends at the eighth grade.

Arizona's Instrument to Measure Standards (AIMS) has four performance levels: Falls Far Below (FFB), Approaches (A), Meets (M), and Exceeds (E). The last two levels are considered proficient.

The same performance levels are used for Arizona's alternative assessments. Two types of alternative tests are administered to the most significantly cognitively disabled students, AIMS-A and AIMS-A level II. Students taking the AIMS-A are included in the growth model; if these students move up a performance level, they are considered to have met their growth target.

**Summary of the Arkansas Growth Model**

The Arkansas growth model uses a *Growth to Proficiency* target based on the difference between a particular proficiency score and the student’s current score. Trajectories change annually to reflect students’ current scores. For students who are not proficient, the growth target is the Grade 8 proficiency scale score. Growth increments are calculated by creating a ratio of the proficient scale score for the next grade and the proficient scale score for the current grade, divided by the difference between the proficient scale score for Grade 8 and the proficient scale score for the current grade. This fraction is then multiplied by the difference between the proficient scale score for
grade 8 and the student’s current scale score. The resulting value is added to the
student’s current scale score and is used as the minimum scale score the student must
attain in the next grade to be deemed as making growth for that particular year.

A school, and the subgroups within the school, is considered to have made AYP if the
percentage of students who are considered to be “making growth” is equal to the AMO
for proficiency. For example, the Arkansas AMO dictates that, for mathematics in grades
K-5, 64.08% of students in each school and subgroup within the school should be
proficient in 2007-2008. Consequently, at least 64.08% of the students must make the
required growth in mathematics for a particular school to make AYP in 2007-2008.

Baselines for grades 3-8 in Reading and Mathematics were established in 2005.
All students are expected to be proficient by the eighth grade. Arkansas’ vertical scale
ends at the eighth grade. The state performance levels are Basic, Proficient, and
Advanced.

Arkansas’ alternate assessments are aligned to alternate achievement standards and
administered to students with the most significant cognitive disabilities. Almost none of
these special needs students will ever change to assessments based on grade level
standards, consequently these students are not included in Arkansas’ growth model.
Summary of the Delaware Growth Model

Delaware uses a Value Table Model for measuring student growth that relies on awarding points to the school based on the change in students’ performance across vertically-articulated achievement standards. Value tables are constructed to award a school (or district) varying amounts of points depending on how much progress, in terms of movement across achievement standards, a student makes from one year to the next. This value table approach, in addition to several conceptual advantages (e.g., it is not based on vertically scaled scores), is quite transparent for school leaders to understand how changes in student performance are translated into changes in school accountability scores.

This method for measuring progress assigns a certain number of points to each of the various combinations of levels student performance in consecutive years. For example, if a student scored at Level 1 in Year 1 and Level 2 in Year 2, that result would earn the student’s school a certain number of points. If, on the other hand, the student were Level 2 in Year 1 and then Level 1 in Year 2, that result would earn the student’s school no points. One aspect of this system is that students who are Level 1 cannot go down, no matter how little progress they might make from year to year, and students who score at Level 5 cannot go up, no matter how much progress they make. This “floor” and “ceiling” effect is an inherent limitation.
Delaware has five levels of performance consistent across grades 3 through 10 in reading, writing, and math. In grade 2, there are three levels of performance for reading and math since the grade 2 assessment has fewer items than the assessments at the other grades.

Delaware will calculate AYP based on status and safe harbor for all schools and subgroups that meet the minimum sample size requirement of 40 (the “traditional model”). Delaware will also calculate AYP for proficiency based on the value table growth model methodology for all schools and subgroups that meet the minimum sample size requirement of 40. The participation rate, other academic indicators, and sanctions from the traditional model will remain the same and will carry over to the growth model. By calculating proficiency both ways, Delaware will have information that will be useful in analyzing how this growth model performs and how the results compare to the AYP traditional model. A school that makes AYP based on either the traditional model or the growth model will be deemed as meeting AYP.

The value table approach is unique in terms of student longitudinal growth based accountability systems because it incorporates explicit state values into the accountability system. The Delaware tables do not value growth above proficient, and growth to proficiency is valued less than being proficient (similar to NCLB). This means that students scoring in performance levels 3 through 5 earn the same number of points. There is no compensation for growth above proficiency and if all of the students in a subgroup were proficient in Year 1 and stayed proficient, the average score for the subgroup would be 300. Likewise, if all of the students in the subgroup moved from
anywhere below proficiency in year 1 to proficiency in Year 2, the average score for the subgroup would be 300 points. The maximum score, therefore, is 300, which is equivalent to 100% proficient, and all students are meeting the standards. Note that this is not a compensatory model that rewards growth above proficiency. It rewards growth towards proficient and staying proficient.

A one-tailed confidence interval, to be determined after further investigation, will be used in the growth model. The rationale for using a confidence interval in the growth model is the same as the rationale for using confidence intervals in the traditional status model. Confidence intervals are used to control for sampling errors or measurement errors, thereby increasing the validity and reliability of classifying schools as making or not making AYP. The confidence interval will be applied to the average growth value.

**Summary of the Florida Growth Model**

Florida’s **Growth to Proficiency Model** is implemented in grades 3-10 for reading and mathematics. The assessment used to measure student academic performance, which is aligned to the Florida Sunshine State Standards (SSS), is the Florida Comprehensive Assessment Test (FCAT) Developmental Scale Score (DSS), which has a continuous scale from grade 3 through grade 10. FCAT is vertically aligned and has a continuous developmental scale score (DSS) ranging from 0 to 3000 for grades 3 - 10. Approximately one percent of students take a reading alternate assessment and approximately one percent of students take a mathematics alternate assessment.
Florida calculates the students’ average annual projected growth rate by taking the difference between the students’ current scale scores and the students’ first scale score and dividing the difference by the number of years the student has been in school over that time. Students must be “on track to proficient” to be added to the proficient numbers for AYP.

For each year, the growth target is a one-third (33.33 percent) decrease in the difference from the baseline score on the state’s growth scale to the score necessary to be proficient on the test three years after entry into a tested grade in Florida. A student will be considered “on track to be proficient” if his or her individual trajectory meets or exceeds the proficiency cut points each year to attain the proficiency threshold in three years or less. A student will not be considered “on track to be proficient” unless their proficiency trajectory will meet or exceed the proficiency threshold in three years or less based on the annual cut point benchmarks.

One of the attributes of the model is that it is easy for the public to understand. It does not rely on complex statistical procedures or imputed values for students; does not use covariance matrices to estimate or project student performance; does not use a multivariate, or multi-level model, to determine a student’s growth trajectory.
The following lists some of the major components of the Florida Growth Model:

- A high performing student cannot compensate for the lack of growth of a low performing student.
- A student who scores proficient is weighted the same as a student who is on track to be proficient only if the proficient student meets his or her annual benchmark.
- Proficient students are included in growth trajectories for AYP purposes but proficient students not meeting their growth targets will not be considered “on track to be proficient.”
- Students who use alternate assessments that are not on the FCAT developmental scale will have growth calculated based on progress or improving achievement levels or maintaining a proficient level.

Currently, Florida is beginning a new six-year cycle of review and revision of the SSS that will focus first on the reading/language arts standards and the mathematics standards. Because the degree of change in the standards is unspecified at this time, the possible change in FCAT, and consequently, on the measurement of growth is unknown.

**Summary of the Iowa Growth Model**

The Iowa Growth Model is a **Growth to Proficiency Model** in which a student’s growth trajectory must cross a category boundary in order to be considered for Adequate Yearly Growth. Adequate Yearly Growth (AYG) is the measure Iowa uses in their computation for AYP. This is intended to insure that regardless of the beginning achievement level, a student will be proficient within four years. In this model, AYG must be greater than expected growth in order to ensure that students are on track to proficiency. It is not compensatory. Only students who are not proficient in Year 1 are eligible for the growth determination in Year 2, if they continue to be non-proficient. Because of the vertical alignment of the score scale for the Iowa Tests, as well as the vertical alignment of content standards, the expectation is that a student scoring at a level that is below
proficient will need to make a substantial improvement in achievement in order to be counted as “on track to be proficient.” Students will have four consecutive years to attain proficient status, beginning with their initial participation in the statewide assessment. Under the proposed model, a particular student may only count for AYG a maximum of two times in their entire testing history. For Iowa, growth targets are defined as movement from one proficiency level to the next.

Iowa currently utilizes three main achievement levels (Low, Intermediate, and High) for AYP decisions plus some additional level divisions. The Low achievement level is further divided into two achievement levels: Weak and Marginal.

The Marginal level contains two regions: Low Marginal (containing the lowest part of the Marginal level) and High Marginal (a region that contains the scale score points that are within one standard error of measurement below the cut score for Proficient). Intermediate achievement level reflects students who are proficient. High achievement level reflects students who have achieved the highest level of proficiency.

Non-proficient students who make “Adequate Yearly Growth” or AYG (defined as a movement from the Weak achievement level to the Marginal level, or from the Low Marginal region to the High Marginal region) will be counted as proficient for AYP purposes. Iowa’s model does not incorporate the use of a regression-type model or a multivariate or multi-level model.
The Iowa Growth Model proposal does not include students with the most significant cognitive disabilities who participate in the Iowa Alternate Assessment. Students who take The Iowa Alternate Assessment are classified in three achievement levels: non-proficient, proficient, and “better than proficient.”

**Summary of the North Carolina Growth Model**

North Carolina employs a Standard Score Trajectory model that expects students to be proficient within four years of entering the tested grades. The trajectory calculations are applied to 3rd through 8th grades. A pretest administered at the start of 3rd grade is used as the baseline.

North Carolina uses a Standardized Scale Approach (SSA) to growth which uses the normative distribution of student performance in the standard setting year of any test as a common basis to build a scale. The SSA system uses a time-locked modified z-scale. The scale is termed a “change scale” or “C-scale.” The C-scale cut score for proficiency on any given test edition at an individual grade level remains constant for the life of the scale and test edition regardless of the changes in the distribution of test scores that might occur as schools change their instructional methods. The state means and standard deviations from the standard setting year are used indefinitely for any given test.

Separate trajectories are built for each individual student for reading and mathematics. The following steps are used to calculate individual growth increments and growth targets by means of a modified z-scale called the C-scale.
• Each scaled score is transformed to the C-scale using the mean and standard deviation of test scores from the year the test was initially administered (baseline year).
• The difference between the C-scale equivalent of the proficiency score for the grade by which the student is expected to be proficient and the C-scale equivalent of the student’s score in the baseline year is divided by the number of years the student has to meet the proficiency target score.
• A student’s growth target is calculated by adding this difference to the C-scale equivalent of the student’s baseline score.
• A student meets their growth target if their C-scale score from the following year is greater than or equal to their C-scale growth target for that year.

Students who lack the pretest or baseline scores and students who take alternate assessments that are not on the growth scale are not included in the growth model calculations. The participation of these students is limited to their absolute status. It is important to note that proficient students are not included in growth trajectories for AYP purposes. However, North Carolina uses its ABCs growth formulas for proficient students to determine the percentages of all students meeting growth expectations at a subgroup and school level. A student who scores proficient receives the same weighting as a student who is deemed to be on target. Therefore, the growth of high-performing students does not compensate for the lack of growth among other students.

Summary of the Ohio Growth Model

Ohio uses a Projection Model as an additional method of meeting AYP. The projection model will calculate student trajectories toward proficiency as a second look, once the conventional assessment of whether all subgroups are meeting the status or Safe Harbor goal for percent proficient is made. Under this provision, if one or more subgroups
(including the *all students* group) falls short of the status goal and safe harbor, but the
school or district demonstrates that students are making gains such that they are on track
to reach or remain proficient by the next grade beyond the school’s grade configuration
(for example, a K-5 school would use the projection measure to determine the likelihood
of students reaching proficiency by grade 6; for high schools, the projection will
determine whether students are on-track to meet proficient levels on the OGT by the end
of 11th grade), then the district or school will meet AYP. Schools and districts that fall
short of the conventional AYP measure *and* are failing to make gains necessary to be
proficient by the next grade level will fail to meet AYP.

Ohio uses a projection model as outlined by Wright, Sanders, and Rivers (2006). The
projection model uses up to five years of available test scores for individual students,
merged longitudinally, to provide the best estimates of future student achievement
trajectories. Ohio incorporate a measure of projected student growth to include students
who are on a path to reach proficiency by the end of the first year at their next school in
the percent of students counted as proficient for AYP. Conversely, proficient students
whose projected growth indicates that they will fall below proficient by the end of the
first year at their next school will be subtracted from the number of students counted as
proficient for AYP. Ohio’s longitudinal database includes state developed and
administered achievement and graduation test scores for students in grades 3-8, 10, and
11, starting with school year 2003-04.
The projection model uses all available test scores for each child to create an individual growth trajectory in the appropriate subject. Student-level projections can be created for any subject and any test metric; however test scores for all subjects, including science and social studies, are analyzed to improve the precision of the estimates. The primary assumption upon which successful modeling rests is that past test achievement is a good predictor of current and future achievement. By using all of an individual student’s previous achievement data to make these projections, the need to adjust for a student’s socioeconomic condition, ethnic background, prior knowledge, etc. is not necessary. This ensures that students with the same prior academic achievement will have the same projection regardless of the neighborhood in which the attending school is located or the student’s demographic characteristics.

While the proposed projection model uses multivariate longitudinal data, this application is fundamentally different from many value-added models. Value-added methodology typically has been applied to compare expected gains to observed gains. Because expected gains were derived from historical achievement patterns of similar students, expected gains differed for different students – the expectations were “conditioned” on the patterns of past performance for similar students. In other words, past applications of value-added methodology compared a student’s academic gains to the typical achievement gains of similar students. Thus, application of value-added modeling is at odds fundamentally with the spirit of NCLB. The core expectation of AYP is that all students reach proficiency and that achievement expectations not differ based on historic performance of similar students.
With the proposed projection model, past performance is employed to identify the likely achievement trajectory for students with similar academic histories. This projected trajectory is then compared to the trajectory needed to reach proficiency by the target grade. For students already proficient, the projected trajectory is compared to the trajectory needed to remain proficient by the target grade. In other words, the model compares each student’s expected trajectory to the trajectory needed to be proficient.

Some additional information about the Ohio growth model includes:

- **Confidence Intervals** – Ohio does not employ confidence intervals or tests of statistical significance when determining percent proficient or safe harbor for purposes of AYP determinations. Consistent with current Ohio practice regarding the determination of percent proficient and safe harbor, Ohio will not employ confidence intervals to determine the group percent on track to proficiency in the future.

- **Uniform Averaging** – Ohio will continue to employ uniform averaging when determining AYP using status and Safe Harbor. Ohio will not average when applying the growth model to the AYP determination.

- **AYP Formula Issues** – Ohio employs a minimum group size of 30 (45 for students with disabilities), which it will continue to apply to the status and Safe Harbor calculations (see the response to Core Principle 4).

- **Assessments** – Ohio employs alternate assessments for the most severely cognitively disabled students. Ohio does not employ native language assessments. Ohio’s projection model will employ projected scores based on alternative assessments against the alternative assessment scale and add these into the appropriate subgroup, school, and district calculations in the same manner that the projection for students based on the regular assessment are included (see response to Core Principle 1).

- **Higher-achieving Students** – As identified in Core Principle 1, when applying the growth model, a trajectory will be calculated for students who are proficient and projected to the target grade. If the already-proficient students are projected to remain proficient, they are counted as proficient for purposes of calculating the percent proficient. If the already-proficient students are projected to fall short of proficient, they are not counted as proficient.
Summary of the Tennessee Growth Model

The Tennessee Projection Model (Wright, Sanders, & Rivers, 2006) uses individual student projection data to determine the percent of students, by subgroup and subject area, who are projected to attain proficiency on the state assessment three years into the future. It uses 7th and 8th grade projections for 4th and 5th grade students, respectively, and uses high school graduation exam projections for 6th – 8th grade students. The model uses current-year scores for 3rd grade students, students new to the state, and students who take alternative assessments.

Schools and districts meet AYP proficiency requirements through the projection model if all subgroups meet the annual measurable objective in both reading/language arts and mathematics.

Of Tennessee’s two growth models – a value-added model that estimates district, school, and teacher effect scores and a projection model that estimates individual students’ projected scores on future assessments – only one is appropriate for the NCLB growth model pilot program. The value-added model, which measures whether districts, schools, and teachers provide sufficient instruction for their students as a group to make one year of progress each year, is an innovative mechanism to drive academic progress for all students but is clearly not aligned with NCLB’s precise goal that each individual student will reach proficiency. The projection model, meanwhile, by predicting each student’s future achievement relative to state standards, is more in line with NCLB.
The projection methodology uses all of an individual student’s prior achievement scores to estimate the student’s achievement level at a future point in time. For each student, all available past TCAP scores, as far back as grade 3, are used as predictors in the projection model. The model’s only predictor variables are the student’s prior test scores. By assuming that the student will have the average Tennessee schooling experience in the future, it includes estimated mean scores for the average school in Tennessee and regression coefficients that are pooled within schools across the state. These coefficients are updated each year as a new student cohort acquires test scores at the projection endpoint. The only source of the model’s complexity is missing data – not all students have prior achievement scores for all subjects at all grades/years.

This methodology does not require vertically linked data nor does it need to assume a linear growth function (or any other specific growth function). Instead, what is required are good predictors of the response variable. The predictors need not be on the same scale with the response or with one another. Potentially, they could be test scores from different vendors and even in different subjects from the response. This gives the methodology considerable flexibility.

The projection model sets a very high standard. Schools and districts may meet AYP proficiency requirements under the projection model only under the following strict conditions:
1. Each subgroup’s projected percentage of students who score proficient or advanced on reading/language arts meets the approved annual measurable objective for reading/language arts; and

2. Each subgroup’s projected percentage of students who score proficient or advanced on mathematics meets the approved annual measurable objective for mathematics.

The projection model assigns school credit for all students who are projected to be proficient three years into the future, whether they are currently below proficient or are currently proficient. It does not assign schools any credit for students who are currently proficient but are projected to score below proficient on the future assessment. It does not assign schools any additional credit for students who score advanced.

The projection model includes current scores for 3rd grade students and other students who are in their first tested year in Tennessee. If the student scores proficient in the current year, he or she will be counted as proficient in the projection model. If the student scores below proficient in the current year, he or she will be counted as below proficient in the projection model.

Additional points:

1. The status model will continue to use uniform averaging across two and three years. The projection model will not use uniform averaging.

2. The minimum group size will continue to be 45 (or 1%, whichever is greater) and the projection model will apply this policy.

3. The confidence interval will continue to be 95% but the projection model will not apply this policy.
4. The projection model will use projected scores for students who took a regular assessment in the current year. It will use current-year alternative assessment scores for students who took these exams, should these scores’ inclusion fall under current policy.

5. The projection model includes projected scores of all students (subject to the exemptions described above). Students whose score is above the cut for proficiency will be counted as proficient. Students whose score is below the cut for proficiency will be counted as below proficient. It does not “credit” schools for students who have projections above proficiency.

6. The State will publicly report data from the projection model in a manner consistent with its traditional reporting of AYP data, substituting aggregate projection scores. It will continue to make individual student projection data available to educators to use in instruction and to share with students and parents.

III. Why Proposals Were Not Approved

Why did initial state growth model proposals fail to be approved? A reason commonly given for rejection of a growth model proposal was that the model did not insure that 100% of the students would reach proficiency by 2014. For example, Utah’s proposal was rejected because the model (U-PASS) required only 75 percent of students to meet the proficient level in reading and math by 2014, rather than the 100 percent target required by the federal law. The state’s plan also proposed combining student subgroups, such as all African-American, Hispanic, or low-income students at a school, into one large, school-level subgroup for accountability calculations. Similar proposals that failed to meet the 100% target requirement were also returned for revision and resubmission.

A second requirement typically not met in early proposals was that the state must have a student identification system (ID) that assigns a unique identification number to each student. The lack of such a state ID system is understandable in that a central database must be maintained which requires a significant investment of resources to develop,
install, and maintain. This limitation is expected to become less of an issue as nearly all states are expected have this capability in place in the near future.

The proposals of some states were not approved, because their system did not meet the overarching USDE requirements. For example, after being turned down during the review of their initial growth model proposal, Oregon reapplied and was again denied. Federal officials cited a recent state revision of standards as an example of instability within the state’s educational system. A second example of a fundamental proposal flaw was found with the state of Delaware, which was initially rejected in March, 2006. In the original proposal, Delaware argued that students demonstrate they have progressed a year academically when they score at the same achievement level for two consecutive years. However, the federal peer-review panel stated that students scoring in the “below the standard” level on the state’s tests must move up one category to a level of proficiency to be considered as making adequate progress. When Delaware resubmitted their application, they increased the size of a test-score increment required for a student to be counted as making adequate academic progress.

The initial proposals of some states were rejected because the proposed growth model included student demographic characteristics (e.g., an HLM model). This was the case for Tennessee, who in order to meet the federal requirements for a growth model abandoned their long established accountability model that included student demographic characteristics and reestablishment of growth target annually. South Carolina’s proposal was rejected because higher thresholds for subgroup sizes were set for students with
disabilities and students with limited English skills compared to other subgroups in determining whether schools and districts met growth targets.

Proposals from other states were not originally approved for reasons that were relatively minor and easier to adjust, such as changing the size of the confidence interval or the minimum sample size for a subgroup. For example, Ohio's growth model was approved on the condition that the state adopt a uniform minimum group size for all subgroups, including students with disabilities and limited English proficient students, in AYP determinations for the 2006-2007 school year.

IV. Differentiated Accountability

Recall that the idea of using growth models is that they theoretically give schools more time to bring all students up to proficiency. However, because of the constraints placed on the growth models (namely that schools still must achieve proficiency for all students by 2014) schools with the lowest performance have unreasonably steep projections to move all students to proficient in the set amount of time. Schools with higher achieving students have much less work to do to get their students over the bar than schools with the lowest achieving students. This makes the GM models, in practice, extremely similar to the original status measure. Indeed, an Education Daily (Sawchuk, 2008) states that “most states using growth models under ED’s program did not dramatically increase the number of schools meeting proficiency targets compared to the status calculation.” Because the GM stipulates that performance must be measured against predetermined
proficiency levels, rather than against relative performance of other students and schools, growth models do not add much, over and above the original (and easier to implement) status measures of AYP. Because the status and projection models are so similar, few schools are making AYP because of “growth” alone.

In an *Education Week* article from June 2008, Michael Weiss wrote:

“While value-added measures of school performance have their own problems, these are the models people typically think of when they envision an accountability system that includes a growth component. If value-added models were used, we would identify quite a few low-status, high-growth schools… Value-added models are not allowed under the growth-model pilot program because they don’t adhere to the core principle of NCLB—to bring all students up to proficiency. But they do represent the fairest (albeit imperfect) way to compare schools’ effectiveness. The dilemma over which measure of school performance to use highlights an inherent tension when designing an accountability system for schools, one between the desire to compare their relative effectiveness (value-added models) while simultaneously holding them accountable for bringing all students up to high achievement levels (status or projection models). Some people thought that the pilot program’s projection models were a happy middle ground. Unfortunately, projection models don’t address the essential tension between status and growth.”

A newer pilot program has been introduced that targets interventions to the lowest performing schools that are working under these seemingly unrealistic growth trajectories. In March 2008, Secretary Spellings announced a program called the Differentiated Accountability pilot (DA), which allows states to distinguish between schools falling below AYP targets who need substantial help, and those who are nearly meeting their goals. The pilot will allow states to vary type of interventions brought to bear depending on whether a school needs comprehensive intervention or more focused help.
As of July 2008, six states had been approved for the DA, including: Florida, Georgia, Illinois, Indiana, Maryland, and Ohio. In January 2009, another three states were added to the program: Arkansas, Louisiana, and New York. While this program cannot be used to circumvent the requirements of NCLB, it does help states to target resources to the schools most in need of significant reform.

It is not yet clear how the reauthorization of NCLB will be conducted, or what role growth models might play in that reauthorization. The topic seems to have moved from a question about whether or not growth models should be used for school accountability to which models would best measure growth for NCLB. The Center for Public Education (2007) highlights a number of important policy questions that will need to be answered including: Which growth models best meet the goals of NCLB? How will growth models affect the goal of one hundred percent proficiency by 2014? How much flexibility should states have when designing growth models for NCLB? Would growth models have a greater effect on improving schools if they were designed to inform instruction rather than for high-stakes accountability?

While growth models are not perfect, several researchers agree that they are probably better than the accountability systems that are currently in place (McCaffrey, et al., 2003). It is clear is that growth models can provide valuable information that is not available when only considering achievement status. What remains to be seen is how growth models can best be utilized for the improvement of schools in our efforts to truly leave no child behind.
References


