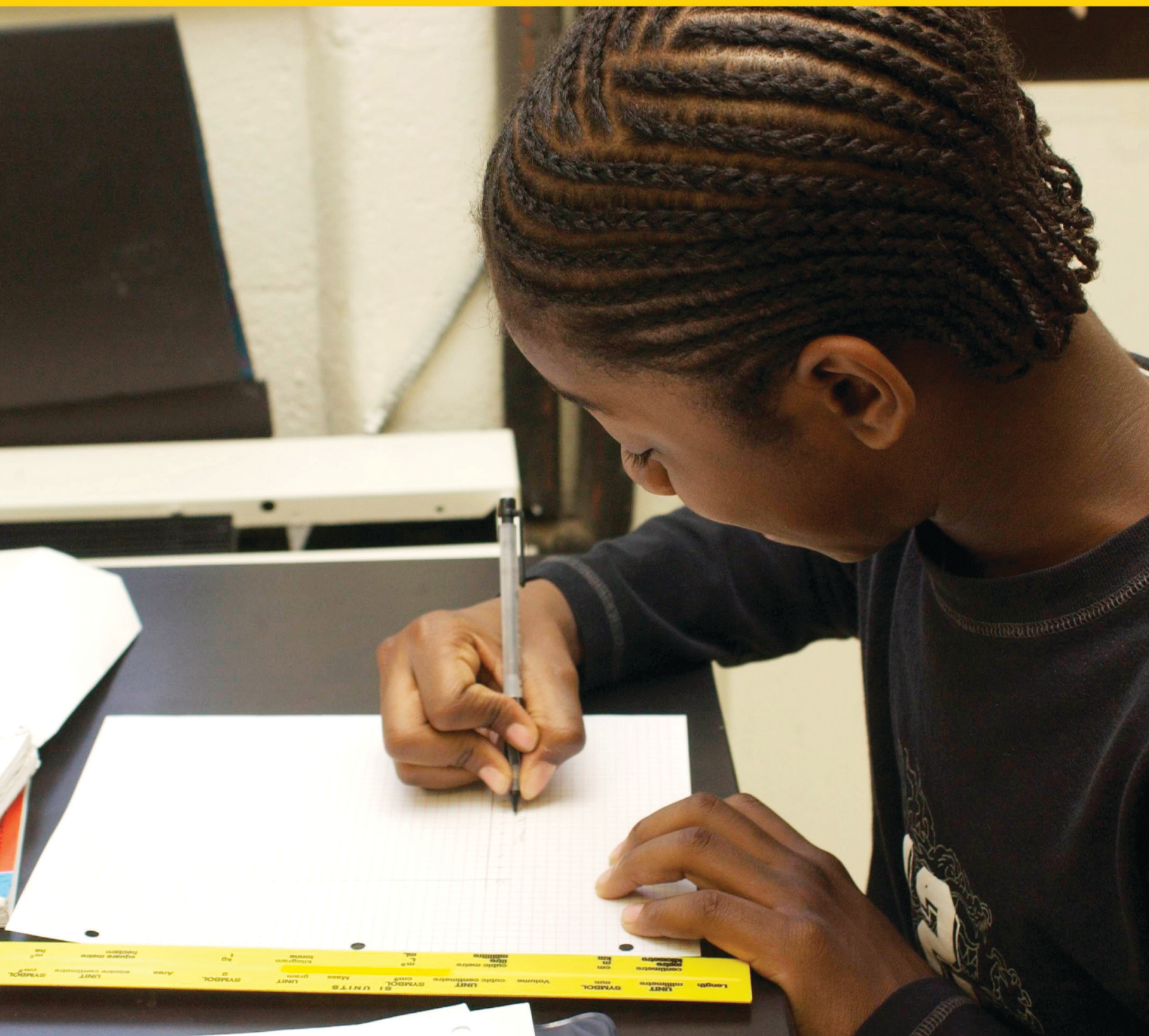
U.S. DEPARTMENT OF EDUCATION

Interim Report on the Evaluation of the Growth Model Pilot Project



Interim Report on the Evaluation of the Growth Model Pilot Project

---

Prepared by:

Thomas B. Hoffer
E. C. Hedberg
Kevin L. Brown
Marie L. Halverson
Sarah-Kathryn McDonald

National Opinion Research Center (NORC)
at the University of Chicago

Prepared for:

U.S. Department of Education
Office of Planning, Evaluation and Policy Development
Policy and Program Studies Service

2010

This report was prepared for the U.S. Department of Education under Contract Number ED-06-CO-0053 with National Opinion Research Center at the University of Chicago. Andrew Abrams served as the contracting officer's representative. 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 is intended or should be inferred.

U.S. Department of Education
Arne Duncan
Secretary

Office of Planning, Evaluation and Policy Development
Carmel Martin
Assistant Secretary

Policy and Program Studies Service
Alan Ginsburg
Director

April 2010

To obtain copies of this report:

Write to ED Pubs, Education Publications Center, U.S. Department of Education, NTIS, P.O. Box 22207, Alexandria, VA 22304

Or **fax** your request to 703-605-6794.

Or **e-mail** your request to edpubs@edpubs.ed.gov.

Or **call** in your request toll free to 1-877-433-7827 (1-877-4-ED-PUBS). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 1-877-576-7734. If 877 service is not available in your area, call 1-800-872-5327 (1-800-USA-LEARN).

Or **order online** at http://edpubs.ed.gov.

This publication is also available at the Department's Web site at http://www.ed.gov/about/offices/list/opepd/ppss/reports.html#title.

On request, this publication is available in alternate formats, such as Braille, large print or computer diskette. For more information, contact the Department's Alternate Format Center at 202-260-0852 or 202-260-0818.

Contents

Exhibits

Acknowledgments

This report benefited from the efforts and thoughtful input of many individuals. Andrew Abrams at the U.S. Department of Education's Policy and Program Studies Service (PPSS) directed the project and provided helpful guidance and encouragement at each stage. David Goodwin, Daphne Kaplan, and Stephanie Stullich, also at PPSS, offered much constructive advice and feedback throughout the project.

The data analyzed in this report were drawn from the U.S. Department of Education's ED*Facts* archive and from each of the separate state departments of education participating in the Growth Model Pilot Project. We are grateful to Ross Santy and Susan Thompson-Hoffman for their generous help and patience with the ED*Facts* data. The Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, and Tennessee state departments of education were very responsive to our various requests for data and information about their pilot programs, and we would particularly like to acknowledge the help of Les Morse, Robert Franciosi, Denise Airola, Robin Taylor, Ed Croft, Thomas Deeter, Kenneth Barbour, and Dan Long.

We were very fortunate to receive thoughtful and timely reviews from several individuals both within and outside the U.S. Department of Education. In particular, we received extremely helpful guidance in the early stages of the project and detailed feedback on early drafts of the report from Patrick Rooney at the Office of Elementary and Secondary Education within ED. We also thank Abigail Potts and Victoria Hammer of OESE for sharing their expertise in the design and review stages.

The project had a Technical Working Group consisting of David Figilo, Thomas Fisher, Pete Goldschmidt, Rachel Quenemoen, and Keith Zvoch. The TWG provided guidance with the study design and analysis plans, and contributed many useful comments and corrections on an early draft of this report.

At NORC, the project has benefited greatly at several key junctures from the insightful perspectives of Stephen Raudenbush and Larry Hedges. In the later stages of drafting the Interim Report, Marie Halverson and Cindy Simko joined the project as project director and associate project director, respectively, and contributed greatly to organizing and implementing the many reviewer comments. The task of formatting this report in each of its iterations was handled with great skill and efficiency by Imelda Demus and Yajaira Gijon.

Executive Summary

The U.S. Department of Education (ED) initiated the Growth Model Pilot Project (GMPP) in November 2005 with the goal of approving up to ten states to incorporate growth models in school adequate yearly progress (AYP) determinations under the *Elementary and Secondary Education Act* (*ESEA*). After extensive reviews, eight states were fully approved for the pilot project in the 2006–07 school year: Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, and Tennessee. Based on analyses of data provided by the U.S. Department of Education and by the pilot grantee states, this report describes the progress these states made in implementing the GMPP in the 2006–07 school year.

GMPP Objectives

Use of growth models for determining AYP is attractive to states and local districts because it offers a means to identify schools making progress even though they may not yet be reaching proficiency standards. Without recognition of their progress, these schools would be subject to school improvement actions that may not be appropriate in light of their demonstrable improvements.

The standard method of determining AYP has been the "status model," in which school performance is mainly evaluated in terms of the proportion of students meeting or exceeding proficiency standards for reading and mathematics. Growth models measure how much students have gained from one year to the next using longitudinal records of individual student achievement in reading and mathematics. The models determine whether each student was "on-track" to reach or exceed the state's proficiency cut points (or thresholds) on the annual tests of reading and mathematics within three or four years or by a specified grade level (usually grade eight or nine) as defined by the state's particular growth model. For purposes of determining AYP, students who were not proficient but on-track could be counted the same as proficient students.

States Included in the Pilot Project

The GMPP began in 2005 with two states—North Carolina and Tennessee—approved to use growth models for *ESEA* accountability in the 2005–06 school year. The project expanded to include eight states approved to implement growth models in 2006–07, 11 states in 2007–08, and 15 states in 2008–09. This interim report focuses only on the eight states approved for 2006–07.[1] A final report will examine results for the 2007–08 school year for the eight states plus Ohio, which first implemented in 2007–08. In 2007 and 2008, six additional states were approved: Michigan and Missouri (beginning in the 2007–08 school year); and Colorado, Minnesota, Pennsylvania, and Texas (beginning in the 2008–09 school year). While the GMPP began in 2005 with a goal of approving up to 10 states, under current Title I regulations, there is no limit now on the number of states that can be approved.

[1] An evaluation by the Office of Elementary and Secondary Education of growth model results in North Carolina and Tennessee for the 2005–06 school year is available at http://www.ed.gov/admins/lead/account/growthmodel/gmeval0109.doc.

Findings

Features of growth models implemented by pilot states

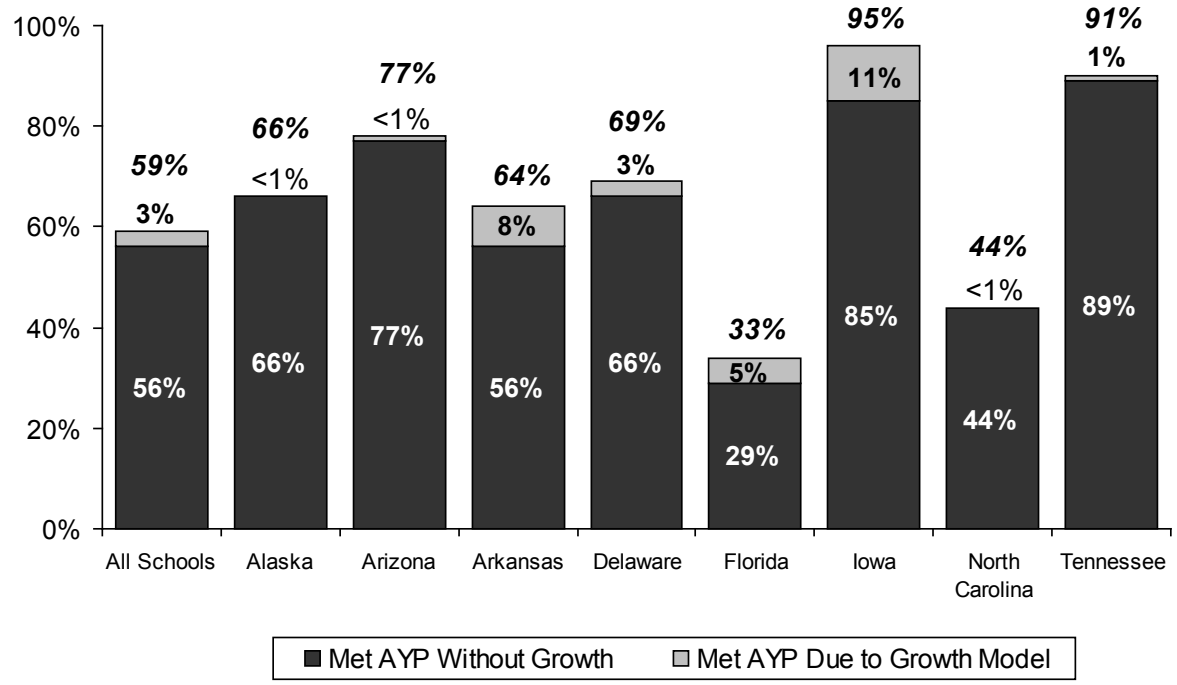
The growth models implemented under the GMPP were all designed to augment rather than replace the standard status model and safe-harbor provisions for determining school AYP. The growth models resulted in more schools making AYP than would have been the case using only status and safe-harbor.

Seven of the eight states used growth criteria only after schools failed to make AYP under the status and safe-harbor provisions. Delaware was the exception, applying growth results before status and safe-harbor. The designs of the pilots in the other seven states applied growth criteria only to *ESEA* reporting groups that did not reach their annual measurable objectives (AMOs) or obtain AYP via safe-harbor provisions. Furthermore, in all but two of those seven states, the growth criteria were applied *within* the *ESEA* reporting groups only to the students who did not reach the proficiency threshold. The number of non-proficient but on-track students was added to the number of proficient students and the reporting group was counted as meeting the AMO if the total was high enough.

Comparing AYP by growth with status and safe-harbor

The growth models in the eight states that provided data for the 2006–07 school year resulted in some schools making AYP that would not have made AYP under status or safe-harbor alone (see Exhibit S.1). Among all schools, 3 percent made AYP uniquely because of the growth model (that is, they did not make AYP by status or safe-harbor). The percentages of all schools that made AYP uniquely by growth varied widely among the states, ranging from zero or one percent of all schools in Alaska, Arizona, North Carolina, and Tennessee to 11 percent of all schools in Iowa.

Exhibit S.1
Percentage of Schools That Made AYP Before and After the Application of the Growth Model, in Eight States, in 2006–07



As the bar chart illustrates, the pilot states varied greatly in the proportion of their schools that made AYP under status or safe-harbor. The schools making AYP uniquely by growth represented a percentage increase in the schools making AYP (the dark gray column in Exhibit S.2) of 6 percent across all states, and ranged as high as 12 percent in Iowa, 14 percent in Arkansas, and 16 percent in Florida.

Another measure of GMPP impact is the extent to which it reduced the number of schools that would not have made AYP had the growth model not been available. From this perspective, the number of schools that did not make AYP by either status or safe-harbor was reduced by 8 percent overall because of the GMPP (the medium gray column in Exhibit S.2). The percentage reduction was highest in Iowa (69 percent), Arkansas (17 percent), and Tennessee (13 percent).

Exhibit S.2
Percent of Schools Making AYP Because of Use of the Growth Model, Percentage Increase in Number of Schools That Made AYP Due to Growth, and Percentage Decrease in Number of Schools That Did Not Make AYP Due to Growth, by State, 2006–07

Pilot States	Number of Schools	Number of Schools Making AYP by Status or Safe-Harbor	Number of Schools Not Making AYP by Status or Safe-Harbor	Number of Schools Not Making AYP by Status or Safe-Harbor That Met by Growth	Percent of All Schools That Met by Growth	Percentage Increase in Schools Making AYP Due to Growth	Percentage Decrease in Schools Not Making AYP Due to Growth
All Eight States	10,883	6,088	4,795	371	3%	6%	8%
Alaska	492	323	169	0	0%	0%	0%
Arizona	1,430	1,096	334	1	<1%	<1%	<1%
Arkansas	894	499	395	69	8%	14%	17%
Delaware	185	123	62	5	3%	4%	8%
Florida	3,209	925	2,284	149	5%	16%	7%
Iowa	1,104	936	168	116	11%	12%	69%
North Carolina	2,207	969	1,238	12	<1%	1%	<1%
Tennessee	1,362	1,217	145	19	1%	2%	13%

Exhibit reads: A total of 10,883 schools in the pilot states included grade levels to which the growth models were applied, and 371 of those schools (3 percent) made AYP because of the growth model. Of the 6,088 schools that made AYP under either status or safe-harbor, use of the growth model increased the number of schools making AYP by 6 percent. Of the 4,795 schools that did not make AYP under either status or safe-harbor, use of the growth model reduced that number by 8 percent.

Source: U.S. Department of Education, ED*Facts* and the Alaska, Arizona, Delaware, North Carolina, Ohio, and Tennessee state departments of education.

Impact of Growth Models on AYP Rates Among High-Poverty Schools

Schools serving disadvantaged populations have been found to make AYP at much lower rates than those serving more affluent populations (U.S. Department of Education, 2007). The growth model pilots may reduce these associations to some extent by identifying high-growth schools serving low-income and minority communities.

The results of this analysis showed that schools serving economically disadvantaged student populations in all pilot states except for Delaware were more likely than more-advantaged schools to make AYP by growth. Across all eight states, the percentage increase in the number of high-poverty schools making AYP as a result of the growth model being available was 8 percent, compared to 3 percent among low-poverty schools. The percentage increases among high-poverty schools in Arkansas, Florida, and Iowa were five to six times greater than those among low-poverty schools.

Hypothetical Results of Using Growth Models Instead of Status and Safe-Harbor

Some have suggested that growth should be the primary accountability indicator for all schools, including those that made AYP by status. The student data were used to assess the extent to which the schools that made AYP by status (excluding safe-harbor) would also have met or exceeded their AMO for reading and mathematics proficiency if the growth criteria of on-track-to-proficiency were used instead of the status criteria.

Overall, 85 percent of the schools that made AYP by status criteria in the eight states also would have met their AMO strictly by using the growth criteria. Results varied widely among the states: only 45 percent of the Arizona schools that made AYP by status also met their AMO using growth criteria alone, but almost all of the North Carolina (99 percent) and Florida (97 percent) schools that made AYP by status also met their AMO with the growth criteria.

The percentages of schools that made AYP by safe-harbor and that also met or exceeded their AMO under the growth-only criteria were also high (87 percent overall) and consistent across the six states with safe-harbor schools (Alaska and Arizona had none), ranging from 75 percent in Tennessee to 99 percent in North Carolina.

Qualifications and Implications

The results of this analysis show that use of growth models will generally add to the number of schools making AYP, especially among schools serving low-income populations, which can help administrators identify schools "on-track" and those that may need help to get "on-track."

The growth targets identified by the growth models are tied to the *ESEA* goal of universal proficiency by 2014. This means that substantial student performance improvements that do not reach the 2014-driven targets are not recognized by the GMPP growth models. However, growth models could be adapted to other targets with the result that more schools would be identified as making AYP than is currently the case.

The generally low rates of making AYP by growth also reflect the impact of the various other (non-growth) methods for determining AYP available in those states for schools to make AYP by status and safe-harbor (e.g., confidence intervals and multiyear averaging), such that the status and safe-harbor methods picked up schools which would have made AYP by growth had those various provisions not been available.

Reasons for the variation among states in the percentages of schools making AYP by growth are not examined systematically in this report but are likely to be some combination of differences in the assessments and proficiency cutpoints used, the features of the growth models implemented, the states' AMOs, and the actual levels of growth realized by the students. The final report will explore these factors in more detail.

An implication for future policy is that states could clarify each school's progress by applying growth criteria to all their schools and groups before status and safe-harbor. The main advantage of applying the growth model before the status and safe-harbor models is that it would identify schools that are realizing adequate progress toward universal proficiency. This would clearly distinguish those schools from schools making AYP under status criteria but not realizing growth

sufficient to continue meeting their AMOs. The latter are probably not a large number but identifying such schools would serve as an early warning mechanism of possible problems. The exploratory analyses in this report also indicate that applying growth criteria before safe-harbor could usefully reclassify most of the current safe-harbor schools as making AYP by growth, and would clearly identify those that are not on-track to proficiency and thus headed for improvement status in the near future.

Another reporting option is to classify each school in terms of both growth and status. Schools making AYP would be distinguished as making AYP by both growth and status, by growth only, by status only, by a mix of status and growth, or by safe-harbor only. This would have the advantage of uniquely identifying different sets of schools (those making AYP in terms of both growth and status).

I. Introduction

A key goal of the *Elementary and Secondary Education Act* (*ESEA*), as amended in 2001, is that all students attain proficiency in reading and mathematics by the year 2014. *ESEA* requires states to: (a) develop grade-level specific proficiency standards in both subjects for grades 3–8 and one or more high school grades, and (b) to assess the performance of all students in those grades each year. The states' proficiency standards progressively increase with all students expected to be proficient by 2014. Based on their students' scores, every public school is evaluated to determine whether or not it is making adequate yearly progress (AYP), and consequences are applied to schools not making AYP for more than two consecutive years.

The U.S. Department of Education initiated the Growth Model Pilot Project (GMPP) in November 2005 with the goal of approving up to ten states to incorporate growth models in school AYP determinations under *ESEA*. Growth models are defined as complements or alternatives to the standard status model for determining school AYP. The status model bases AYP on the proportion of a school's students attaining proficiency in reading and mathematics in a given year. Growth models, in contrast, base AYP in part on some measure of how much additional proficiency students gain in these subjects from one year to the next relative to the target of 100 percent proficiency by the year 2014. Growth models promise to provide a fuller understanding of school effectiveness and the progress each school's students are making toward their proficiency goals. The main objectives of the GMPP are to help states develop and implement models for determining school-level AYP that incorporate measures of student growth.

This interim evaluation of the GMPP is restricted to the eight states approved for participation in the pilot project as of September 2006.[2] It is designed to answer three questions:

1) How have states in the pilot project implemented growth models?
2) How does each pilot state's growth model affect the number and kinds of schools that make AYP?
3) What are the implications of the pilot project experience for extending and strengthening growth models within the context of *ESEA*?

[2] The states approved for participation in the pilot project as of September 2006 were: North Carolina and Tennessee (approved for implementation beginning in the 2005–06 school year); and Alaska, Arizona, Arkansas, Delaware, Florida, and Iowa (approved for implementation beginning in the 2006–07 school year). Ohio was approved to begin implementation in the 2007–08 school year and will be discussed in the final report. In June 2008 and January 2009 Secretary Spellings announced the approval of six additional pilot growth models, in Michigan and Missouri (approved for use in the 2008–09 school year), and Colorado, Minnesota, Pennsylvania, and Texas (beginning in the 2008–09 school year). While the GMPP began in 2005 with a goal of approving up to 10 states, under current Title I regulations, there is no limit now on the number of states that can be approved. See "U.S. Secretary of Education Margaret Spellings Approves Additional Growth Model Pilots" retrieved June 2008 and January 2009 from http://www.ed.gov/news/pressreleases/2008/06/06102008.html and http://www.ed.gov/news/pressreleases/2009/01/01082009a.html.

The remainder of this chapter considers how the GMPP models compare with status models in approaches to evaluating student achievement. Chapters II and III consider, for each of eight pilot grantee states in the 2006–07 school year, the impact of the state's GMPP model on its AYP determinations. Chapter IV addresses a number of hypothetical questions about how results might change if the data collected as part of the pilot project were used differently. A final project report is planned to address the same research questions with data collected for the 2007–08 school year for all eight states plus Ohio, as well as additional questions about the impact of technical features of the models which are not addressed in the current report.

The status model of accountability under *ESEA*

Under *ESEA*, each state develops a standards-based system of student achievement measures and targets, and conducts annual assessments to determine whether its schools and local education agencies (LEAs) are making AYP.[3] Each school has a certain percentage of students who score proficient or higher each year on the mathematics and reading or language arts achievement tests and this constitutes an annual measure of a school's performance. That percentage is expected to reach 100 percent by the end of the 2013–14 school year in incremental steps. In addition, *ESEA* requires each school to meet or exceed statewide standards on one or more "other academic indicators," typically defined in terms of average daily attendance for elementary schools and graduation rate for high schools.

Each step in the path to achieving universal proficiency in reading and mathematics is known as the "annual measurable objective" or "AMO".[4] The AMO is the standard that schools and districts use to determine whether or not they are making adequate yearly progress. AMO's do not increase uniformly across states, including the states in this study. Some increase in a consistent linear fashion toward 2014, while others increase more in the years closer to 2014 than in those closer to the 2002 start of the AYP requirements.

In order for the school to make AYP under the *ESEA* status model, several conditions must be met. These conditions are required by the law and are intended to improve the reliability and validity of the accountability results. First, the school must test at least 95 percent of its students in each of the *ESEA* reporting groups in both reading or language arts and mathematics. The *ESEA* reporting groups consist of all students plus the major racial and ethnic subgroups, students with disabilities, limited English proficient students, and students from low-income households. Within each school, a reporting group may be excluded from federal accountability requirements if the number of "full academic year" students from that group is below a minimum "n" size. Most states define a full academic year as starting in the fall when enrollments are finalized (typically around Oct. 1) and extending through the end of the testing window in the

[3] See "What is adequate yearly progress (AYP)?" (retrieved May 2008 from http://answers.ed.gov/cgi-bin/education.cfg/php/enduser/std_adp.php?p_faqid=6&p_created=1095256734&p_sid=biCz1Gli&p_lva=&p_sp=cF9zcmNoPSZwX3NvcnRfYnk9JnBfZ3JpZHNvcnQ9JnBfcm93X2NudD0xMTUmcF9wcm9kcz0mcF9jYXRzPSZwX3B2PSZwX2N2PSZwX3BhZ2U9MQ**&p_li=&p_topview=1); U.S. Department of Education, Office of the Secretary, Office of Public Affairs. (2003). *No Child Left Behind: A Parents Guide*, Washington, D.C. (retrieved May 2008 from http://www.ed.gov/parents/academic/involve/nclbguide/parentsguide.pdf).

[4] AMOs are defined by the states and have different targets for each school year, and can be different within each year across grades and subjects. In any case, states were required to follow strict federal guidelines in setting the AMOs.

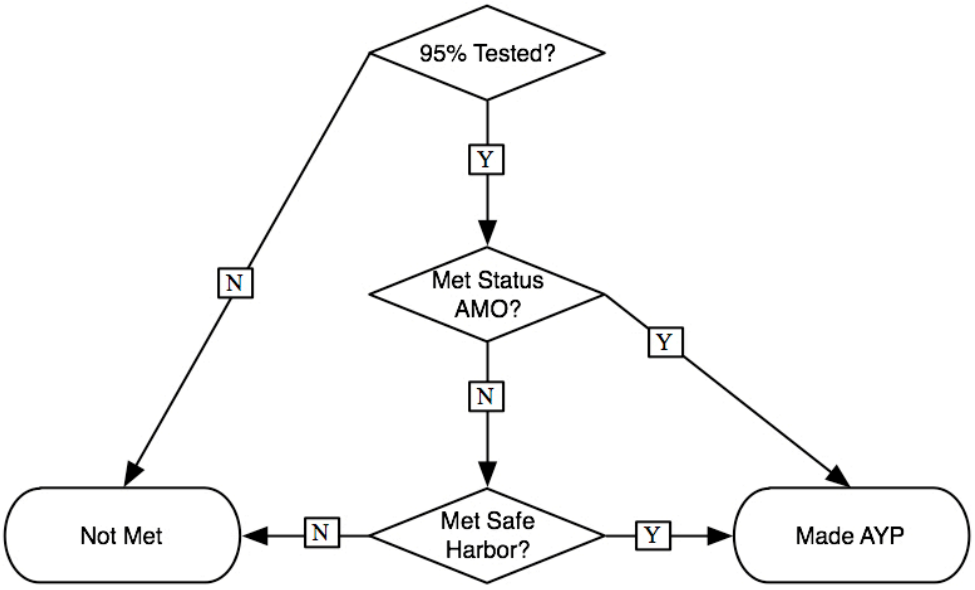
spring, while the state-defined minimum n sizes for reporting groups range from a low of no minimum to 50 students. Second, the percentage of tested students scoring proficient or higher must meet or exceed the AMO in both subjects for eligible reporting groups. These percentages are calculated only for students enrolled for the full academic year. If a single subgroup fails to achieve the AMO, the school does not make AYP.

In order to reduce the chances of incorrectly classifying schools as not making AYP, the states are allowed to apply additional steps within the status model. If any one or more of the *ESEA* reporting groups did not make AYP, the school may:

- Apply confidence intervals to the group's percent proficient and compare the upper bound to the AMO. Analogous to the sampling errors typically reported on results from political and other opinion polls based on random samples, a confidence interval represents the range of values within which the true value is expected to fall at a given level of statistical certainty. The higher the standard of certainty asked for, the wider the confidence interval. For a given standard of certainty, the smaller the number of students, the wider the confidence interval. If the higher limit of that confidence interval is greater than the AMO, then the subgroup is considered to make AYP.
- Average the test results for the group over two or three years and compare the average to the AMO (this is often referred to as "multiyear averaging").
- Apply safe-harbor, whereby the group makes AYP if the percentage of non-proficient students in the group decreased by 10 percent or more from the prior year.
- Apply safe-harbor but assess whether the reduction in the percentage non-proficient was 10 percent or more from the average percent non-proficient over the prior two or three years.

This basic method of determining school AYP is used in all states and is referred to as the status model. A simplified version of the decision tree (excluding the full academic year, minimum n, confidence intervals, and multiyear averaging conditions) is illustrated in Exhibit 1.

Exhibit 1
Determining AYP Under Status Model



Limitations of the status model

One characteristic of the status model is that it does not recognize real improvements in student achievement unless they result in higher percentages of students meeting or exceeding proficiency standards in a given year. Schools whose students are demonstrating learning gains, but whose gains overall fall short of the AMO standard will not be judged to be making adequate yearly progress. Conversely, schools whose students meet the AMO standard will be judged to be making AYP even if little is done to advance student achievement beyond that minimum proficiency target.

The fact that schools either make AYP or not, with no credit given for improvement other than safe-harbor, has raised concerns that, in the short term, instructional resources may be focused on students who are closest to the proficiency threshold. Students less likely to attain proficiency from a given amount of instructional effort—e.g., those farthest below the proficiency threshold—may receive less attention. There is not a substantial amount of empirical research supporting this hypothesis and, in any case, this type of strategy becomes less relevant as AMOs converge on the target of universal proficiency by 2014. On the other side of the ledger, students unlikely to fall below proficiency may also receive less attention because there are no statutory consequences for failing to improve achievement among proficient students. This possible

tendency is tied to the use of a minimum threshold and would thus not be affected by the 2014 target.[5]

Another characteristic of the status model is that it does not take account of changes in a school's student composition from one year to the next. Thus a school classified as not making AYP in one year could be judged to make AYP in the next if more proficient students enrolled or if less-proficient students left. Conversely, a school making AYP one year may not reach the AMO standard the following year if student composition shifted the other way.

The Growth Model Pilot Project (GMPP)

The U.S. Department of Education (ED) initiated the Growth Model Pilot Project (GMPP) in November 2005 with the goal of approving up to 10 states to incorporate growth models in school AYP determinations under *ESEA*. Growth models are defined as complements or alternatives to the standard status model for determining school AYP; they base AYP on some measure of how much students have gained from one year to the next.

Growth models are intended to recognize schools' progress moving students toward proficiency. As suggested above, school-level student growth patterns may overlap or may diverge from the school-level assessments of students' proficiency generated by the application of the status model; these are illustrated in Exhibit 2. Ideally, all schools would be in cell A, meeting the status requirements and realizing high rates of annual progress. The basic goal of the GMPP is to identify schools in cell B, that is, schools with high numbers of students making progress but not yet attaining the grade-level proficiency thresholds necessary to meet AYP standards. Schools in cell C—those with low rates of progress but still making AYP—are also of interest to the GMPP but, as will be discussed further at various points in this report, were not targeted in the project. For the *ESEA* as a whole, the overarching goal is to ensure that no schools are in cell D, not meeting the status model requirements and not making sufficient gains to be on-track to meet the requirements in the near future.

Exhibit 2
Conceptual Map of How Growth Patterns Compare With Status Model AYP Designations

Achievement Growth Pattern in the School	School AYP Designation Under the Status Model	
	Made AYP	Did Not Make AYP
High rates of progress	A	B
Low rates of progress	C	D

[5] Research on the extent to which schools have adopted these sorts of strategies has been conducted by Naomi Chudowsky, Victor Chudowsky, and Nancy Kober (2009) "Is the Emphasis on "Proficiency" Shortchanging Higher- and Lower-Achieving Students? Retrieved on July 30, 2009 from http://www.cep-dc.org/index.cfm?fuseaction=document_ext.showDocumentByID&nodeID=1&DocumentID=280 ; Derek Neal and Diane Whitmore Schantzenbach (2008), "Left Behind By Design: Proficiency Counts and Test-Based Accountability" retrieved on September 26, 2008 from http://derek.a.neal.googlepages.com/restat_rev_final_200809.pdf ; and Jennifer Booher-Jennings (2005), "Below the Bubble: 'Educational Triage' and the Texas Accountability System," *American Educational Research Journal*, 42: 231–268.

The U.S. Department of Education used a rigorous peer review process to evaluate the adequacy of the technical aspects of the proposed models and to ensure that the models aligned with seven core principles.[6] In general, these core principles required all pilot states to set annual "growth targets" for ensuring proficiency by 2014 and to track individual students across schools and measure their progress across grades in both reading and mathematics. The first principle requires that the growth model, like the status model, be applied to each targeted subgroup as well as all students in the school. This means that growth outcomes are to be monitored separately, or "disaggregated," for the main racial and ethnic groups, limited English proficient (LEP) students, special education students, and low-income students.

The second principle stipulates that growth expectations cannot be based on student background or school characteristics. This is consistent with the *ESEA* rule that proficiency targets must be the same for all students in a given grade and cannot be modified for different kinds of students or schools. In the growth model context, this excludes use of models that would give AYP credit to schools or subgroups within schools strictly on the basis of realizing average or even higher-than-average annual growth rates. The key criterion is "meeting grade-level proficiency." Students scoring below proficiency must not only gain more per year than one grade-level equivalency, but those gains must also point to attaining proficiency standards within a specified time frame.[7] A full list of the seven core principles of the Growth Model Pilot Project is provided in Exhibit 3.

[6] See U.S. Department of Education (Nov. 18, 2005) "Press Release: Secretary Spellings Announces Growth Model Pilot, Addresses Chief State School Officers' Annual Policy Forum in Richmond" (retrieved May 2008 from http://www.ed.gov/news/pressreleases/2005/11/11182005.html), and U.S. Department of Education (July 2007) "No Child Left Behind: Growth Models—Ensuring Grade-Level Proficiency for All Students by 2014" (retrieved May 2008 from http://www.ed.gov/admins/lead/account/growthmodel/proficiency.pdf).

[7] This guiding principle precludes the use of growth models referred to as "value added" models (VAM), at least when they include statistical adjustments for student social background variables or school characteristics. In general terms, value-added formulations seek to separate the impacts of various factors on student growth in order to make attributions about the marginal effectiveness of the factors on achievement gains. Inputs are variously defined, depending on the purpose of the analysis, but often include student social background variables, teachers, and instructional programs. If used for evaluation of schools, value-added models estimate growth for a common "average type" of student in each school and the schools are then compared in terms of that background-standardized estimate. In that sense, VAM are typically measures of *relative school effectiveness*, not absolute student proficiency. The latter is the purpose of GMPP, and VAM are generally inappropriate for that purpose.

Exhibit 3
Seven Core Principles of the Growth Model Pilot Project

States approved for participation in the GMPP were required to meet seven core principles in the *ESEA* accountability plans they submitted for incorporating growth models in their AYP measurements:

1. Ensure that all students are proficient by 2014, and set annual goals to ensure that the achievement gap is closing for all groups of students;
2. Set expectations for annual achievement based on meeting grade-level proficiency, not on student background or school characteristics;
3. Hold schools accountable for student achievement in reading or language arts and mathematics;
4. Ensure that all students in tested grades are included in the assessment and accountability system, hold schools and districts accountable for the performance of each student subgroup, and include all schools and districts;
5. Include assessments in each of grades 3–8 and in high school for both reading or language arts and mathematics, and ensure that they have been operational for more than one year and receive approval through the *NCLB* peer review process. The assessment system must also produce comparable results from grade to grade and year to year;
6. Track student progress as part of the state data system; and
7. Include student participation rates and student achievement on a separate academic indicator in the state accountability system.

Source: See "Peer Review Guidance for the *NCLB* Growth Model Pilot Applications" (retrieved May 2008 from http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf).

Other significant features of pilot growth models were allowed to vary, as long as the technical specifications passed review by a panel of nationally recognized experts.[8] Reviewers had a series of meetings to discuss the Peer Review Guidance document and the Department's expectations for the process. Proposals that passed an initial review by Department staff and a round of clarification by states were forwarded to the peer reviewers. The Department then set up conference calls between the states and the peer review team, after which the reviewers met again to discuss the proposals and make recommendations for use by the secretary in deciding which proposals to approve for the pilot project.

Of the 20 states that submitted proposals by February 2006, 13 asked for approval to use growth models for immediate use in the 2005–06 school year while the others proposed to start in 2006–07. Eight of these proposals passed the initial evaluation, and revised proposals were forwarded to the peer reviewers.[9] Two states, North Carolina and Tennessee, received approval to use growth models for 2005–06, while Delaware, Arkansas, and Florida revised their

[8] See "Peer Review Guidance for the *NCLB* Growth Model Pilot Applications" (retrieved May 2008 from http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf).

[9] These reviewers included Eric Hanushek, Stanford University, Chris Schatschneider, Florida State University, David Francis, University of Houston, Margaret Goertz, University of Pennsylvania, Kati Haycock, The Education Trust, William Taylor, Citizens Commission on Civil Rights, Sharon Lewis, Council of Great City Schools (retired), Robert Mendro, Dallas Independent School District, Jeff Nellhaus, Massachusetts Department of Education, and Mitchell Chester, then at the Ohio Department of Education.

proposals and were approved in a subsequent peer review to implement their proposed models for the following school year (2006–07). A second round of proposals was solicited in October 2006. Six previous applicants submitted revised proposals along with submissions from three new states.[10] Of these, Alaska, Arizona, and Iowa received immediate or conditional approval to use their growth models beginning with the 2006–07 school year, while Ohio was approved to begin in the 2007–08 school year.[11]

The approved pilot states were all required to calculate whether each student in the target grade levels was on-track to be proficient within a specified number of years or by a particular grade level. Students who were on-track to be proficient could be counted as proficient for purposes of AYP determinations. However, the states were given a variety of options on how to incorporate the student growth indicator in their AYP determinations.[12] These included using:

- only the growth measure to calculate the percentage on-track to proficiency for AMO assessment;
- both the status and the growth measures to calculate the percentage proficient or on-track for AMO assessment;
- the status measure to calculate the percentage proficient, applying safe-harbor provisions if needed, and using the growth measure either in conjunction with the status measure or alone if AYP not met with safe-harbor; and
- safe-harbor provisions and the growth measure for AMO assessment.

While the approved models differ from one another in a number of important ways, all use state-specific assessment data to measure student progress and proficiency, and the method of incorporating growth outcomes in AYP determinations was generally the same. Seven of the eight pilot states proposed to apply growth criteria only *after* schools failed to make AYP under the status and safe-harbor provisions rather than determining AYP solely on the basis of student improvement. More specifically, the designs of the pilots in these seven states applied growth criteria only to those students who were members of *ESEA* reporting groups that did not reach their AMOs or attain AYP via safe-harbor provisions. This basic method of augmenting status

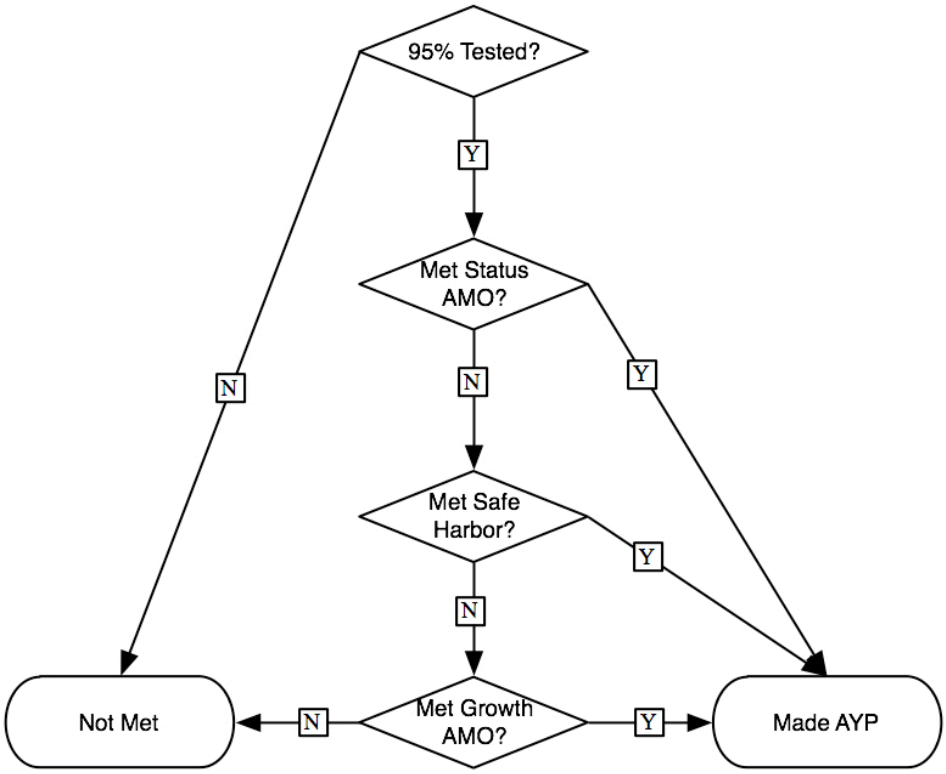
[10] See "Secretary Spellings Approves Additional Growth Model Pilots for 2006–2007" retrieved June 2008 from http://www.ed.gov/news/pressreleases/2006/11/11092006a.html). This panel of reviewers included Anthony Bryk, Stanford University, Harold Doran, American Institutes for Research, Chrys Dougherty, National Center for Educational Accountability, Lou Fabrizio, North Carolina Department of Public Instruction, Tom Fisher, Independent Consultant, Pete Goldschmidt, University of California at Los Angeles, Sharon Lewis, Council of Great City Schools (retired), Margaret McLaughlin, University of Maryland, Robert Mendro, Dallas Independent School District, Jeff Nellhaus, Massachusetts Department of Education, Ann O'Connell, University of Connecticut, Dianne Piché, Citizens Commission on Civil Rights, Sandy Sanford, Riverside County Office of Education, Chris Schatschneider, Florida State University, William Taylor, Citizens Commission on Civil Rights, and Martha Thurlow, University of Minnesota.

[11] In December 2007, the Department moved to expand participation in the growth model project and extended an invitation to all states to apply for inclusion. At the time of writing, 15 states have been approved to implement growth models in AYP determinations.

[12] See "Peer Review Guidance for the *NCLB* Growth Model Pilot Applications" (retrieved May 2008 from http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf)

with growth results is shown in Exhibit 4 (again, this is a simplification in that provisions for using confidence intervals and multiyear averaging prior to applying growth results are not represented).

Exhibit 4
Determining AYP Under Status, Safe-Harbor, and Growth



Delaware adopted a different order, applying growth first and then applying status and safe-harbor only to schools and subgroups that did not meet AYP under the growth criteria. This procedure thus classified schools as making AYP by growth even if they also would have made AYP by status or safe-harbor criteria. In contrast, schools in the other states were identified as making AYP by growth only in cases in which AYP was not met by status or safe-harbor.

Types of growth models implemented in the pilot

The states approved for the pilot study proposed to employ either a status-augmented-with-growth or (in Delaware) a growth-augmented-with-status method of determining AYP, but the models used by the pilot states varied in how they established growth targets that define whether individual students were "on-track" to reach proficiency in the allotted time frame. We have identified three basic types of growth models being used in the GMPP: the *transition matrix* model, the *trajectory* model, and the *projection* model.[13]

[13] See the CCSSO's *Implementer's Guide to Growth Models* for an alternative, more extensive, typology of growth models (retrieved May 2008 from http://www.ccsso.org/content/pdfs/IGG%20Final%20AP.pdf).

Transition Matrix. This type of model evaluates student progress from year to year in terms of a relatively small set of discrete performance levels. The levels are defined in general terms that are applied to all grades (e.g., below proficient, proficient, advanced). Student growth is indexed by movement ("transitions") from lower to higher categories. Delaware and Iowa used models of this type.

An illustration of the transition matrix model used in Iowa is shown in Exhibit 5. Students who scored below the proficiency threshold in year 1 (the first three rows) were classified as "on-track" to proficiency if their test performances improved enough in year 2 to move up at least one level. All students who were proficient or advanced in year 1 and who were proficient or advanced in year 2 were classified as on-track. For the purpose of determining AYP, the Iowa model counted students who were on-track toward proficiency as being fully proficient. All students in the gray-shaded cells are not on-track and do not count as proficient for AYP purposes.

Exhibit 5
Illustration of How the Iowa Transition Matrix Model Classifies Students

Year 1 Performance Level	Year 2 Performance Level				
	Weak	Low Marginal	High Marginal	Proficient	Advanced
Weak	Off-track	On-track	On-track	On-track	On-track
Low marginal	Off-track	Off-track	On-track	On-track	On-track
High marginal	Off-track	Off-track	Off-track	On-track	On-track
Proficient	Off-track	Off-track	Off-track	On-track	On-track
Advanced	Off-track	Off-track	Off-track	On-track	On-track

Delaware used a somewhat different type of transition matrix model; it is illustrated in Exhibit 6. It used a point system that gave a maximum of 300 points to all students who attained proficiency in year 2 but gave only partial credit to students who scored below proficiency in year 1 and who made gains but did not reach the proficiency standard in year 2. Students who scored below the proficiency threshold in year 1 and did not move up to a sufficiently higher level in year 2 (the gray-shaded cells) were assigned no points for determining AYP. Points were summed within subgroups and divided by the number of students in each subgroup to calculate an "average growth value" which was compared to annual growth targets.

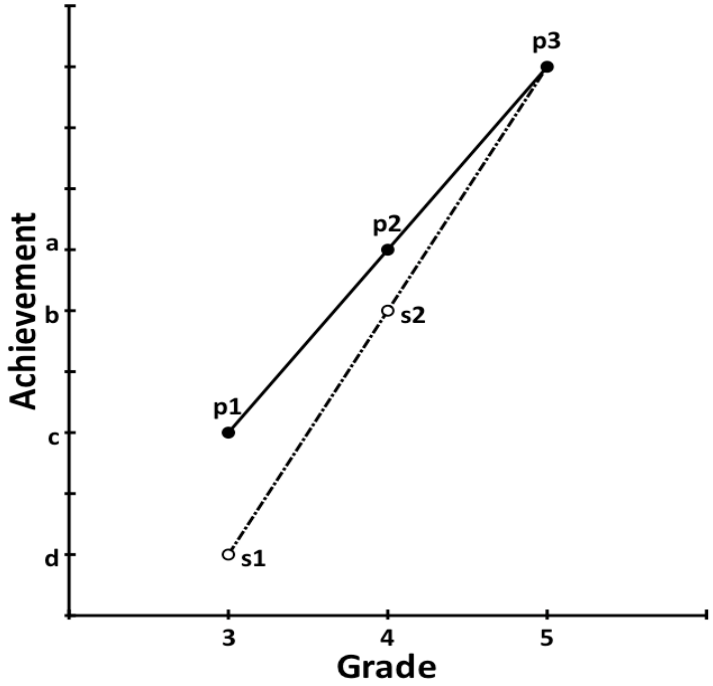
Exhibit 6
Illustration of How the Delaware Transition Matrix Model Assigns Points to Students' Gains

Year 1 Performance Level	Year 2 Performance Level				
	PL 1A	PL 1B	PL 2A	PL 2B	PL 3, 4, and 5
PL 1A: lowest level below proficiency	0	150	225	250	300
PL 1B	0	0	175	225	300
PL 2A	0	0	0	200	300
PL 2B: highest level below proficiency	0	0	0	0	300
PL 3, 4, and 5: Proficient or higher	0	0	0	0	300

Trajectory. This type of model uses the gap between a student's baseline test score and a performance standard several years out to calculate the amount of growth he or she must attain to become proficient. This performance "trajectory" is divided up into annual growth targets that, taken together, put the student on-track to proficiency within the allotted years. Alaska, Arizona, Arkansas, Florida, and North Carolina used trajectory models.

An example of a trajectory model is shown in Exhibit 7. The Y axis represents student achievement on a vertically aligned scale, and the X axis represents successive grades. The large solid dots indicate the level of achievement necessary to be considered proficient at each time point ("p1" for proficiency at time 1, "p2" for proficiency at time 2, and "p3" for proficiency at time 3). The hollow dots represent a student's actual achievement at time points 1 and 2, marked "s1" and "s2" respectively.

Exhibit 7
Illustration of How Trajectory Models Set Targets for Students Scoring Below Proficiency Thresholds



At grade 3, the student is not considered proficient because his or her achievement (at level "d" on the Y axis) is lower than what is necessary at that time (level "c"). Under a trajectory model, each student has a growth target that must be met to be considered "on-track" and count as proficient. In the illustration above, this is done by drawing a line from the student's achievement at grade 3 ("s1") to the proficiency point for grade 5 ("p3"). This line intersects point "s2" at grade 4, indicating that "s2" is what the student must achieve in grade 4 to be considered "on-track" to proficiency. The level of achievement targeted by the trajectory model is lower than what is expected under the status model. Under the status model, a student must have a level of achievement at or above point "p2".

This illustration shows that trajectory models allow students to score below the proficiency thresholds between the year they first miss proficiency ("s1") and the year when they must be proficient under status criteria ("p3"). Because expected growth is based on a trajectory from the student's initial achievement and the target year, trajectory models distribute the growth required to meet status expectations among the intervening years. Finally, while this illustration used a linear trajectory, some states, such as Arkansas, use a nonlinear method to calculate growth targets.

Projection. This type of model uses current and past test scores to statistically predict performance several years ahead based on how all students in the state or school with similar patterns of scores generally perform. Such "projections" utilize multiple test scores to estimate a prediction equation using multiple regression techniques and use that equation to predict how each student will score at the end of the time frame. If that predicted score is equal to or greater

than the proficiency cut point, the student is classified as on-track to proficiency. Ohio (starting in 2007–08) and Tennessee are using projection models.

Exhibit 8 illustrates a simple projection model. The projection model determines a projected level of achievement for a future time point for each student. Depending on the specifics of a state's model, students who do not make the proficiency threshold under a status model at grade 5 may still be counted as proficient if the projection model predicts that they will make proficiency by grade 6. Like the trajectory model illustration, the Y axis represents student achievement on some vertical scale. The X axis represents grades. Again, the solid dots represent grade-level proficiency thresholds at grades 3 (p1) through 6 (p4), and the hollow dots represent a student's achievement over consecutive grades. The solid line represents the projection equation based on the current and past data for that student (points "s1" through "s3"). Though illustrated here with data for just one student, both the Ohio and Tennessee models estimate this projection equation from a regression analysis of current and past data for all students in the school so that the projection represents statistically expected scores for students similar to the focal student. The dotted line is the extension of that projection equation to a future time point specified by the state's growth model.

Under a status model, a student who reached the level of "s3" would not be considered proficient at grade 5, because her score is lower than that of point "p3". However, under a growth model using projections based on student data, that student would be considered "on-track" because her projected score qualifies as proficient at grade 6 (the dotted line is above "p4" in the figure).

Exhibit 8
Illustration of How Projection Models Evaluate Students Scoring Below Proficiency Thresholds

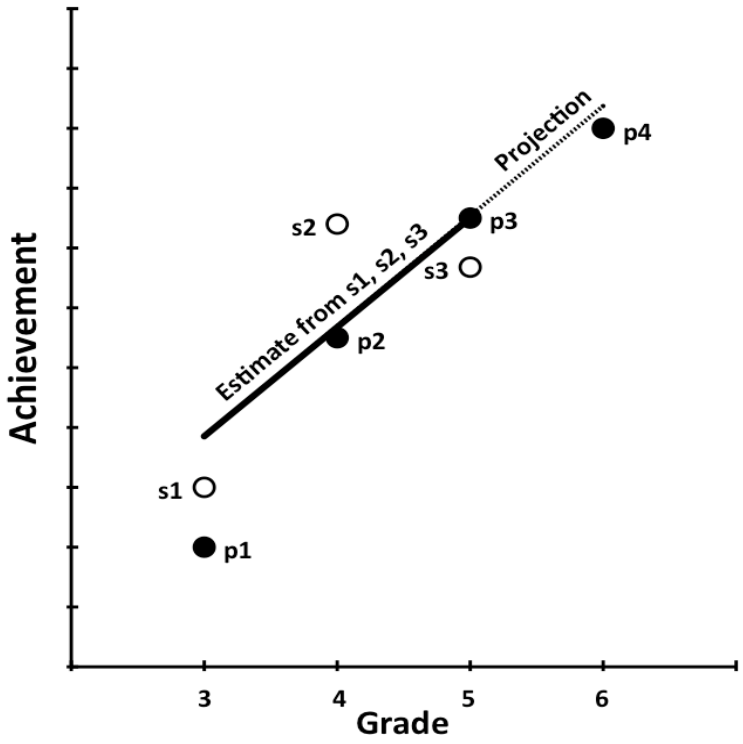


Exhibit 9 presents each pilot state's growth model organized by type of model used and provides brief model summaries along dimensions that may affect final AYP outcomes. These include which grades receive growth calculations, how many years of growth are allowed, what tests are used to measure achievement, the standards applied for determining adequate yearly growth, and the growth required for students to be recognized as "on-track-to-proficiency" and therefore counted positively for school AYP determinations. More detailed information about the technical features of each model can be found in the growth model summaries appended to this report (Appendix A).

Exhibit 9
Overview of Growth Models Approved for the GMPP in the 2006–07 School Year

State	Type of Model	First Year Used	Grades Included	Minimum Subgroup Size	Years to Proficient	Achievement Measures Used	Growth Standards Applied	Student On-Track-to-Proficiency Determined
Delaware	*Transition Matrix*	2006–07	3–10	40	N/A	Scale scores for reading and math via the Delaware Student Testing Program	Four levels of "below the standard" are used to categorize non-proficient students	Points are awarded for proficiency and for movement to higher levels of below proficient and are then averaged to make AYP determinations
Iowa	*Transition Matrix*	2006–07	4–8	30	4	Scale scores based on biannual Iowa Test of Basic Skills (ITBS) math and reading results	Three categories of performance are used to classify non-proficient students	Students who move to a higher category of non-proficiency make Adequate Yearly Growth (AYG) and are counted as proficient for AYP determinations
Alaska	*Trajectory*	2006–07	4–9	25	4	Adjusted scale scores for math and language arts (reading + writing) on the Standards Based Assessment (SBA) Test	Annual growth targets for each student are based on the test score gains needed to reach proficiency in four years or by tenth grade. Each student's targets are reset every year to reflect his or her latest score.	Students who close the score gap by one-fourth, one-third, and one-half annually (depending on the distance from target year) are "on-track" to reach proficiency and are counted as proficient for AYP determinations
Arizona	*Trajectory*	2006–07	4–7	40	3	Scale scores for math and reading on the Arizona Instrument to Measure Standards (AIMS) tests	Growth targets are set using the increment in base year score needed to reach proficiency in three years or by eighth grade	Students who test "below the standard" but are estimated to meet their growth targets are counted as proficient for AYP determinations

Exhibit 9, continues next page

Exhibit 9
Overview of Growth Models Approved for the GMPP in the 2006–07 School Year
continued from previous page

State	Type of Model	First Year Used	Grades Included	Minimum Subgroup Size	Years to Proficient	Achievement Measures Used	Growth Standards Applied	Student On-Track-to-Proficiency Determined
Arkansas	*Trajectory*	2006–07	4–7	40	4	Scale scores for literacy and math on the Arkansas Benchmark Exams	Growth targets are set using the annual increment in current test score needed to reach proficiency by eighth grade	Non-proficient students meeting their growth targets are counted as proficient for AYP determinations
Florida	*Trajectory*	2006–07	3–10	30	3	Developmental Scale Scores (DSS) for math and reading on the Florida Comprehensive Assessment Test (FCAT)	Growth targets are based on the gap between initial score and proficiency cutoffs three years later	Students who close the original gap by one-third (one-half for ninth-graders) annually are counted as proficient for AYP determinations
North Carolina	*Trajectory*	2005–06	3–7	40	4	Scale scores on third grade pretest and/or annual North Carolina End-of-Grade Math and Reading Tests	Growth targets are based on closing the gap between baseline test score and proficiency cutoff in four years or by eighth grade	Students who close the test score gap by one-fourth each year meet growth targets and are counted as proficient for AYP determinations
Tennessee	*Projection*	2005–06	4–8	45	4	Scale scores for math and reading/ language arts from the Tennessee Comprehensive Assessment Program (TCAP)	Proficiency standards for each grade are based on cutoffs for proficiency three years later using test results on the TCAP starting in 2002–03	Students with predicted scores above the cutoff, using regression methods to calculate predicted scores, are counted as proficient for AYP determinations

How growth models are used for AYP determinations

The pilot states used the student-level indicators of on-track-to-proficiency generated by the GMPP growth models to augment the status indicators of proficiency in order to make AYP determinations for their schools. An overview of how the growth models' on-track-to-proficiency data fit into the overall scheme in most of the eight pilot states is shown in Exhibit 10. This exhibit resembles a checklist, showing all the criteria that must be met in order to make AYP and (at the bottom of the table) how schools are classified in terms of making AYP.

The key point is that when a school is designated as "making AYP under growth," it means that use of the growth model changed the designation for one or more targeted subgroups. Further, within the affected subgroups for a given school, the growth criterion is usually applied only to the students who did not achieve at or above the proficiency level. Simply stated, "making AYP under growth" does not mean that all students are on-track to proficiency, and it can mean as few as one non-proficient student is on-track if a sufficient number of others in the subgroup are proficient.

Seven of the eight states with data available for 2006–07 used the status-and-safe-harbor-augmented-with-growth methodology shown in Exhibit 10. In contrast, Delaware used a growth-augmented-with-status-and-safe-harbor method for determining AYP, and many of the schools listed as making AYP under growth would also have made AYP under status. As discussed more fully in Chapter II, the particular type of growth model used in Delaware actually functioned similarly to the other states' models in terms of classifying students as proficient, on-track to proficiency, and not on-track.

Exhibit 10
Overview of How Schools Are Classified as Making AYP Under the Status Augmented With Growth System

ESEA Outcomes	Exempt by Minimum N Criterion?	Met AMO by Status? (Applies Only if "Exempt" is No)	Met AYP by Safe-Harbor? (Applies Only if "Met by Status" is No)	Met AMO by Growth? (Applies Only if "Met by Safe-Harbor" is No)
Reading or Language Arts				
All students	Not Applicable	Y/N	Y/N	Y/N
Asian	Y/N	Y/N	Y/N	Y/N
American Indian	Y/N	Y/N	Y/N	Y/N
Black	Y/N	Y/N	Y/N	Y/N
Hispanic	Y/N	Y/N	Y/N	Y/N
White	Y/N	Y/N	Y/N	Y/N
LEP	Y/N	Y/N	Y/N	Y/N
Low income	Y/N	Y/N	Y/N	Y/N
Students w/ disabilities	Y/N	Y/N	Y/N	Y/N
Mathematics				
All students	Not Applicable	Y/N	Y/N	Y/N
Asian	Y/N	Y/N	Y/N	Y/N
American Indian	Y/N	Y/N	Y/N	Y/N
Black	Y/N	Y/N	Y/N	Y/N
Hispanic	Y/N	Y/N	Y/N	Y/N
White	Y/N	Y/N	Y/N	Y/N
LEP	Y/N	Y/N	Y/N	Y/N
Low income	Y/N	Y/N	Y/N	Y/N
Students w/ disabilities	Y/N	Y/N	Y/N	Y/N
Other Academic Indicator (e.g., Average Daily Attendance)	Not Applicable	Y/N	Not Applicable	Not Applicable
School AYP Classification	School still receives AYP classification from status, safe-harbor, or growth based on the "all students" group even if all subgroups are exempt.	"By status" if all except exempt groups are Yes	"By safe-harbor" if one or more are Yes and the rest are Yes by status or exempt	"By growth" if one or more are Yes and the rest are Yes either by safe-harbor or status or exempt

Data sources and availability

Data employed for the analyses in this report were provided by the U.S. Department of Education and the pilot grantee states. The data from the Department were extracted from the ED*Facts* database. ED*Facts* is the main repository for school, district, and state data on *ESEA* requirements related to making AYP determinations, as well as enrollment and demographic data. All states are required to submit standard data elements on all their districts and K-12 schools each year to ED*Facts*. For the 2006–07 school-year data that are the focus of this report, the standard reporting variable on whether the school made AYP was modified to collect information on whether each school made AYP because of the growth model. This variable was defined in ED*Facts* with a set of mutually exclusive categories: "made AYP by regular determination," "made AYP by growth," or "did not make AYP."

The "made AYP by regular determination" category included all methods of making AYP except by growth criteria. The main methods included regular status criteria, status with confidence intervals, status with multiyear averaging, and safe-harbor. Use of confidence intervals and multiyear averaging were not reported in ED*Facts*, but safe-harbor was reported with respect to subgroup AMO results for reading and mathematics. The standard ED*Facts* reporting variable for AMO results included the mutually exclusive categories of "met by status," "met by safe-harbor," "exempt by minimum n," and "did not meet" for each subject area. For purposes of this report, schools were classified as "made AYP by safe-harbor" if one or more subgroups were classified in ED*Facts* as "met by safe-harbor" for either the reading or mathematics AMO.

The eight GMPP states approved for 2006–07 school year were additionally required as a condition for continued participation in the pilot to collect and make available for evaluation purposes data on results of using their approved growth models. The data obtained directly from the pilot states consisted of scale scores and proficiency designations in reading or language arts and mathematics for each student in the grades involved in the GMPP. Of particular importance to this study were the proficiency designations, for these included the indicator of whether the student was "on-track" to achieve proficiency within the time frame specified by the approved growth model. Additional data elements included various background characteristics including school identifier codes, grade level of current enrollment, and *ESEA* subgroup memberships.

Five of the eight states had provided both the ED*Facts* and student-level data required to address the two study questions which are the focus of this interim report. The other three states (Arizona, North Carolina, and Tennessee) had not reported GMPP results to the ED*Facts* archive but provided school-level indicators of which schools were classified as having made AYP by growth in 2006–07.

Calculation of hypothetical growth-only percentages on-track to proficiency

As noted above, states had the options of making AYP determinations using (a) only growth results and (b) growth criteria before status and safe-harbor criteria. While these options were generally not exercised, it is possible to estimate what would happen if they had been. That information may prove useful to states in the process of developing growth models, as well as current pilot states contemplating changes to their models. Toward that end, a number of hypothetical "what if" questions are addressed in Chapter IV using the indicators of whether

students are on-track to proficiency. These data are used to calculate the percentages of on-track students in reading and mathematics for each *ESEA* subgroup in each school, and these percentages are compared to the state's "annual measurable objectives," or AMOs. These results are then used to assess the extent to which schools currently making AYP by status and safe-harbor would be able to reach their AMOs using growth-only percentages of students on-track to proficiency.

Data limitations

The data described above have a number of limitations that are important to note at the outset and which will be reiterated at various points in this report. First, the ED*Facts* data for 2006–07 only reported growth-model results in terms of how each school was classified for AYP reporting. ED*Facts* did not identify which *ESEA* reporting groups (all students and the eight subgroups) made their AMOs for reading or mathematics by growth.

As noted above, both the school AYP and subgroup AMO results were defined in ED*Facts* with a set of mutually exclusive categories. It was not possible to determine from these data, for the schools classified as having made AYP by growth, whether particular subgroups made their AMOs as a result of the GMPP growth provisions. An additional category of "met by growth" was added to subgroup AMO reports beginning in 2007–08 and this should overcome the limitation.

A second limitation is that the mutually exclusive AYP categories reported by ED*Facts* made it impossible to determine the extent to which schools that made AYP under status and safe-harbor would also have made AYP if *only* growth criteria were considered. The student-level on-track-to-proficiency indicators described in the preceding subsection can be used to calculate growth-based AMO determinations for each *ESEA* reporting group and for the school as a whole. However, it should be emphasized that these growth-based AMO determinations are based solely on information regarding students' reading or language arts and mathematics performances and do not necessarily indicate whether a school would make AYP under a growth-only system. Under *ESEA*, schools must also meet additional conditions to make AYP. These conditions include the requirements that the school (a) meet or exceed a minimum level on an "other academic indicator," typically average daily attendance for elementary and middle schools; and (b) realize at least a 95 percent participation rate on the annual assessments. We were unable to consider these criteria in our calculations because the requisite data were not available. (The "other academic indicator data" are school-level determinations that were not provided in the student files and were unevenly reported in ED*Facts*. The student files from some states also did not include the information needed to calculate participation rates.[14]) This means that some

[14] Participation rates are calculated on the basis of students enrolled for the "full academic year" (FAY), but FAY and non-FAY students were not distinguished in some of the state data files. And some states were only able to provide student data on students with test results from both 2005–06 and 2006–07 (i.e., students for whom growth could be calculated). The FAY indicators as well as test score data and results for all current students (not just the ones with prior-year test scores) are being requested for the final report, which will analyze the 2007–08 school year results.

schools meeting the AMO on the achievement outcomes using the growth criteria identified in this report may have officially missed AYP on the additional criteria.[15]

Given the potential discrepancies between the growth-based AMO designations and the official *ESEA* designations, the results presented in Chapter IV of this report should be regarded as suggestive rather than definitive. With that caveat, these comparisons allow an assessment of the extent to which schools making AYP by status or safe-harbor criteria might also have made AYP if only growth criteria were used.

[15] The number of such "false positives" is probably small. National data from 2003–04 show that about 3 percent of all schools did not make AYP solely because of not making acceptable levels on the other academic indicator or low (below 95 percent) participation rates for the achievement testing (U.S. Department of Education, 2007, p. 43.)

II. Effects of Growth Models Implemented Under the GMPP on School AYP Determinations

The data available from the 2006–07 school year allow for a variety of analyses that address the guiding research questions. Before examining the data, it is useful to provide more detail on how the approved growth models in general, and each pilot state's model in particular, fit into the accountability framework of *ESEA*. This chapter provides an overview of how test results (in reading or language arts, mathematics, and other subjects) and other criteria are used to make school AYP determinations, and how student-level data from the growth models fit into those decisions. While each state has a model that is unique in important ways, the approved models generally group into three types. Recognizing key distinguishing features of these three basic approaches is helpful both in clarifying the details of each state's model and understanding the results it produces.

This chapter presents overviews of models implemented in the eight pilot states, focusing on the methods they employ to (1) set expectations, (2) measure growth, and (3) incorporate these growth results into their AYP determinations. These summaries are based on a review of extant documentation, including the approved GMPP proposals, decision letters from the U.S. Department of Education, and correspondence with the Department concerning feedback and suggested revisions from the peer review panels; growth model descriptions found in state accountability workbooks, school report cards, and other online technical documentation; and edits by state officials to draft model summaries provided to each pilot state at a December 2008 summit meeting in Washington, D.C. Any inconsistencies in model summaries were then resolved through e-mail correspondence and follow-up phone calls with state officials over the course of the evaluation. More detailed technical summaries of the states' growth models are included in Appendix B.

Following the overview of each state's model, we present data on the impact of the growth model on AYP outcomes. These analyses address the research question "How many schools made AYP under the growth model that would not have made it under the *ESEA* status model?" This question can be answered for all states except Delaware by analyzing data reported by states to the federal ED*Facts* system plus (in three of the states) additional information from the state departments of education on which schools made AYP by growth; as discussed below in the section focused on the Delaware model, the Delaware Department of Education provided the additional data required to address this question for that state.

State growth models and their effects on schoolwide AYP results

The eight states approved for the GMPP in the 2006–07 school year provided data on the results of their pilot growth models to the ED*Facts* system. It is these eight states, then, which are the subject of this interim report, and we summarize here the key features of these eight pilot grantee states' growth models.

Alaska's growth model

Alaska was formally accepted into the Growth Model Pilot Project on July 3, 2007. Alaska's growth model includes fourth- through ninth-graders only, with all other students expected to be

proficient under the status model criteria. The growth model defines targets for each student at each grade based on the student's scores on the Standards Based Assessment (SBA) tests for mathematics and language arts. The SBA test was first given in the 2004–05 school year to students in grades 3–9. The SBA is scaled such that students in every grade must score 300 or above to be considered proficient in math and a combined reading and writing score of 600 or above to be considered proficient in language arts.[16]

Starting in the 2006–07 school year, fourth- through sixth-grade students scoring below the proficiency cutoff were counted as proficient for AYP purposes if they were on-track to reach proficiency by the seventh grade, and eighth- and ninth-graders were counted as proficient if they were on-track to reach proficiency by the tenth grade. Students are always classified only by the status model in grades 3 and 7. Annual growth targets for all fourth-, fifth-, sixth-, eighth-, and ninth-graders who scored below proficiency were set by dividing the difference between the student's "baseline" test scores and the proficiency cutoffs (300 in math and 600 in language arts) by the number of years allowed.

The baseline scores were defined as those from the first year the student scored below the proficiency cutoff. For students in grades 4 and 5 in the 2006–07 school year, this could go back to the third grade. The baseline score for sixth-graders in 2006–07 could go back to the fourth grade (i.e., to 2004–05 when the SBA was first administered). Concretely, for fourth-grade students who scored below proficiency in both the third and fourth grades, the baseline is the third-grade score. For fifth-grade students who scored below proficiency in the third and fourth grades, their third-grade scores would also be the baselines used to define their trajectories to proficiency by grade 7. The baseline for students in grades 8 and 9 who score below proficiency can go back to seventh grade.

Thus a third-grade student who scored below proficiency has four years of allowable growth and must make up one-quarter of the gap between adjusted score and proficiency cutoff in the first year, one-third of this proficiency gap the second year, one-half of the gap the third year, and must score at or above the cutoff the fourth year in order to be counted as proficient for AYP determinations. Similarly, a seventh-grader scoring below proficiency must close the gap by one-third in grade 8 and by half in grade 9.[17]

The Alaska growth model only applies to students who score below proficient on the SBA in their first year of eligibility. This means that students achieving a proficient score in the third through eighth grades but scoring below the cutoff in following years are not eligible for the growth model. Students who change schools or districts carry their baseline score with them.

The Alaska growth model uses a "status-plus-growth" method for determining whether reporting groups meet the AMOs for language arts and mathematics. That is, first, the reporting group is assessed in terms of the percentage of students who scored at or above proficiency. If that

[16] Alaska estimates a "true score" for the SBA by adjusting for the annual reliability of the tests. This has the effect of raising scores that are below average. The estimated true score is calculated by subtracting an estimated "reliable deviation" of the student's observed score from the statewide average for students in the same grade.

[17] Note that a ninth-grader must close the proficiency gap by 100 percent by the tenth grade, meaning that the growth model is equivalent to the status model for all ninth-graders who score below proficiency.

percentage met the AMO, the group was classified as "met by status." If the percentage did not meet the AMO but the percentage not proficient was 10 percent or more less than the percentage not proficient in the prior year, the group was classified as "met by safe-harbor." If the group did not meet the AMO by status or pass by safe-harbor, then the number of below-proficient students who are on-track to proficiency was added to the number of proficient students and this sum was divided by the total number of students in the reporting group who took the test. If this percentage met the AMO, the group was classified as "met by growth." If one or more reporting groups met by growth and the other AYP criteria noted below were also met, the school as a whole was classified as "made AYP by growth."

Alaska's *ESEA* accountability rules require a reporting subgroup to have more than 25 students before it is included in school AYP determinations. Subgroups larger than 40 students will count only if 95 percent of them participate in testing, while smaller subgroups require that no more than two students fail to participate. Schools with students and subgroups who meet the AMO within a 99 percent confidence interval, which varies by school size, are determined to have made AYP.

The AYP results for Alaska in 2006–07 are shown in Exhibit 11. Of the 323 schools that met AYP requirements for that year, representing 66 percent of all eligible schools, no school made AYP via the state's pilot growth model.

Exhibit 11
Alaska School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	323	66%
Met with Safe-Harbor	0	0%
Met with Growth	0	0%
Not Met	169	34%
All Eligible Schools	492	100%

Exhibit reads: For Alaska's schools overall, 323 met AYP under status, which was 66 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Arizona's growth model

Arizona's participation in GMPP was approved on July 3, 2007, for use beginning in the 2006–07 school year. The Arizona growth model includes students in grades 4 through 7 and uses scores on the Arizona Instrument to Measure Standards (AIMS) tests for reading and mathematics, which is vertically scaled for grades 3 through 8. Proficiency levels are set for each grade, and a student scoring "below the standard" for either reading or math has three years to reach the cutoff or by eighth grade, whichever comes first. Growth targets are set by dividing the difference between initial score and proficiency cutoff three grades later into equal parts. In

order to be counted as proficient for AYP purposes, a student in grades 3 through 5 must make up one third of the shortfall in each of the next three years. A sixth-grader must cover half of the shortfall in each of the two years of eligibility remaining.[18] Students who leave the Arizona school system before reaching proficiency have new growth targets set upon their return.

For the purposes of its growth model, Arizona adjusts AIMS scores using past scores to correct for improvement that might be due to chance. These "predicted" scores are then subject to a 95 percent confidence interval, so that a score just below the growth target can still be counted as proficient if it is within the range of statistical error due to the prediction itself. The state also uses a 99 percent confidence interval around annual AMOs for AYP determinations.[19] *ESEA* subgroups are counted in such determinations only if they have at least 40 students and if no fewer than 95 percent of these students participate in annual testing. Students who have estimated scores above the proficiency cutoff but who also fail to meet their growth targets continue to be counted as proficient for AYP purposes.

Arizona applied the results of the growth model after applying status and safe-harbor criteria. Growth model results were only applied to *ESEA* reporting groups that did not make AYP by status or safe-harbor. For those groups the number of students identified as "on-track to proficiency" was added to the number of proficient students and that sum was divided by the number of test-takers in the reporting group. That percentage was then compared to the AMO. If the school met all AYP criteria and had one or more reporting group meeting the AMO because of the addition of on-track students, the school as a whole was classified in ED*Facts* as making AYP by growth.

ED*Facts* data for the 2006–07 school year indicate that Arizona's growth model resulted in only one school making AYP that would have missed if only status and safe-harbor were applied (Exhibit 12). The other 1,096 schools that met AYP requirements did so using status criteria alone.

[18] Note that an eighth-grader must close the proficiency gap by 100 percent in the first year, meaning that the growth model is equivalent to the status model in this grade.

[19] Confidence intervals around the AMO (as opposed to around the percent proficient) are calculated using the number of full academic year students in the reporting group as the basis for the estimate. If the percent proficient in the group meets or exceeds the lower bound of this confidence interval, the group is classified as meeting the AMO.

Exhibit 12
Arizona School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	1,096	77%
Met with Safe-Harbor	0	0%
Met with Growth	1	<1%
Not Met	333	23%
All Eligible Schools	1,430	100%

Exhibit reads: For Arizona's schools overall, 1,096 met AYP under status, which was 77 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Arkansas' growth model

Arkansas received approval to implement its proposed growth model in the 2006–07 school year. Arkansas uses a trajectory model that calculates growth for students in grades 4 through 7 using results of the Arkansas Benchmark Exams for mathematics and literacy, which are administered in grades 3 through 8 (plus grade 11 for literacy). Proficiency levels for these vertically scaled exams are set for each grade, and growth targets are based on the annual exam score increment needed to reach the proficiency standard in eighth grade. The annual increment that a student must attain in order to be classified as on-track to proficiency is calculated using grade-specific growth target multipliers of 0.295 in fourth grade, 0.319 in fifth grade, 0.385 in sixth grade, and 0.542 in seventh grade. These multipliers indicate the proportion of the total difference between the eighth-grade standard (a score of 700) and the student's current score that the student must gain over the next year in order to be on-track for eighth-grade proficiency. They contrast with the multipliers in other trajectory states (which all represent fractions of whole years – e.g., one-fourth, one-third, one-half) because Arkansas Benchmark Exams are not scaled to have a linear progression from one year's cut point to the next (see Appendix B for more detail and illustrations).

An important feature of this model is that it resets the growth target every year rather than setting a series of annual targets based on the first below-proficient exam score. For example, a third-grader who scores a 480 on the Arkansas Benchmark Exams is 20 points below the standard for that grade. Since the eighth-grade standard is 700, the student must gain 65 points (= (700–480) * 0.295) to be counted as on-track to proficiency. If the student scores a 539 in fourth grade (again 20 points below the standard), he or she fails to reach this threshold and now needs to gain 51 points (=(700–539) * 0.319) in fifth grade to be counted as on-track to proficiency. Arkansas students who make sufficient growth are counted as proficient for AYP determinations.

Arkansas' growth model requires students who score above the proficiency threshold to improve toward the eighth-grade standard of 700 at the same basic grade-level rate as non-proficient students in order to count as on-track to proficiency. This means that above-proficient students

who continue to score above the proficiency threshold can nonetheless fail to realize adequate growth to count toward school AYP determinations under the growth model.

Like Arizona, Arkansas applies confidence intervals to the grade-specific AMOs. In contrast, the other states in the GMPP apply confidence intervals to the percentages of proficient students. The confidence interval is 95 percent and is applied by using the number of students enrolled in the school's tested grades for each reporting subgroup.

The data reported to ED*Facts* by Arkansas for the 2006–07 school year show that 24 percent of the schools made AYP according to the status model, 32 percent made AYP under safe-harbor provisions, and 8 percent made AYP under the growth model (Exhibit 13). These 8 percent represent schools that would not have made AYP had the GMPP not been available.

Exhibit 13
Arkansas School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	212	24%
Met with Safe-Harbor	287	32%
Met with Growth	69	8%
Not Met	326	36%
All Eligible Schools	894	100%

Exhibit reads: For Arkansas' schools overall, 212 met AYP under status, which was 24 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

The percentage of Arkansas schools making AYP with safe-harbor (32 percent) is much higher than in any of the other GMPP states, but it is not clear why this occurred.

Delaware's growth model

Delaware also received approval to use its proposed growth model starting with the 2006–07 school year. Delaware's model includes students in grades 3 through 10 and does not limit the number of years students have to make proficiency. The Delaware Student Testing Program (DSTP) yields vertically aligned scale scores for mathematics and language arts (a combination of reading and writing). Mathematics and reading tests are administered starting in grade 2 and growth can be assessed in grade 3 and above, but the writing component of the language arts scores do not start until grade 3.

The Delaware model is an example of the transition matrix type. Delaware uses a "value table" method for AYP determinations that assigns points for students depending on the type and extent of changes between the performance levels (see Exhibit 6). Points in the value table increase with the level of proficiency, so that a student scoring at the bottom receives fewer points for

moving up one level (150) than a student moving from that level to the next (175). Conversely, students who move up two levels are awarded more points (225) than a student who moves to the same level in one step (175 or 200). All students who surpass their grade level cutoff receive 300 points regardless of how high they score or whether their underlying scores actually declined from the prior year. AYP is determined by calculating the average points for the school and its subgroups in the two required subject areas and comparing these averages with the state's AMO levels.

Delaware was unique in the pilot project in that it applied both the status model and the growth model to all schools, and classified all schools that made AYP under the growth model as "met AYP by growth" even if they also met AYP by status. The other states employed the status-plus-growth model shown in Exhibit 4. Delaware's arrangement makes it impossible to determine from ED*Facts* whether the schools making AYP under growth also made AYP under status, and thus to answer the question of how many Delaware schools failed to make AYP under status but did make it under growth.

Drawing on the state reports in ED*Facts*, 83 of 180 schools in Delaware made AYP under the growth model in the 2006–07 school year (Exhibit 14). As a result of using the procedure of employing the growth model results before applying the status and safe-harbor provisions, the percentage of Delaware schools reported in ED*Facts* as making AYP under growth (46 percent) is much higher than any other state. Many of these 83 schools would have made AYP under status or safe-harbor, had those criteria been applied before (as was done in the other GMPP states) the growth criteria.

Exhibit 14
Delaware School AYP Determinations With Growth Model Results Augmented With Status and Safe-Harbor Results, 2006–07

School AYP Under Growth-Plus-Status Model	Number	Percent
Met with Status but not Growth	21	12%
Met with Safe-Harbor but not Growth	19	11%
Met with Growth	83	46%
Not Met	57	31%
All Schools	180	100%

Exhibit reads: For Delaware's schools overall, 21 met AYP under status but not growth, which was 12 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

The ED*Facts* reporting on school AYP outcomes used mutually exclusive categories of "by status" and "by growth" and it was thus not possible to address the question of how many Delaware schools that made AYP under their growth model would have missed AYP under the status model or by safe-harbor. However, data provided by the state directly to the evaluation team indicate that only five schools, or 3 percent of the schools in the state, made AYP because

of the growth model provisions (Exhibit 15).[20] These were the only schools that were reclassified from not making AYP to making AYP as a direct result of the Delaware GMPP growth criteria.

Exhibit 15
Delaware School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	101	55%
Met with Safe-Harbor	22	12%
Met with Growth	5	3%
Not Met	57	31%
All Eligible Schools	185	100%

Exhibit reads: For Delaware's schools overall, 101 met AYP under status, which was 55 percent of all eligible schools.

Source: Delaware Department of Education.

Florida's growth model

Florida's growth model proposal was approved for use starting with the 2006–07 school year. The growth model applies to students in grades 3 through 10 using the Developmental Scale Scores (DSS) from the Florida Comprehensive Assessment Tests (FCAT) for mathematics and reading. The state uses a trajectory model that bases growth targets on the score required for proficiency three years after the first year tested (normally, grade 3). To be counted as proficient, a student who was not proficient at the baseline must close the gap between their baseline score and the proficiency cutoff three grades later by one-third each year.[21] Students who continue to score below the cutoff after three years start the process over and have new growth targets set.

Florida retains third-graders who score below proficient, and the growth model incorporates such students by using their two third-grade scores on the FCAT DSS. Otherwise, only students in the fourth grade or later will have the two scores required to calculate growth.

For determining school AYP, Florida applies growth-model results for students only in *ESEA* reporting groups that do not meet their AMOs by status or safe-harbor. An important feature of Florida's accountability model is that, for groups not meeting their AMO, the growth-model data

[20] These were final (summer) data for 2006–07 school year provided to NORC by Robin Taylor, associate secretary of education for the Delaware Department of Education's Assessment and Accountability Branch, on Aug. 14, 2008. The data differed from the ED*Facts* reports in that both status and growth results were included for all schools and *ESEA* reporting groups. The data also included five schools not reported in ED*Facts*.

[21] This means that a student can only make proficiency by growth for two years since the gap must be fully closed in the third year, i.e., a student must meet or exceed the minimum proficiency score in year 3. A student first enrolled in grade 9 must close the gap by half to be counted as "on-track."

are used exclusively. That is, the number of non-proficient but on-track students is not added to the number of proficient students as is done in several other pilot states. Instead, the growth model results for both the proficient and non-proficient students are used. This means that students who are proficient but who are not on-track to maintain proficiency in three years are counted as "not proficient" for AYP purposes in these reporting groups. Groups not meeting AMO by status or safe-harbor must thus meet the AMO entirely on the basis of "on-track" students.

The AYP results for Florida in 2006–07 are shown in Exhibit 16. A total of 34 percent of Florida schools made AYP, with 24 percent making it under status, 5 percent by safe-harbor, and 5 percent by growth. The percentage of schools that made AYP in Florida is much lower than the other GMPP states, but this may reflect higher standards for proficiency rather than lower performance.

Exhibit 16
Florida School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	770	24%
Met with Safe-Harbor	155	5%
Met with Growth	149	5%
Not Met	2,135	67%
All Eligible Schools	3,209	100%

Exhibit reads: For Florida's schools overall, 770 met AYP under status, which was 24 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Iowa's growth model

Iowa's growth model was approved for use with students in grades 3 through 8 beginning in the 2006–07 school year. Iowa uses third-grade math and reading scores on the Iowa Test of Basic Skills (ITBS) as a baseline, so growth calculations begin in fourth grade. Iowa's growth model is a type of transition matrix model (see Exhibit 5). To calculate growth, Iowa uses two categories of proficient (Intermediate and High) and three categories of below proficient (Weak, Low Marginal, and High Marginal). Proficiency thresholds for each category are set using national percentile ranks for each grade, with ITBS scale scores in the 40th percentile considered below proficient and 10th percentile scores considered Weak. Non-proficient students can still make Adequate Yearly Growth (AYG) if they improve to a higher category of below proficient within four years of the first year tested.

An important feature of Iowa's model is that students cannot fall back to a non-proficient category and still make AYG. This means that High Marginal students who decline to Low Marginal or Weak cannot make AYG simply by regaining High Marginal status, and that Low

Marginal students who decline to Weak must score in the High Marginal category to make adequate growth. Iowa counts all students making AYG as proficient for *ESEA* reporting groups that fail to make AYP when the status and safe-harbor provisions have been applied.

The 2006–07 AYP results for Iowa show that 11 percent of the schools made AYP with the growth model in the status-plus-growth framework (Exhibit 17). This was on top of 75 percent of the schools making AYP with status and 10 percent via safe-harbor. The percentage making AYP by growth was the highest of the eight pilot states, and the overall rate of AYP among Iowa schools (above 95 percent) was higher than any of the other GMPP states. The concurrence of high rate of making AYP by status and by growth suggests there is not necessarily a "ceiling effect" on making AYP, at least with Iowa's implementation.

Exhibit 17
Iowa School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	831	75%
Met with Safe-Harbor	105	10%
Met with Growth	116	11%
Not Met	52	5%
All Eligible Schools	1,104	100%

Exhibit reads: For Iowa's schools overall, 831 met AYP under status, which was 75 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

North Carolina's growth model

North Carolina received approval on May 17, 2006, to use its growth model for the 2005–06 school year. This model calculates growth in grades 3 through 7 using vertically equated North Carolina end-of-grade tests for mathematics and reading. Third-graders in North Carolina have four years to grow to proficiency because they take a test upon entering the third grade and a test at the end of the year. All other students take end-of-grade tests and thus have only three years to become proficient or until eighth grade, whichever comes first.

For students that take the third-grade pretest, growth targets are set by dividing the difference between the initial test score and the proficiency cut score for sixth grade into four equal parts. Thus third-grade students who make up one-fourth of the shortfall between the baseline score and sixth-grade standards by the end of third grade are considered to be "on-track" to proficiency. Students who enter the school system after third grade must close this gap more quickly since they will have fewer than four years until eighth grade, when all students are expected to make a proficient score.

For AYP purposes, North Carolina applies growth-model results after status and safe-harbor criteria. For reporting groups not meeting their AMO by status or safe-harbor, the number of non-proficient students who are on-track to proficiency per the growth model is added to the number of proficient students in the group. A group that reaches the state's AMO with the inclusion of "on-track" students is considered to have met the AMO by growth and the school as a whole is classified as having made AYP by growth. North Carolina applies a 95 percent confidence interval before using the growth model for AYP determinations and only includes subgroups with at least 40 full academic year (starting on Oct. 1 of the current school year) students. The standard 95 percent participation rule also applies to subgroups and schools and is based on the full set of students enrolled in the current school year.

Exhibit 18 shows that the pilot growth model resulted in 12 more North Carolina schools making AYP in 2006–07 than would have made AYP through status (661) and safe-harbor (308) alone. More than half (56 percent) of schools in North Carolina did not meet their AYP requirements under the GMPP.

Exhibit 18
North Carolina School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	661	30%
Met with Safe-Harbor	308	14%
Met with Growth	12	1%
Not Met	1,226	56%
All Eligible Schools	2,207	100%

Exhibit reads: For North Carolina's schools overall, 661 met AYP under status, which was 30 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Tennessee's growth model

Tennessee received approval on May 17, 2006, to use its growth model beginning in the 2005–06 school year.[22] The state calculates growth for students in grades 4 through 8 using vertically aligned scores for mathematics and reading or language arts from the Tennessee Comprehensive Assessment Program (TCAP). Students are first tested in the third grade, and those who score below proficient are given until ninth grade to reach proficiency by growth. Tennessee's model establishes growth targets for students by estimating a regression equation to predict each student's ninth-grade achievement level on the basis of their current and prior test scores. The regressions are estimated using data from all students at each grade level in the state.

[22] This approval stipulated that Tennessee work with the U.S. Department of Education to ensure that results from alternate assessments using different achievement standards were included by the 2006–07 school year.

Students who score below proficiency but who are predicted to be above the TCAP cut scores by grade 9 are counted as proficient for AYP growth determinations.

For students past the fifth grade, proficiency cut scores are based on the TCAP high school graduation (or Gateway) assessments. Another important feature of Tennessee's growth model is that projections are possible even if a student is missing prior achievement scores for some grades or subjects. Also, the model recalculates projections every year and calculates projections for students who score above proficiency.

For AYP purposes, the growth-model results are applied only to *ESEA* reporting groups that did not meet their AMOs by status or safe-harbor. As in Florida, Tennessee uses growth-model results for all students in groups not meeting their AMO. That is, proficient students in those groups who are projected to score below the cutoff in three years count are not counted as proficient. Tennessee includes subgroups in AYP determinations if they have at least 45 full academic year students and applies a 95 percent confidence interval to the AMO (like Arizona and Arkansas) instead of to the percent proficient. A school in Tennessee making AYP is considered to have made it by growth if including students projected to be proficient allows all non-proficient subgroups to meet the AMO.

Exhibit 19 shows that 1,217 (or 89 percent) of 1,362 Tennessee schools met AYP requirements in 2006–07 through status and safe-harbor criteria. An additional 19 schools (or one percent) that failed to meet these criteria made AYP by also applying the pilot growth model.

Exhibit 19
Tennessee School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	1,149	84%
Met with Safe-Harbor	68	5%
Met with Growth	19	1%
Not Met	126	9%
All Eligible Schools	1,362	100%

Exhibit reads: For Tennessee's schools overall, 1,149 met AYP under status, which was 84 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Impact of GMPP on AYP

The pilot states varied in the proportion of their schools that made AYP under status or safe-harbor, and the differences can obscure the impact of their growth models on AYP. One way to assess the impact is to calculate the percentage increase in the number of schools making AYP due to use of the growth model. The schools making AYP uniquely by growth represented a percentage increase in the schools making AYP of 6 percent across all states, and ranged as high as 12 percent in Iowa, 14 percent in Arkansas, and 16 percent in Florida (Exhibit 20).

Another indicator of the impact of the GMPP is the extent to which the growth models affected the pool of schools identified as failing both the status and safe-harbor criteria for AYP. Across all eight states, 8 percent of the 4,795 schools that did not make AYP by status or safe-harbor made AYP by growth (Exhibit 20). These rates ranged from a high of 69 percent of eligible schools in Iowa making AYP by growth to 1 percent or fewer of eligible schools in Alaska, Arizona, and North Carolina making AYP by growth.

Exhibit 20
Percentage Increase in Number of Schools That Made AYP Due to Growth, and Percentage Decrease in Number of Schools That Did Not Make AYP Due to Growth, by State, 2006–07

Pilot States	Number of Schools Making AYP by Status or Safe-Harbor	Number of Schools not Making AYP by Status or Safe-Harbor	Number of Schools Not Making AYP by Status or Safe-Harbor That Met by Growth	Percentage Increase in Schools Making AYP Due to Growth	Percent of Schools Not Making AYP by Status or Safe-Harbor That Met by Growth
All Eight States	6,088	4,795	371	6%	8%
Alaska	323	169	0	0%	0%
Arizona	1,096	334	1	<1%	< 1%
Arkansas	499	395	69	14%	17%
Delaware	123	62	5	4%	8%
Florida	925	2,284	149	16%	7%
Iowa	936	168	116	12%	69%
North Carolina	969	1,238	12	1%	1%
Tennessee	1,217	145	19	2%	13%

Exhibit reads: The 371 schools that made AYP by growth increased the number of schools making AYP from 6,088 to 6,459 schools, which was a percentage increase of 6 percent. Of the 4,795 schools that did not make AYP under either status or safe-harbor, 371 or 8 percent made AYP using the growth model.

Source: U.S. Department of Education, EDFacts and the Alaska, Arizona, Delaware, North Carolina, and Tennessee state departments of education.

Discussion

The idea of using growth models to assess school academic performance is attractive and the data collection systems and database architecture needed to support such models are rapidly becoming available. The pilot models were strictly add-ons to the traditional status-plus-safe-harbor models for determining AYP in all eight states. Even Delaware, the only state that applied the growth model and then augmented with the status model, used the growth model as an add-on instead of a replacement. While states had the option of applying growth criteria instead of status and safe-harbor, none exercised that option. Growth criteria were instead added to status and safe-harbor criteria in ways that could only increase the number of schools making AYP.

States also had the option of applying growth criteria before status and safe-harbor were applied, but Delaware was the only state that did so. Delaware's growth-plus-status procedure had the effect of officially recognizing about 40 percent of their schools as making AYP by growth, which was a much higher percentage than under the status-plus-growth procedure used in the other states. While this may have had a positive effect on the public visibility and understanding of the model in Delaware, the actual number of growth schools that would not have made AYP under status or safe-harbor was small.

As shown in Exhibits 11 through 19, the percent of schools that made AYP using the growth provisions of the GMPP, and which would not have made AYP without those provisions, ranged from a high of 11 percent of schools in Iowa to 0 percent in Alaska. Another perspective on these rates is gained when the number of schools making AYP under growth is compared not to the total number of schools in the state but instead to either the number of schools in the state that made AYP under either status or safe-harbor (the percentage increase due to growth) or the number that did not make AYP under either status or safe-harbor (percentage decrease in non-AYP due to growth). The percentage increases in schools making AYP ranged from a high of 16 percent in Florida to less than 2 percent in Alaska, Arizona, North Carolina, and Tennessee. The growth model decreased the number of schools identified as failing under the status-plus-safe-harbor model by 69 percent in Iowa but by less than 1 percent in Alaska, Arizona, and North Carolina. The rate of making AYP by growth after missing by status and safe-harbor ranged from 7 percent to 17 percent in the remaining four pilot states.

Reasons for the variation among states in the percentages of schools making AYP by growth will be examined systematically in the final report but are likely to be some combination of differences in the assessments used and the proficiency cutpoints, the features of the growth models implemented, the states' AMOs, and the actual levels of growth realized by the students. It is noteworthy that the percentages making AYP by growth are not correlated with the percentages making AYP by status or safe-harbor (both of which vary greatly among these states) or to whether the state adopted a transition matrix, trajectory, or projection type of growth model. The exceptionally high rate in Iowa may be at least partly due to a more inclusive definition of growth toward proficiency used there (that is, any improvement of one or more performance levels). In contrast, the very low rates in other states may reflect the impact of the various other (non-growth) methods for determining AYP available in those states for schools to make AYP by status and safe-harbor (e.g., confidence intervals and multiyear averaging), such that the status and safe-harbor methods picked up schools which would have made AYP by growth had those various provisions not been available.

III. School Characteristics Associated with AYP Outcomes

This chapter addresses the question of whether certain school characteristics are associated with making AYP under growth but not status or safe-harbor. The school characteristics are organizational and demographic variables that include *ESEA* improvement status, poverty level, minority concentration, locality, and size. Of particular interest here is how pilot growth models affect the AYP designation of schools serving disadvantaged populations. Because such schools have been found to make AYP at much lower rates than those serving more affluent populations (U.S. Department of Education, 2007), the GMPP may function to lessen the AYP gap on this dimension.

The question addressed in this chapter is whether certain types of schools are more or less likely to make AYP by the growth criteria. To answer this question, the following analyses look at the percentage increase of schools of each type that made AYP because of the availability of the growth model under the GMPP. The percentage increase is calculated by dividing the number of schools making AYP by growth but not by status or safe-harbor by the number that made AYP by either status or safe-harbor. Any difference between comparison groups in the percentage increases of schools making AYP by growth will suggest that the GMPP had a disproportionate effect on certain types of schools.

The U.S. Department of Education report *State and Local Implementation of the No Child Left Behind Act, Volume II—Accountability Under NCLB: Interim Report* (Le Floch, Martinez, O'Day, et al. 2007) found that certain school demographic characteristics were associated with the likelihood of making AYP. For example, the study found that high-poverty, high-minority, larger, and urban schools were less likely to make AYP by status criteria (Le Floch, Martinez, O'Day, et al. 2007: 39, 40). Another important characteristic considered here is the *ESEA* improvement status classifications of schools. This chapter uses the data reported in ED*Facts* to characterize schools that made AYP under growth but not under status or safe-harbor. As discussed in Chapter II, Delaware reported results to ED*Facts* based on growth model results augmented with status and safe-harbor results. In order to identify the incremental impact of the growth model on schools' likelihood of making AYP, we substitute data using status and safe-harbor augmented with growth for the Delaware ED*Facts* report.

ESEA *School Improvement Status.* An important issue for *ESEA* accountability is the extent to which schools identified for improvement are, despite their relatively low levels of performance, actually making progress toward the goal of universal proficiency. The growth model pilot is in part intended to identify such schools. States are required under *ESEA* to identify for improvement Title I schools that do not meet AYP for two consecutive years and to initiate a process of interventions designed to improve student outcomes; most states have applied a similar process for non–Title I schools. Schools not making AYP for two consecutive years are officially identified for improvement. If a school does not make AYP after two years in the improvement status, it is identified for corrective action. If the school misses AYP for an additional year, it moves into restructuring status (Le Floch, Martinez, O'Day, et al. 2007: 3-6).

The overall results shown in Exhibit 21 indicate that the percentage increase in schools making AYP due to the growth model was greater among schools identified for improvement or corrective action (19 percent) than schools not identified for improvement or corrective action (5

percent). The pilot states varied greatly in the impact of the GMPP on the different types of schools. No schools identified for improvement or corrective action in Alaska, Arizona, or Delaware made AYP by growth. In contrast, Arkansas and Florida had substantial numbers of schools identified for improvement or corrective action that made AYP by growth (21 and 37 schools, respectively). These counts translated into percentage increases in AYP among schools identified for improvement or corrective action attributable to the GMPP of 27 percent in Arkansas and 34 percent in Florida.

Exhibit 21
Numbers of Schools Making AYP Under Status-Plus-Safe-Harbor and Growth, and Percentage Increase Due to Growth, by *ESEA* School Improvement Status, 2006–07

Pilot States	Made AYP Based on Status or Safe-Harbor			Made AYP Due to Growth Model			Percentage Increase Due to Growth		
	Improvement or Corrective Action	Restructuring	Not Identified	Improvement or Corrective Action	Restructuring	Not Identified	Improvement or Corrective Action	Restructuring	Not Identified
All Eight States	351	23	5,618	66	3	302	19%	13%	5%
Alaska	68	9	246	0	0	0	0%	0%	0%
Arizona	36	5	1,055	0	0	1	0%	0%	<1%
Arkansas	78	0	421	21	0	48	27%	0%	11%
Delaware	5	0	118	0	0	5	0%	0%	4%
Florida	110	1	814	37	2	110	34%	200%	14%
Iowa	1	0	935	5	0	111	500%	0%	12%
North Carolina[a]	33	0	851	1	1	10	3%	0%	1%
Tennessee	20	8	1,178	20	8	1,178	10%	0%	1%

Exhibit reads: Of the schools across all eight pilot states that were either identified for improvement or under corrective action, 351 made AYP under status or safe-harbor in 2006–07 and another 66 schools made AYP under growth in 2006–07. The schools that made AYP under growth increased the number of schools that made AYP in 2006–07 by 19 percent.

[a] Results for North Carolina are from 2005–06 since the state did not provide data on school improvement status in 2006–07.

Source: U.S. Department of Education, ED*Facts*.

The 2007 Interim Report on the implementation of *NCLB* found that schools' likelihood of making AYP under the status model was strongly correlated with student socioeconomic status and racial or ethnic minority representation (Le Floch, Martinez, O'Day, et al. 2007: 39).[23] Exhibit 22 (below) shows that for all eight states combined, the percentage increase in schools making AYP due to the growth models was greater among high- and medium-poverty schools 8 and 7 percent increases, respectively) than among low-poverty schools (3 percent increase). This pattern held in Arkansas, Florida, Iowa, North Carolina, and Tennessee but not in Alaska, Arizona, or Delaware, where no high-poverty schools made AYP by growth. The differential

[23] Poverty level is defined here in terms of the percentages of students eligible for the federal free and reduced-price lunch program (FRPL). Low-poverty schools enroll 25 percent or fewer FRPL-eligible students, medium-poverty schools enroll 26 percent to 75 percent FRPL-eligible students, and high-poverty schools enroll more than 75 percent.

impact of the pilot growth models on making AYP between low- and high-poverty schools was greatest in Arkansas, Florida, and Iowa.

Exhibit 22
Numbers of Schools Making AYP Under Status-Plus-Safe-Harbor and Under Growth, and Percentage Increase in AYP Due to Growth, by School Poverty Concentration, 2006–07

Pilot States	Made AYP Based on Status or Safe-Harbor			Made AYP Due to Growth Model			Percentage Increase Due to Growth		
	High Poverty	Medium Poverty	Low Poverty	High Poverty	Medium Poverty	Low Poverty	High Poverty	Medium Poverty	Low Poverty
All Eight States	737	3,674	1,606	62	265	44	8%	7%	3%
Alaska	50	183	90	0	0	0	0%	0%	0%
Arizona	174	444	478	0	1	0	0%	0%	0%
Arkansas	43	403	53	16	50	3	37%	12%	6%
Delaware	3	87	33	0	3	2	0%	3%	6%
Florida	75	515	335	24	109	16	32%	21%	5%
Iowa	22	618	296	8	86	22	36%	14%	7%
North Carolina	99	665	172	3	8	1	3%	1%	1%
Tennessee	271	759	149	11	8	0	4%	1%	0%

Exhibit reads: Among high-poverty schools in the eight pilot states, 737 made AYP based on status or safe-harbor in 2006–07 and an additional 62 schools made AYP due to the application of growth models, resulting in an 8 percent increase in the number of high-poverty schools that made AYP in those states.

Source: U.S. Department of Education, ED*Facts*.

Results for school minority composition are shown in Exhibit 23.[24] The percentage increase in high-minority schools making AYP by growth across all eight states (7 percent) was slightly higher than the percentage increase among low-minority schools (4 percent) and slightly lower than the percentage increase among medium-minority schools (9 percent). This pattern again held in all states except Alaska, Arizona, and Delaware, states where none of the high-minority schools made AYP by growth. The differential impact of the pilot growth models on AYP rates of low- and high-minority schools was greatest in Iowa, Arkansas, and Florida.

[24] Minority level is defined here in terms of the percentages of nonwhite or Hispanic students. Low-minority schools enroll 25 percent or fewer nonwhite or Hispanic students, medium-minority schools enroll 26 percent to 75 percent nonwhite or Hispanic students, and high-minority schools enroll more than 75 percent.

Exhibit 23
Numbers of Schools Making AYP Under Status-Plus-Safe-Harbor and Under Growth, and Percentage Increase in AYP Due to Growth, by School Minority Concentration, 2006–07

Pilot States	Made AYP Based on Status or Safe-Harbor			Made AYP Due to Growth Model			Percentage Increase Due to Growth		
	High Minority	Medium Minority	Low Minority	High Minority	Medium Minority	Low Minority	High Minority	Medium Minority	Low Minority
All Eight States	865	2,032	3,189	59	173	139	7%	9%	4%
Alaska	102	116	105	0	0	0	0%	0%	0%
Arizona	274	544	277	1	0	0	0%	0%	0%
Arkansas	24	123	351	7	38	24	29%	31%	7%
Delaware	8	97	18	0	3	2	0%	3%	11%
Florida	174	391	360	38	88	23	22%	23%	6%
Iowa	3	87	846	2	32	82	67%	37%	10%
North Carolina	92	423	454	4	6	2	4%	1%	0%
Tennessee	188	251	778	7	6	6	4%	2%	1%

Exhibit reads: Among high-minority schools in the eight pilot states, 865 made AYP based on status or safe-harbor in 2006–07 and an additional 59 schools made AYP due to the application of growth models, resulting in a 7 percent increase in the number of high-minority schools that made AYP in those states.

Source: U.S. Department of Education, ED*Facts*.

School location in an urban, suburban, or rural area is another variable that previous studies have found related to AYP results, with rural and suburban schools generally more likely to make AYP under status or safe-harbor than urban schools. Exhibit 24 shows that, in most states, the growth model moved relatively more urban schools than rural schools into AYP. Delaware was the only state among the eight to exhibit the opposite trend, and Alaska and Arizona had no schools making AYP by growth in any category. Except for Arkansas and Iowa, the percentage increases in urban and suburban schools making AYP were of similar magnitude.

Exhibit 24
Numbers of Schools Making AYP Under Status-Plus-Safe-Harbor and Under Growth, and Percentage Increase in AYP Due to Growth, by School Urbanicity, 2006–07

Pilot States	Made AYP Based on Status or Safe-Harbor			Made AYP Due to Growth Model			% Increase Due to Growth		
	Urban	Suburban	Rural	Urban	Suburban	Rural	Urban	Suburban	Rural
All Eight States	1,446	2,035	2,440	104	174	84	7%	9%	3%
Alaska	57	65	200	0	0	0	0%	0%	0%
Arizona	479	371	178	0	0	0	0%	0%	0%
Arkansas	61	119	303	16	18	32	26%	15%	11%
Delaware	17	75	31	0	1	4	0%	1%	13%
Florida	202	553	146	36	91	18	18%	16%	12%
Iowa	126	232	545	37	54	24	29%	23%	4%
North Carolina	165	262	520	5	4	3	3%	2%	1%
Tennessee	339	358	517	10	6	3	3%	2%	1%

Exhibit reads: Among urban schools in the eight pilot states, 1,446 made AYP based on status or safe-harbor in 2006–07 and an additional 104 schools made AYP due to the application of growth models, resulting in a 7 percent increase in the number of urban schools that made AYP in those states.

Source: U.S. Department of Education, ED*Facts*.

Discussion

The answer to the question, "Is the likelihood of making AYP under growth models associated with school characteristics?" varied across the six pilot states where the growth models had an impact on AYP determinations. Looking first at schools identified for improvement under *ESEA*, the percentage increase of those schools making AYP due to the growth model was greater than among schools not identified for improvement. Turning to demographic characteristics of the schools, the growth component was more likely to identify schools as making AYP among those with higher concentrations of poverty-level and minority students compared to those with lower concentrations in Arkansas, Florida, Iowa, North Carolina, and Tennessee. In Delaware, more advantaged schools were more likely to have made AYP by growth than by status or safe-harbor. The effect of growth models on AYP was most pronounced among schools enrolling high proportions of students from poverty-level households. Across all pilot states combined, the percentage increase in AYP because of the growth model was greater among high-poverty schools than among low-poverty schools, and the differences in the increases were particularly large in Arkansas, Florida, and Iowa.

IV. Using Student Growth Indicators to Assess School Outcomes

This chapter addresses the hypothetical question of how AYP determinations would be affected if the data collected by the GMPP states were used in various different ways. Four specific questions are addressed: (1) How do the models compare in terms of the number of schools that made AYP by status criteria and also made their state-defined "annual measurable objective" (AMO) by growth criteria? (2) How do the models compare in terms of the number of schools that made AYP by safe-harbor and also reached their AMO by growth criteria? (3) How do the models compare in terms of the demographic characteristics of schools that made their AMO by status as well as by growth criteria? and (4) To what extent did the pilot states have test scores for two or more years for all students, and did the non-matched students differ from their matched counterparts?

In theory, growth models could provide a more useful way of assessing school academic performance than the traditional status-plus-safe-harbor-model. The GMPP growth models are designed to assess whether or not each student is on-track to reach or exceed the proficiency standards within a specified time frame. Aggregated to the school level, the individual-student growth assessments could provide a better picture of whether the school is meeting the *ESEA* goals of progress toward universal proficiency in two respects. First, the growth models can identify some schools as making AYP that would not have made AYP under status-plus-safe-harbor alone. Second, some schools making AYP under status-plus-safe-harbor might not make AYP if growth criteria were applied instead. Schools making AYP under status may not make AYP under growth if sufficient numbers of proficient students were not improving their scores enough to stay on-track for their states' proficiency standards in subsequent years. Similarly, schools making AYP under safe-harbor may not have sufficient numbers of students on-track to proficiency to make AYP in the years to come. Moving to the theoretically more attractive model would thus result in some changes both among the set of AYP schools and the set of non-AYP schools.

Underlying the school-level AYP determinations analyzed in Chapters II and III are AYP determinations for each of the subgroups targeted by *ESEA*, insofar as the various subgroups are represented in the school. And underlying the subgroup determinations are results for each student that is a member of a subgroup in question. The growth model results for subgroups were not collected for the ED*Facts* repository, nor were any individual student data. However, the GMPP states were required to compile growth model results for all students in the participating grade levels and those data make it possible to calculate subgroup and school-level "growth only" results. Specifically, each of the pilot states included in this Interim Report provided student-level data with variables indicating (1) whether the student scored at the proficient level or higher on the state's reading and mathematics tests, and (2) whether the student was on-track to attain (or maintain) proficiency in the two subjects according to the growth model. The analyses presented in this section draw on the indicators of whether each student was on-track to proficiency in reading and mathematics. These data are used to calculate the proficiency level of all students in the school as well as each subgroup, and these rates were used to determine whether the school met the AMO by the growth criteria for 2006–07.

As discussed in Chapter I with respect to schools, it is theoretically possible for a student to have different proficiency determinations for growth and status. These possibilities are illustrated in Exhibit 25, which outlines the four possible patterns. A student can be proficient under status and growth (cell "A"), or can be proficient under growth but not status (cell "B"), or proficient under status but not growth (cell "C"), or not be proficient under status nor growth-only (cell "D"). The primary goal of the GMPP growth models is to identify students in cell B and to allow the schools to count those students the same as proficient students for accountability purposes.

Exhibit 25
Conceptual Map of How Growth-Only Proficiency Designations Compare With Status Model Proficiency Designations for Students

Student Proficiency Under the Growth-Only Model	Student Proficiency Under the Status Model	
	Proficient	Not Proficient
On-Track to Reach or Maintain Proficiency	A	B
Not On-Track to Reach or Maintain Proficiency	C	D

The percentages of students in each cell are shown in Exhibit 26. The percentages in cell A range from lows of 36 percent in Florida and 39 percent in Arkansas to 85 percent in Tennessee and more than two-thirds in Alaska, Delaware, and Iowa. The extent of overlap in the proficiency and on-track designations is measured by the sum of cells A and D. These sums ranged from a high of 94 percent in Iowa to a low of 75 percent in Delaware, with an average of 83 percent across the eight states. To the extent that a goal of the GMPP growth models is to provide information that goes beyond the proficiency designations under the status model, this high overlap, or redundancy in designations may represent a limitation of the growth models as currently allowed, particularly insofar as it entails relatively low rates of identifying non-proficient students who are on-track to proficiency. That is, the main expectation is that the growth model will identify some students in cell B (not proficient but on-track) and thus increase the percentage of students that contribute to positive AYP determinations (the sum of cells A and B). Comparing the percentages of students in cell B, the states ranged from highs of 25 percent of all students in Delaware and 20 percent in Alaska to lows of 3 percent in North Carolina and 5 percent in both Arkansas and Iowa. The variations among states in the cell B percentages are likely to reflect a combination of factors, including the effectiveness of the states' systems in moving below-proficient students to the on-track level as well as differences in the pool of eligible (non-proficient) students and the assessment instruments and standards.

It is clear that the percentages in cell B do not correlate with the percentages of schools that made AYP by growth documented in Chapter II. Alaska, for example, had no schools classified as making AYP by growth despite having a relatively high rate of non-proficient students identified as on-track to proficiency. The lack of correlation between the student and school results generally means the states differed in how the cell B students figured into the AYP determinations. For example, the pilot states differed in how they used safe-harbor provisions, such that the proportions of non-proficient but on-track students enrolled in safe-harbor schools may have varied from state to state.

Exhibit 26
Distribution of Students According to How Their Status and Growth Proficiency Classifications Compare, 2006–07

Pilot States	Proficient and On-Track (Cell A in Exhibit 25)	Not Proficient but On-Track (Cell B in Exhibit 25)	Proficient but Not On-Track (Cell C in Exhibit 25)	Neither Proficient nor On-Track (Cell D in Exhibit 25)	Number of Students
Alaska	67%	20%	1%	11%	63,984
Arizona	52%	10%	10%	27%	283,434
Arkansas	39%	5%	12%	44%	156,722
Delaware	68%	25%	N/A	7%	39,312
Florida	36%	8%	13%	43%	1,409,426
Iowa	69%	5%	N/A	25%	119,104
North Carolina	54%	3%	14%	30%	392,549
Tennessee	85%	7%	0%	7%	352,945

Exhibit reads: Of the 63,984 students in the state of Alaska, 67 percent were proficient and on-track to remain at or above the proficiency cut point, 20 percent were not proficient but were on-track to reach proficiency, one percent were proficient but not on-track to remain at or above the proficiency cut point, and 11 percent were neither proficient nor on-track to proficiency.

Source: U.S. Department of Education, ED*Facts*, Alaska Department of Education, Arizona Department of Education, Arkansas Department of Education, Florida Department of Education, Tennessee Department of Education.

Comparisons of status and growth-only proficiency determinations for schools

As discussed in Chapter I, the pilot states had the option to apply growth criteria to school AYP determinations in a number of ways: after status and safe-harbor, before status and safe-harbor, without status or safe-harbor, etc. Of the eight pilot states in 2006–07, all but Delaware elected to apply growth criteria after status and safe-harbor. Delaware, in contrast, applied growth criteria first, followed by status and then safe-harbor. If all three criteria (status, safe-harbor, and growth) are applied, the total number of schools making AYP is not affected by changes in the order in which the criteria are applied. Nonetheless, the quality of the information about the schools provided by the AYP data is arguably affected by the order of application, depending on the manner in which the AYP data are reported.

Theoretically, schools that make AYP by growth-only are more likely to attain or maintain AYP in the future than schools that make AYP by status but not growth, and also more likely to attain or maintain AYP in the future than schools that make AYP by safe-harbor but not by status or growth. If the AYP results are reported in a way that shows the criteria under which a school first achieves the AMO benchmark, applying growth criteria before status and safe-harbor could thus provide policymakers and the public with better information about school performance even though the overall numbers of schools making AYP would be the same.

The results in Chapter II indicated that relatively few additional schools made AYP by growth after status and safe-harbor criteria were applied. Referring to the cells in Exhibit 27, the analysis in Chapter II focused on the number of schools that made AYP by growth (cells C plus G). However, those results do not provide a comprehensive account of the growth model data

collected through the GMPP. First, it was possible for the schools that made AYP by status or by safe-harbor to also have had sufficient numbers of students classified by the growth model as on-track to proficiency to also meet their AMOs strictly under the growth criteria. Those schools would be located in cells A and B in Exhibit 27. Second, it was also possible for the schools that were classified as having made AYP by status-plus-growth to either meet (cell C) or not meet (cell G) the AMO if the students' on-track to proficiency data were exclusively used.

Exhibit 27
How the Growth Model On-Track-to-Proficiency Designations Can Compare with AYP Designations for Schools

School AMO Designations Under a Growth-Only Model	School AYP Designations in *ESEA* Reporting, Based on Status-Plus-Growth Determinations			
	Made AYP by Status	Made AYP by Safe-Harbor	Made AYP by Growth	Did not Make AYP
Met AMO with growth-only	A	B	C	D
Did not meet AMO with growth-only	E	F	G	H

As noted in Chapter I, the ED*Facts* data classified schools in terms of AYP and reporting groups in terms of AMO with sets of mutually exclusive categories that do not allow one to address hypothetical questions of what would happen if the student growth model results were used differently. In order to assess the possible magnitude of the cells in Exhibit 27, the best resource is the status and growth proficiency data provided in the student files compiled by each of the pilot states. These data do not provide sufficient information to implement the states' methodologies for determining whether schools made AYP, but they do contain information on the key dimensions of reading and mathematics proficiency and thus allow assessments of whether reporting groups and schools as wholes met their AMOs.

In Delaware and Iowa, all students who scored at proficiency or higher on the status criteria were also classified as "on-track" to proficiency per their growth models. The other six states, in contrast, applied their growth models to all students and identified some proficient students who were not on-track to continue to score at or above the proficiency levels in later grades.

The results shown in Exhibit 28 indicate that for all eight states combined, 85 percent of the schools that made AYP by status also met their reading and mathematics AMOs using just the growth criteria. The states varied considerably, ranging from only 45 percent in Arizona and 67 percent in Arkansas to over 90 percent in Alaska, Delaware, Florida, Iowa, North Carolina, and Tennessee.

Exhibit 28
Number and Distribution of Schools That Made AYP Based on Status Criteria, by Whether They Met AMOs Based on Growth Criteria, 2006–07

Pilot States	Number			Distribution	
	Number of Schools That Met AMO With Status Criteria	Met AMO With Growth-Only Results (Cell A in Exhibit 27)	AMO Not Met by Growth-Only (Cell E in Exhibit 27)	Met AMO With Growth-Only Results (Cell A in Exhibit 27)	AMO Not Met by Growth-Only (Cell E in Exhibit 27)
All Eight States	4,804	4,083	721	85%	15%
Alaska	319	301	18	94%	6%
Arizona	1,047	473	574	45%	55%
Arkansas	209	141	68	67%	33%
Delaware	96	96	0	100%	0
Florida	770	745	25	97%	3%
Iowa	780	780	0	100%	0
North Carolina	445	439	6	99%	1%
Tennessee	1,138	1,108	30	97%	3%

Exhibit reads: Of the 4,804 schools that made AYP based on status criteria, 4,083 schools (85 percent) met the state AMO under the growth criteria, while 721 schools (15 percent) did not.

Source: U.S. Department of Education, ED*Facts*, Alaska Department of Education, Arizona Department of Education, Arkansas Department of Education, Florida Department of Education, Iowa Department of Education, North Carolina Department of Education, and Tennessee Department of Education.

Taken at face value, these results indicate that very few schools that made AYP by status in Alaska, Delaware, Florida, Iowa, North Carolina, and Tennessee were not on-track to meet their AMO in the near term, whereas over half of the Arizona schools that made AYP by status were not on-track to maintain their AYP standing.

These findings provide some evidence that the low rates of making AYP by growth found in some states (most notably Alaska, North Carolina, and Tennessee) and documented in Chapter II were not the result of low rates of students being on-track to proficiency and schools not having sufficient growth to meet their AMOs using growth-only. This, in turn, suggests that the low rates of making AYP by growth in these states with high rates of meeting AMOs by growth-only was more a reflection of how their status models were designed rather than of an absence of growth. With status applied before growth, the large numbers of schools that could have met their AMOs by growth-only were obscured.

Growth-only proficiency results of schools making AYP by safe-harbor

The safe-harbor provision of *ESEA* was designed to identify schools making progress toward meeting the AMO even though their reading or mathematics proficiency levels, or both, are still below the AMO. Specifically, safe-harbor recognizes schools and subgroups that have decreased the number of students scoring below proficiency thresholds by 10 percent or more from one

year to the next. Identifying progress among low-performing schools is, of course, also the intention of the growth models, but the growth models are distinguished by (1) basing progress estimates on longitudinal student-level data, and (2) evaluating progress in terms of whether it is sufficient to reach the AMO within a specific time frame (e.g., by the time the student completes grade 8).

By virtue of both these distinguishing features, schools meeting their AMOs by growth should be performing better than schools meeting their AMOs by safe-harbor but not by growth. Nonetheless, all of the pilot states except Delaware applied safe-harbor provisions before growth criteria in their AYP determinations. In order to estimate the number of safe-harbor schools that also met their AMOs using just the growth criteria, we used the student-level data to categorize the schools.

The results in Exhibit 29 indicate that 87 percent of the safe-harbor schools in the six states that had any safe-harbor schools also met their AMO using the growth criteria. The percentages exceeded 70 percent in all states. Again, reasons for the differences among states will be explored in the final report, but the main point from this analysis is that the great majority of safe-harbor schools also met their AMOs by growth criteria.

Exhibit 29
Number and Distribution of Schools That Made AYP by Safe-Harbor Criteria, by Whether They Met AMOs Based on Growth Criteria, 2006–07

Pilot States	Number			Distribution	
	Number of Schools That Met AMO With Safe-Harbor	Met AMO With Growth-Only Results (Cell B in Exhibit 27)	AMO Not Met by Growth-Only (Cell F in Exhibit 27)	Met AMO With Growth-Only Results (Cell B in Exhibit 27)	AMO Not Met by Growth-Only (Cell F in Exhibit 27)
All Six States	901	785	116	87%	13%
Arkansas	285	233	52	82%	18%
Delaware	22	20	2	91%	9%
Florida	155	140	15	90%	10%
Iowa	105	76	29	72%	28%
North Carolina	288	285	3	99%	1%
Tennessee	68	51	17	75%	25%

Exhibit reads: Of the 901 schools in the six states that made AYP by safe-harbor, 785 schools (87 percent) met the state AMO under the growth criteria while 116 schools (13 percent) did not meet the state AMO under the growth criteria.

Note: No schools from Alaska or Arizona made AYP by safe-harbor in 2006–07.

Source: U.S. Department of Education, EDFacts., Arkansas Department of Education, Delaware Department of Education, Florida Department of Education, Iowa Department of Education, North Carolina Department of Education, and Tennessee Department of Education.

One implication of the overlap of safe-harbor and growth-only outcomes is that safe-harbor appears to work largely as intended, in that it identifies schools that are improving. However, the growth-only percent of students on-track could be a better gauge of actual test score improvement than safe-harbor. This is because the growth-only measure relies on the longitudinal student records rather than the percentages proficient in each year. A second implication is thus that growth criteria could be usefully applied to AYP determinations before safe-harbor is applied and the information about school improvement would be enhanced.

Demographic characteristics of schools that made their AMO by status and growth criteria

As shown in Chapter III, schools enrolling higher proportions of low-income and minority students were more likely than more affluent and lower-minority schools to make AYP by growth in the status-plus-safe-harbor-plus growth system. It is also interesting to examine the demographic characteristics of schools that reached their AMOs with both the status and growth criteria. These schools are theoretically the most likely to continue to reach the accountability milestones leading to universal proficiency by 2014.

To examine this relationship, the student data are used to determine both whether the school met the AMO by growth and by status. Using the student data for the status calculations has the advantage of clarifying the comparisons by removing schools that made AMOs by safe-harbor, confidence intervals, and multiyear averaging, all of which could produce different results than status-based AMO determinations built entirely from the student records.

The comparisons of status and growth proficiency results are restricted to six states because the Delaware and Iowa models defined all students proficient by status to also be on-track to proficiency by growth. The results in Exhibit 30 indicate that schools enrolling higher proportions of poverty students were less likely than schools with medium and low rates of poverty student enrollments to reach AMOs on both the status and growth criteria. The relationship of school poverty with meeting the AMOs on both criteria was stronger in Arizona and Arkansas than in Alaska, Florida, North Carolina, and Tennessee.

Exhibit 30
Percentage of Schools That Met the AMO Under Status That Also Met the AMO Under Growth-Only, by School Poverty Level, 2006–07

Pilot States	Number That Met AMO Under Status				Percent That Also Met AMO by Growth-Only			
	Low Poverty	Medium Poverty	High Poverty	All Schools	Low Poverty	Medium Poverty	High Poverty	All Schools
All Six States	1,203	2,371	571	4,145	80%	87%	71%	83%
Alaska	90	183	50	323	96%	95%	92%	94%
Arizona	478	444	174	1,096	56%	45%	31%	48%
Arkansas	24	167	21	212	63%	72%	43%	68%
Florida	314	409	47	770	97%	98%	85%	97%
North Carolina	153	440	38	631	100%	100%	92%	99%
Tennessee	144	728	241	1,113	97%	100%	93%	98%

Exhibit reads: Of the 1,203 schools in the six states that met the AMO under status that had low enrollments of children from poverty-level households, 80 percent also met the AMO under growth in 2006–07.

Source: U.S. Department of Education, ED*Facts*, Alaska Department of Education, Arizona Department of Education, Arkansas Department of Education, Florida Department of Education, North Carolina Department of Education, Tennessee Department of Education.

The relationship between minority enrollment and making AMO by both criteria (not tabulated) was in the same direction but somewhat weaker than the school poverty correlations. These results for high-poverty and high-minority schools show they were less likely to meet proficiency standards by both status and growth-only criteria. In contrast, the results in Chapter III for poverty and minority enrollments showed that the less-advantaged schools were more likely to make AYP by growth in the status-plus-growth framework generally implemented under the GMPP. An implication of these contrasting results is that if the growth-only method were adopted for actual accountability purposes while retaining the current core GMPP principles, the number of less-advantaged schools making AYP would decrease.

Extent of longitudinal matching

The final aspect of growth models examined here is the extent to which the state assessment systems were able to calculate growth indicators for their students. Growth indicators are generally based on at least two years of achievement data. However, some students in each school year will not have prior test score data because they transferred in from out of state or from a private school, were absent on the test dates, or were excluded from the prior year testing for one reason or another. The percentage of students with growth indicators is referred to as the match rate (all of the pilot states except Tennessee required two consecutive years of test scores in order to calculate growth indicators).

The match rate has potentially important implications for the validity of a growth model: if a growth model is applied to all students in an *ESEA* reporting group in order to assess whether the AMO was met, then only the matched students are used to determine proficiency rates. Among the eight states examined here, only Delaware, Florida, and Tennessee applied their growth models to all students in *ESEA* reporting groups. In Delaware, growth model results were

considered first in making AYP decisions for each *ESEA* reporting group. Florida and Tennessee applied growth results after status and safe-harbor but used the growth model results for *all students* within *ESEA* reporting groups that did not make AYP by status or safe-harbor. The other states, in contrast, only used growth results for non-proficient students within reporting groups that did not make AYP by growth or safe-harbor.

In order to gain some information on how the results of growth models like the ones piloted in Delaware, Florida, and Tennessee might be affected by match rates, matched and non-matched students were compared in five of the states (Alaska, Arkansas, Florida, Iowa, and Tennessee; Arizona, Delaware, and North Carolina provided data only on matched students). Exhibit 31 presents the match rates and average test score differences among matched and non-matched students in the four states that provided scale-score data to the evaluation project. The match rates are the percentages of students who have the two years (2005–06 and 2006–07 school years) of test scores needed to calculate growth.

The five states generally had match rates of 90 percent or higher based on the supplied student data (Exhibit 31). Tennessee's rate of 99 percent reflects the fact that the projection model it used is able to estimate projections even with missing data and almost all students thus have valid growth indicators.

Exhibit 31
Differences Between Matched and Unmatched Grade-Standardized Student Scores for Reading and Math for Alaska, Arkansas, Florida, Iowa, and Tennessee, 2006–07

Pilot States	N Students in Grades 4–8	Percent With Test Scores From 2005–06	Difference (SDs) in Reading	Difference (SDs) in Math
Alaska	41,881	93%	0.10	0.09
Arkansas	175,051	90%	0.41	0.30
Florida	924,485	92%	0.43	0.47
Iowa	119,182	100%	0.65	0.73
Tennessee[a]	356,496	99%		

Exhibit reads: Of the 41,881 students in Alaska, 93 percent had matched scores. The mean achievement difference in scale scores between those who were matched and those who were unmatched was 0.1 standard deviations in reading, and 0.09 standard deviations in mathematics.

[a] Tennessee did not provide the students' scale scores in reading and mathematics.

Source: Alaska Department of Education, Arizona Department of Education, Arkansas Department of Education, Florida Department of Education, Iowa Department of Education, and Tennessee Department of Education

In four of the states, reading and mathematics scale scores were available and the mean scores of matched and unmatched students could be compared. Unmatched students had substantially lower average test scores than matched students. Taken together, these results indicate that while the match rates were high, the matched population may have significantly higher levels of achievement than the unmatched group and thus be more likely to be proficient or on-track to proficiency. To that extent, reliance on growth-only data for AYP determinations may increase the likelihood a school will reach its AMO. While the match rates in the pilot states that provided the information were generally very high, the potential for a few exclusions to affect AYP results suggests that further study is needed to determine the validity of using non-matched status results rather than excluding them from the growth-only applications.

Conclusions

This chapter addressed a number of hypothetical questions about the pilot growth models. The first question addressed was the extent to which the results of using only the growth models' on-track-to-proficiency results would diverge from the status model results among the schools that made AYP by the status model in 2006–07. This analysis identified the schools currently performing sufficiently well to make AYP by status criteria but, according to the state's growth model, not making sufficient annual gains to reach the AMO by growth-only criteria. These schools would then be expected not to continue to make AYP if their pattern of low growth continued. For all eight states combined, 85 percent of the schools that made AYP by status also met their reading and mathematics AMOs using just the growth criteria. The states varied considerably, ranging from 45 percent in Arizona to over 90 percent in Alaska, Delaware, Florida, Iowa, and Tennessee.

A similar analysis of the schools that made AYP by safe-harbor found that overall 87 percent of the safe-harbor schools in the six states that had any safe-harbor schools would have also met their AMO using the growth-only criteria. The results were more consistent across states, with the great majority of safe-harbor schools in the pilot states also meeting their AMOs by growth criteria.

The relationship of school demographic characteristics with meeting AMOs both by status and growth-only criteria was examined to see whether divergence of the two indexes was concentrated in certain types of schools. The main finding here was that schools enrolling higher proportions of poverty students were less likely than schools with medium and low rates of poverty student enrollments to reach AMOs on both the status and growth-only criteria.

The final aspect of growth models examined here was match rates. In this chapter we estimated the extent to which the pilot states had reading and mathematics test scores from both 2006–07 and 2005–06 available, and whether the 2006–07 test scores differed for the students with and without the 2005–06 scores. The results indicate that while the match rates were high (90 percent or greater in most states), the matched population may have significantly higher levels of achievement than the unmatched group and thus be more likely to be proficient. However, it is not necessarily the case that reliance on growth-only data for AYP determinations would increase the likelihood a school will reach its AMO because the proficient students may not be on-track to maintain proficiency. But because proficient students are more likely than non-proficient students to be on-track, reliance on the matched cases would likely increase the odds

of reaching the AMO. If excluded students are concentrated in certain reporting groups, the impact of exclusions on school AYP determinations could be further increased. These points underscore the importance of maintaining high match rates if growth-only outcomes are applied to accountability decisions.

Conclusions

This interim report was designed to answer two main questions about the implementation of the Growth Model Pilot Project (GMPP) under the *Elementary and Secondary Education Act* as amended by the *No Child Left Behind Act of 2001*. The two questions and brief summaries of the study results based on analyses of data provided by eight pilot grantee states from 2006–07 that are the subject of this evaluation (Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, and Tennessee) are as follows:

How have states in the pilot project implemented growth models?

While the models approved for the eight pilot grantee states that are the subject of this evaluation differ from one another in a number of important ways, all use state-specific assessment data to measure student progress and proficiency, and the method of incorporating growth outcomes in adequate yearly progress (AYP) determinations was generally the same. Seven of the eight pilot states applied growth criteria only *after* schools have failed to make AYP under the status and safe-harbor provisions, rather than determining AYP solely on the basis of student improvement—a "status plus growth" model. Delaware, in contrast, applied growth criteria first, and then applied the status model and safe-harbor provisions, respectively, to schools that did not make AYP under the growth model.

When a school is designated as "making AYP under growth," this means that use of the growth model changed the designation for one or more targeted *ESEA* subgroups. Within the affected subgroup(s) for a given school, the growth criterion is usually applied only to the students who did not achieve at or above the proficiency level. Simply stated, "making AYP under growth," as defined by the GMPP, does not mean that all students are on-track to proficiency, and it can even mean only one non-proficient student is on-track if a sufficient number of others in the subgroup are proficient. Seven of the eight states whose 2006–07 school year data were available for analyses used the status-plus-growth model (Alaska, Arizona, Arkansas, Florida, Iowa, North Carolina, and Tennessee). Delaware used a growth-plus-status model for determining AYP.

The models approved for the pilot study vary in how they (1) establish growth expectations for students, and (2) determine whether individual students are "on-track" to reach proficiency in the allotted time frame. We have identified three basic types of growth models being used in the GMPP: the *transition matrix* model (which evaluates student progress from year to year in terms of a relatively small set of discrete performance levels), the *trajectory* model (which uses the gap between a baseline test score and a performance standard several years out to calculate the amount of growth required to become proficient), and the *projection* model (which uses current and past test scores to statistically predict performance several years ahead, using new test scores to update projections of student performance).

There was little evidence that the type of model selected had an impact on the extent to which schools were identified as making AYP by growth within the GMPP framework. Further analyses of the strengths and weaknesses of the various models are planned for the final report, including side-by-side comparisons of generic versions of the three types of models using a common dataset. However, the results presented here suggest that impacts of differences in the

ways the models identify students as on-track to proficiency or not may be reduced by (a) the constraints of the GMPP guiding principles and particularly the requirement of universal proficiency by 2014, and (b) the various provisions within *ESEA* for recognizing AYP under the status model. The latter are generally designed to reduce the chance of incorrectly identifying schools as not making AYP, and the growth models were primarily adopted to further reduce that chance. But if the status provisions were already capturing most of the schools that might have been identified as making AYP by growth, then the numbers of schools identified as growth schools by the GMPP would be correspondingly reduced.

How did each pilot state's growth model affect the number and kinds of schools that make AYP?

The eight pilot states in 2006–07 that are the subject of this evaluation provided both school-level (ED*Facts*) and the student-level data to the evaluation project and these data were used to address five subquestions.

1. ***How many schools made AYP under the growth model that would not have made it under the* ESEA *status model?***

   The designs of the growth models in these states included only those students who (a) did not reach proficiency levels in reading or language arts and mathematics, and (b) were members of *ESEA* reporting groups that did not reach their Annual Measurable Objectives (AMOs) or obtain AYP via safe-harbor provisions. The pilot models simply added growth criteria to the traditional status plus safe-harbor model for determining AYP; thus growth criteria could only increase the number of schools making AYP. The number of schools identified as making AYP by growth under the GMPP in 2006–07 ranged from 1 percent or fewer of all schools in Alaska, Arizona, North Carolina, and Tennessee, to 3 percent in Delaware, 5 percent in Florida, 8 percent in Arkansas, and 11 percent of all schools in Iowa. The schools making AYP uniquely by growth represented percentage increases in the numbers of schools making AYP that ranged from highs of 16 percent in Florida, 14 percent in Arkansas, and 12 percent in Iowa to lows of 2 percent or fewer in Alaska, Arizona, North Carolina, and Tennessee. Expressed as percentages of the schools that did not make AYP under status or safe-harbor, the impact of the GMPP on identifying additional schools as making AYP ranged as high as 17 percent in Arkansas and 69 percent in Iowa.

2. ***Are AYP outcomes under the growth models related to school demographics and organizational characteristics?***

   Schools enrolling higher proportions of low-income and minority students were more likely to make AYP under growth in the status-plus-growth framework than were schools enrolling higher proportions of more affluent and nonminority students. However, if growth were the sole criterion for determining AYP, schools enrolling higher proportions of low-income and minority students would be more likely to move from making AYP to not making AYP.

3. ***How many schools that made AYP under the* ESEA *status model would also have made it if the growth criteria were used exclusively for assessing whether schools met their AMOs?***

   This question was addressed by drawing on the student data provided by the pilot states to calculate whether each school met its AMOs for reading and mathematics using the on-track to proficiency indicators for each student. For all eight states combined, 85 percent of the

schools that made AYP by status also met their reading and mathematics AMOs using just the growth criteria. The states varied considerably, ranging from 45 percent in Arizona and 67 percent in Arkansas to over 90 percent in Alaska, Delaware, Florida, Iowa, North Carolina, and Tennessee.

4. ***What is the relationship between AYP status under the growth model and* ESEA *safe-harbor provisions?***

   Six of the eight pilot states had at least one school classified in ED*Facts* as having made AYP by safe-harbor provisions. Overall, 87 percent of the schools making AYP by safe-harbor provisions in the six states also met their states' AMOs using the growth model criteria. Arkansas and Iowa each had about 80 percent of the total number of safe-harbor schools make AYP by growth. Delaware and Florida had roughly 90 percent of the total number of safe-harbor schools that also met their AMO using the growth criteria.

5. ***To what extent are longitudinal student data required by the growth models available?***

   We estimated the extent to which the pilot states had reading and mathematics test scores from both 2006–07 and 2005–06 available, and whether the 2006–07 test scores differed for the students with and without the 2005–06 scores. While the match rates were high (90 percent or greater in all states), the matched population may have significantly higher levels of achievement than the unmatched group and thus be more likely to be proficient or on-track to proficiency. To that extent, reliance on growth data with their requirement for matched longitudinal records for AYP determinations may increase the likelihood a school will reach its AMO.

Implications for future policy

As used for school AYP purposes in all states except Delaware, the growth information for students was used only for reporting groups that did not make AYP under status criteria or safe-harbor provisions. Consequently, it was generally not possible to determine from the official ED*Facts* reports the extent to which a school designated as making AYP by growth was actually realizing growth among its students. Delaware applied growth criteria before status and safe-harbor criteria and all schools identified as making AYP by growth actually met their AMOs on the basis of Delaware's growth model. An implication for future policy is that other states could clarify their schools' progress by applying growth criteria before status and safe-harbor. Two general ways of doing that can be identified and each has advantages and disadvantages. Both of these options are methods for *reporting* of schools' results related to AYP determinations, and do not suggest changes that would actually affect whether schools are determined to make AYP. As described here, neither of the methods would affect a school's AYP status.

1. ***Applying growth criteria before status and safe-harbor criteria to all schools' AYP determinations***

A different possible ordering would involve first using the growth results to assess whether the reporting group reached the AMO, then, if not, using the union of growth and status results to estimate the percentage of students on-track or proficient. Safe-harbor might only then apply to reporting groups that did not reach the AMO through either of these calculations.

The main advantage of applying growth before status and safe-harbor is that it would identify schools that are realizing adequate progress toward universal proficiency. This would clearly distinguish those schools from schools making AYP under status criteria but not realizing growth sufficient to continue meeting their AMOs. The latter are probably not a large number but identifying such schools would serve as an early warning mechanism of possible problems. The exploratory analyses in this report also indicate that applying growth criteria before safe-harbor could usefully reclassify most of the current safe-harbor schools as making AYP by growth and would clearly identify the minority that are not on-track to proficiency and thus headed for sanctions in the near future.

One disadvantage is that schools that are making AYP both by status and growth (arguably the strongest schools) would not be uniquely identified. Another possible disadvantage of applying growth first follows from the fact that the AYP-by-growth determination would be based on students with at least two consecutive years of test data. Matched students are likely to have higher test scores and thus make it more likely that their subgroups and school reaches the AMO. While the match rates in the pilot states that provided the information were generally very high, the potential of a few exclusions to affect AYP results suggests that states consider using non-matched students' status results rather than excluding them from otherwise growth-only applications.

2. Applying both growth and status criteria to all schools' AYP determinations

The GMPP required states to compile student-level data on both status and growth criteria, and these data could be used to classify each school simultaneously on both criteria for AYP-reporting purposes. That is, each school could be classified as making AYP by both growth and status, by growth-only, by status-only, by a mix of status and growth, or by safe-harbor only. This would have the advantage of uniquely identifying different sets of schools (those making AYP in terms of both growth and status). A possible hierarchy for reporting purposes would be:

- Met AMO by growth and by status (best)
- Met AMO by growth but not status
- Met AMO by status but not growth
- Met AMO by combining status and growth results, so that status results are used for students not on-track to proficiency
- Did not make AMO (worst)

A disadvantage is that some states (e.g., Delaware and Iowa) classified all students who were proficient using the status criterion as also being on-track to proficiency on the growth criterion, regardless of their prior year scores. This would have the effect of increasing the number of "met by growth and status" schools and would prevent comparisons with states that measured growth for proficient and non-proficient students.

References

Booher-Jennings , J. (2005). "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal*, 42: 231–268.

Chudowsky, N., Chudowsky, V., and Kober, N. (2009). *Is the Emphasis on "Proficiency" Shortchanging Higher- and Lower-Achieving Students?* Center for Education Policy, Washington, D.C. Retrieved July 2009 from http://www.cep-dc.org/index.cfm?fuseaction=document_ext.showDocumentByID&nodeID=1&DocumentID=280.

Council of Chief State School Officers (CCSSO) (2008). *Implementer's Guide to Growth.* Washington, D.C. Retrieved May 2008 from http://www.ccsso.org/content/pdfs/IGG%20Final%20AP.pdf.

Neal, D., and Schantenback, D. (2008). *Left Behind By Design: Proficiency Counts and Test-Based Accountability.* Retrieved September 2008 from http://derek.a.neal.googlepages.com/restat_rev_final_200809.pdf.

U.S. Department of Education, Office of Elementary and Secondary Education (Jan. 15, 2009). E*valuation of the 2005-06 Growth Model Pilot Program*. Washington, D.C. Retrieved Feb. 2009 from http://www.ed.gov/admins/lead/account/growthmodel/gmeval0109.doc.

U.S. Department of Education, Office of Elementary and Secondary Education (July 2007). *No Child Left Behind* Fact Sheet: "Growth Models: Ensuring Grade-Level Proficiency for All Students by 2014." Washington, D.C. Retrieved May 2008 from http://www.ed.gov/admins/lead/account/growthmodel/proficiency.pdf.

U.S. Department of Education (Jan. 25, 2006)."Peer Review Guidance for the *NCLB* Growth Model Pilot Applications." Washington, D.C. Retrieved May 2008 from http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf.

U.S. Department of Education, Office of the Secretary, Office of Public Affairs (2003). *No Child Left Behind: A Parent's Guide*. Washington, D.C. Retrieved May 2008 from http://www.ed.gov/parents/academic/involve/nclbguide/parentsguide.pdf.

U.S. Department of Education (2008). "U.S. Secretary of Education Margaret Spellings Approves Additional Growth Model Pilots for 2007–2008 School Year", press release, June 10, 2008. Retrieved June 2008 from http://www.ed.gov/news/pressreleases/2008/06/06102008.html.

U.S. Department of Education, Office of Communications and Outreach (2009). "U.S. Secretary of Education Margaret Spellings Approves Additional Growth Model Pilots for 2008–2009 School Year", press release, Jan. 8, 2009. Retrieved January 2009 from http://www.ed.gov/news/pressreleases/2009/01/01082009a.html.

U.S. Department of Education, Office of Communications and Outreach (2006). "Secretary Spellings Approves Additional Growth Model Pilots for 2006–2007", press release, Nov. 9, 2006. Retrieved June 2008 from http://www.ed.gov/news/pressreleases/2006/11/11092006a.html.

U.S. Department of Education, Office of Communications and Outreach (2005). "Secretary Spellings Announces Growth Model Pilot, Addresses Chief State School Officers' Annual Policy Forum in Richmond", press release, Nov. 18, 2005. Retrieved May 2008 from http://www.ed.gov/news/pressreleases/2005/11/11182005.html.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service (2007). *State and Local Implementation of the No Child Left Behind Act, Volume III—Accountability Under NCLB: Interim Report*, by K. LeFloch, F. Martinez, J. O'Day, B. Stecher, J. Taylor, and A. Cook. Washington, D.C.

Appendix A.
Comparison of GMPP Growth Models with State Accountability Systems

Of the eight states approved for the GMPP for the 2006–07 school year, four already had implemented formal measures of student growth for state accountability purposes before their GMPP models were approved. A fifth state included growth as an optional measure, while the remaining two had begun developing growth models before joining the GMPP (one of which has since been implemented). In comparison, a 2007 survey by the Council of Chief State School Officer's (CCSSO) Accountability Systems and Reporting State Collaborative on Assessment and Student Standards (ASR-SCASS) found that only six of the 23 responding states had fully operational non-AYP growth models and fewer than half calculated any form of growth at all.[25] This suggests that states with growth modeling experience were better positioned to apply for and receive approval to participate in the federal program. In fact, the two states first approved for GMPP had the most prior experience using student progress to grade schools. Exhibit A.1 provides brief descriptions of each state's pre-GMPP growth model.

[25] See http://edmeasure.com/ASR/doku.php?id=asr:survey_results (retrieved May 2008). A GAO survey in March 2006 found that 26 of 50 states were using growth models, of which seven measured individual student growth (GAO–06–948T). Three of those states are in the GMPP, and our review of documentation suggests that three other GMPP states were using (or about to use) student growth information for state accountability purposes.

Exhibit A.1 Overview of Growth Components of State Accountability Systems

State	Non-AYP Growth Model	Year Implemented	Grades and Subjects Covered	Growth Standards Applied	School Progress Measured	Other Features
Alaska	School Performance Incentive Program	2007–08	3–10 Reading/Writing/ Math	Growth is measured using seven categories of performance	Points are awarded for student movement among performance categories and are averaged to create a school index score	Four performance levels are used to reward schools according to their growth index scores
Arizona	Measure of Academic Progress (MAP)	1999–2000, revised 2005–06	4–8 Reading/Math	Growth is measured using annual differences in scale scores and are averaged to calculate school growth	Points are awarded for meeting expected growth (by quartile) and are added to other measures of school performance (AZ LEARNS)	Expected growth is calculated using a value-added model adjusted to ensure proficiency by seventh grade and controlling for mobility/ceiling effects
Arkansas	Act 35 Annual School Ratings System	2003–04, growth added in 2008–09	3–11 Reading/Writing/ Math	Student growth is based on changes in eight performance levels across adjacent years.	Value points are added if a student moves up a performance category and are lost if a student moves down a category	School are evaluated using five performance levels for both "improvement" (Category One) and "status" (Category Two)
Delaware	n/a	n/a	n/a	n/a	n/a	n/a
Florida	Grading Florida Public Schools	2001–02	3–10 Reading/Math/ Science/Writing	Growth is measured by movement among five performance levels	Grades schools by awarding points for the number of students meeting proficiency standards plus those making annual learning gains	Points also are awarded for students who maintain high performance, for the percent of students in the lowest quartile making gains, and if half or more of such students make gains
Iowa	*Optional*	n/a	K–12 Reading/Math/ Social Studies/ Science	Growth for student profiles is measured using changes in ITBS and ITED scores	n/a	n/a

Exhibit A.1 continues next page

Exhibit A.1 Overview of Growth Components of State Accountability Systems continued from previous page

State	Non-AYP Growth Model	Year Implemented	Grades and Subjects Covered	Growth Standards Applied	School Progress Measured	Other Features
North Carolina	ABCs Growth Model	1996–97, revised 2005–06	K–8, HS Reading/Math/ Science	Growth is the average of standardized scores from two previous years adjusted for regression to the mean and is expected to be "0" or above, indicating at least one year's worth of gain	A mean growth score is calculated for the school, though schools can also meet expected growth if 60 percent or more of students test at grade-level proficiency	Schools get credit for "high growth" if students who are stable or growing outnumber those declining by at least three to two [or a ratio of 3:2]
Ohio	Value-Added Assessment	2006–07	4–8, plans for K–10 Reading/Math	Growth is measured as mean NCE gains and is categorized as meeting or as being a standard error "above" or "below" the "expected" gain	LRC designations can be raised if schools are above expected growth for at least two years and lowered if below expectations for at least three years	Value-added is based on the 2006–07 distribution of test scores and will be used for accountability beginning in 2007–08
Tennessee	Tennessee Value-Added Assessment System (TVAAS)	1992–93	3–8 Reading/Math/ Language/ Science/Social Science	Growth in TCAP results is measured as mean NCE gains relative to a state growth standard (1998 Terra Nova) or a state three-year average	School reports include average student gains by grade and whether achievement level meet growth standards or are within one, two, or more standard deviations	Specific school effects also are estimated using up to five years of student data to predict scores for the average student in a given school; teacher effect data is restricted access

However, these state accountability growth models had to be modified in significant ways for use in GMPP. One basic difference between models approved under *ESEA* and those used for state accountability purposes is the definition of expected growth. The seven core principles require that the *ESEA* growth models place limits on the amount of time students have to reach proficiency, usually by the last grade tested as applied, resulting in growth expectations that put a student "on-track" to meet grade-level standards in a few years. Most state growth models do not require students to be on-track to proficiency but, for example, simply expect students to exhibit a year's worth of growth per year of instruction. The principle of universal proficiency by 2014 also translates into rising AMOs, a requirement that is absent from state accountability systems. The goal of states in most cases is to determine whether individual schools and the system as a whole are improving rather than to see if a given student will be proficient in a few years or if a school will have all students proficient by 2014.

Another major difference is that growth components of state accountability systems typically calculate an average for the entire school instead of basing school performance on the results of each subgroup separately, and may average over subjects as well. This means that a school can still make adequate growth for state accountability purposes even if one or more subgroups fails to grow or if the school exhibits strong growth in one subject but not another. Of course this does not mean that growth results for subgroups and subjects are not reported, and states often measure growth for more than reading and math. The goal is to create a composite growth score that can then be combined with other measures of school performance (including AYP) both for reporting purposes, such as a "grade" on school report cards, and for rewarding schools that meet or exceed state standards.

The result of these differences is that more schools tend to make adequate growth for state accountability purposes than would make AYP under the GMPP. For example, 779 of the 1,298 (or 60 percent) of the North Carolina schools that failed to make AYP in 2006–07 made expected growth under the state guidelines. Of these, 187 (or 14 percent) met the state standards for high academic growth. The federal growth model used by North Carolina sets the bar higher for both expected growth and percent proficient (AMO) than does the growth model used for state accountability, suggesting that the universal proficiency requirement is driving most if not all of this difference. Florida also drops the proficiency deadline in its grading system, although the transition matrix model used sets expectations that may differ from a single year's worth of growth.[26] The school itself can make adequate growth if half or more of low achievers meet expected growth, which is a different standard than the AMO used for AYP purposes. Thus about 84 percent of Florida schools made adequate growth in both reading and math for 2006–07, compared to the 34 percent of schools that made AYP under the GMPP.

Offsetting this tendency of the 2014 deadline to lower the number of schools making adequate growth is that state accountability systems do not give schools credit if enough students grow to push a subgroup or school over the proficiency threshold as is done in the status-plus-growth models implemented under the GMPP.[27] This means that schools classified as making AYP by

[26] Note that students who score in the two lowest FCAT achievement levels still can make expected growth if they make a year's worth of growth as determined by FCAT developmental scores.

[27] Note that states often give growth credit to schools for *maintaining* proficiency, which is similar to a growth-plus-status model, and may grade schools on proficiency status criteria as well.

growth may not be showing much growth for the student body as a whole. Thus of the 12 North Carolina schools that made AYP by growth in 2006–07, only three met the state's high growth standard and one did not make the lower standard of expected growth. Similarly, 10 of the 19 Tennessee schools that made AYP by growth in 2006–07 grew at a rate below the state three-year average for math, while 14 schools grew at a below average rate for reading. One school even experienced negative growth for both reading and math. Growth models are only one component of federal accountability under the GMPP and thus do not provide much information about whether a school is growing as currently reported. However, results from growth measures used for state accountability purposes suggest that many more schools would make AYP if the first of the seven core principles of the *ESEA* project was relaxed.

Appendix B: State GMPP Model Summaries

The core of each growth model is the method of determining whether a student is "on-track" to being proficient. While states' methods can be categorized as one of three types of growth models (trajectory, transition matrix, or projection), each state's method has unique features. This appendix offers additional details about the growth models used in Alaska, Arizona, Arkansas, Delaware, Florida, North Carolina, and Tennessee. All the relevant detail of Iowa's growth model is included in the main text.

Alaska

Alaska uses a trajectory model that calculates growth for students in grades 4 through 6, grade 8 and grade 9, using results of the Alaska Standards Based Assessment Test mathematics and language arts.

Proficiency levels for these tests are scaled so that thresholds are 300 for math and 600 for literacy within each grade. The literacy score is cutoff at 600 because it is a combination of reading and writing, each of which has a cutoff of 300 points. For this summary we will consider a generic cutoff score of 300 for each grade.

Students are always classified by the traditional status model in grades 3 and 7. In grades 4 through 6 it is possible to use the growth model. In the Alaska growth model the first year a student is below proficient is considered their base year, and growth targets for subsequent years are calculated by evenly dividing the difference between the base year score and the cutoff in grade 7. In grades 8 and 9, the first year a student is below proficient is considered the base year, and growth targets for subsequent years are calculated by evenly dividing the difference between the base year score and the cutoff in grade 10.

For example, if a student scores below proficient in some grade q (between grades 3 and 6), growth targets for each subsequent grade is a function of their base score in grade q, Y_q, plus the difference between this score and 300 weighted fraction of the years that have elapsed between the previous grade and grade 7. Thus, a student in grades 4 through 6 is considered to be on-track toward proficiency in grade k if their score, Y_k, is greater than or equal to this function, $Y_k \geq Y_q + \left(300 - Y_q\right) \times \left(\left(k-q\right)/\left(7-q\right)\right)$, assuming k > q and $Y_q < 300$.

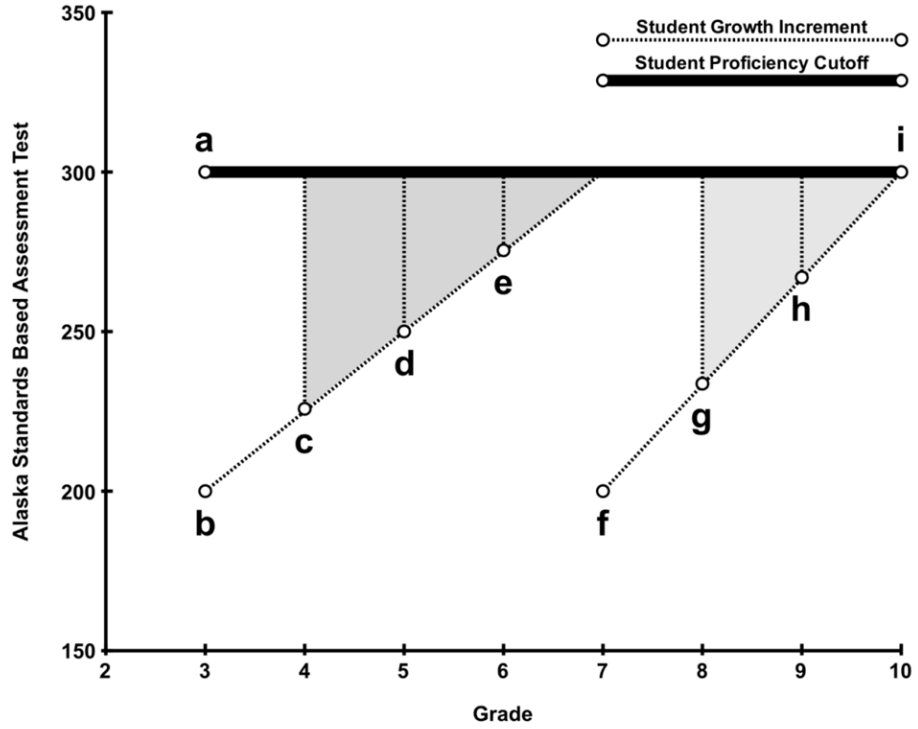
For students who score below proficient in some grade q between grades 7 and 10, growth targets for each subsequent grade is a similar function of their base score in grade q, Y_q, plus the difference between this score and 300 weighted fraction of the years that have elapsed between the previous grade and grade 10. Thus, a student in grades 8 or 9 is considered to be on-track toward proficiency in grade k if their score, Y_k, is greater than or equal to this function, or,

$Y_k \geq Y_q + \left(300 - Y_q\right) \times \left(\left(k-q\right)/\left(10-q\right)\right)$, assuming k > q and $Y_q < 300$.

These targets, once established by the baseline year, remain the growth targets for subsequent years.

The illustration in exhibit B.1 below showcases this process. This figure shows the proficiency standard for each grade (300) along the solid line connecting points a to i. The shaded polygons illustrate the necessary growth between years to be on-track toward proficiency for a set of hypothetical students.

Exhibit B.1
Illustration of Alaska's Method for Determining Whether a Student Is On-Track Toward Proficiency



Label	a	b	c	d	e	f	g	h	i
Grade	3	3	4	5	6	7	8	9	10
Score	300	200	226	250	275	200	234	267	300

Student Z is below proficient in the third grade, he scored 200 (point b) instead of the proficiency standard of 300. In applying the formulas above, Alaska's growth model would set a new target score of 226 in the fourth grade (point c), 250 in grade 5 (point d), and 275 in grade 6 (point e), to be considered on-track for those grades. While that student cannot be used for growth model AMO calculations in the third grade, that student will be expected to make those targets calculated for subsequent years to be considered as on-track in grades 4 through 6. Similarly, student X is below proficient in the seventh grade, he scored 200 (point f) instead of the proficiency standard of 300. In applying the formulas above, Alaska's growth model would set new target scores of 234 in the eighth grade (point g), and 264 in grade 9 (point h), to be considered on-track for those grades. While that student cannot be used for growth model AMO calculations in the seventh grade, that student will be expected to make those targets calculated for subsequent years to be considered as on-track in grades 8 and 9.

Arizona

Arizona uses a trajectory model to set growth targets for students in grades 3 through 8 using Arizona's Instrument to Measure Standards (AIMS) test. The score cutoffs for reading and math are presented in Exhibit B.2.

Arizona, like other states using trajectory models, sets growth targets for students who are not proficient based on a baseline score. Growth targets are set by determining the necessary improvement to reach proficiency and setting evenly spaced improvement targets over a given set of years. In Arizona, growth targets are set based on the first non-proficient score. Growth targets calculated from a baseline score in K through third grade require the student to reach proficiency by the sixth grade. Growth targets calculated based on a fourth grade score require the student to be proficient by seventh grade. Growth targets calculated based on scores in grades 5, 6, and 7 require the student to be proficient by the eighth grade. Growth targets are reset each year based on student performance and movement in and out of the Arizona public school system. However, the specific grade required to reach proficiency does not change when targets are reset.

Exhibit B.2
Arizona Proficiency Standards Test Cutoffs for Grades 3 Through 8

Grade	Reading Cutoff	Math Cutoff
3	431	420
4	450	448
5	468	476
6	478	496
7	489	517
8	499	537

While each student is assigned growth targets based on actual performance, Arizona is unique compared to other growth model pilot states in that schools are evaluated on each student's estimated score instead of students' actual score. However, in calculating the proportion of students within a school who are considered to be on-track toward proficiency, students are classified based on the lower bound of their predicted scores. These predicted scores are generated from a statewide statistical model which fit current year scores to scores from the previous years.

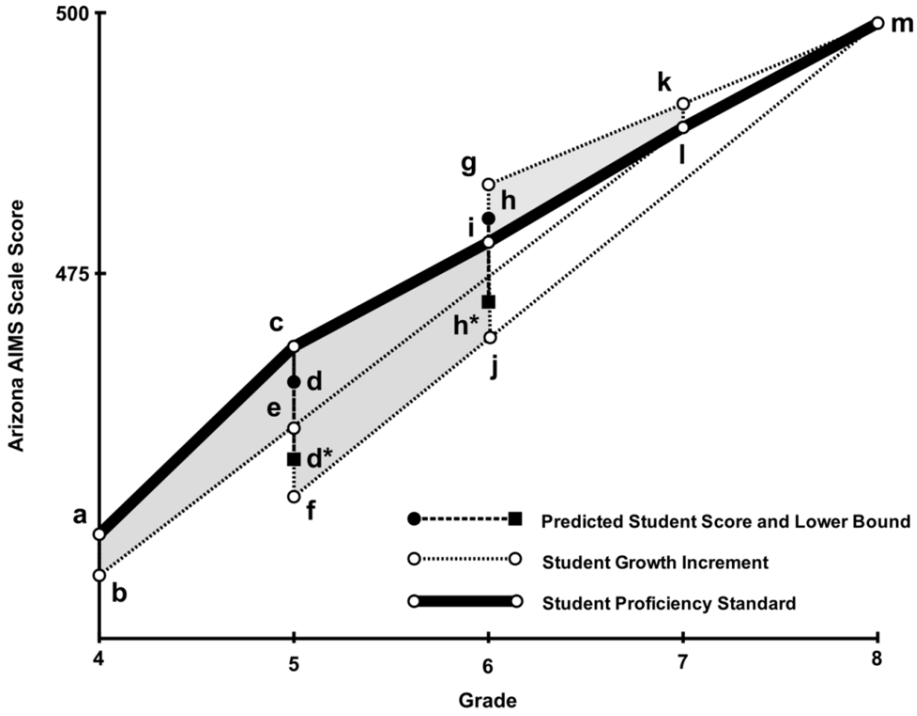
Arizona estimates a regression model predicting the current year scores of all students by their previous year's score. Let Y_{it} be the score from student i in grade k in school j at the current year t, $Y_{it} = \alpha + \beta Y_{it-1} + e_i$. From this model, Arizona predicts each student's current score, $Y^*_{it} = \alpha + \beta Y_{it-1}$. There is a statewide standard error, $SE_{Y^*_t}$, associated with each year's predicted score, which typically equals about 4 or 5 scale score points. The score used to evaluates a student's growth, Y^*_{it}, is then a lower bound of a confidence interval estimated around the predicted score, or $Y^{Lower}_{it} = Y^*_{it} - \left(1.96 \times SE_{Y^*_t}\right)$. As a result, each student is evaluated not on their actual score, but on this lower bound of an estimated predicted score.

The illustration in exhibit B.3 below showcases this process. This figure shows the proficiency standards for grades 4, 5, 6, 7, and 8 along the solid line connecting points a, c, i, l, and m. The shaded polygons illustrate the necessary growth between years to be on-track toward proficiency for a of hypothetical student.

Student Z is below proficient in the fourth grade, he scored 446 (point b) instead of the proficiency standard of 450 (point a). Arizona's growth model would set a new target score of 460 in the fifth grade (point e) to be considered on-track. Student Z then scored exactly 454 in the fifth grade (point f), which is below the necessary growth target. However, the student predicted score is 465 (point d), which is above the growth target. Yet, the school will be unable to count that student as on-track because the lower bound of the predicted confidence interval is 457 (point d*), still better than the actual score, but also still below the growth target. The student actual fifth grade score is then used to estimate a new growth target of 469 in the sixth grade (point j).

Student Z then did well in the sixth grade, scoring 484 (point g), well above the normal cutoff of 478 (point i). However, that student's predicted score is 480 (point h) with a lower bound 472 (point h*). Because the lower bound is still greater than the growth target, that student can then be counted as on-track. Under a growth-only model, that student's growth target for seventh grade would be higher than the normal cutoff (point k instead of point l) because that student's actual sixth-grade score was greater than the normal cutoff as well.

Exhibit B.3
Illustration of Arizona's Method for Determining Whether a Student Is On-Track Toward Proficiency



Label	a	b	c	e	d	d*	f	g	h	h*	i	j	k	l	m
Grade	4	4	5	5	5	5	5	6	6	6	6	6	7	7	8
Score	450	446	468	460	465	457	454	484	480	472	478	469	491	489	499

Arkansas

Arkansas uses a trajectory model that calculates growth for students in grades 4 through 7 using results of the Arkansas Benchmark Exams for mathematics and literacy, which are administered in grades 3 through 8 (plus grade 11 for literacy).

Proficiency levels for these vertically scaled exams are set for each grade, and growth targets are based on the annual exam score increment needed to reach the proficiency standard in eighth grade. These proficiency standards are presented in table B.4 below.

Exhibit B.4 Arkansas Benchmark Exam Proficiency Standards for Grades 3 Through 8	
Grade	**Proficiency Standard**
3	500
4	559
5	604
6	641
7	673
8	700

If a student does not achieve proficiency in a base year, the growth targets necessary to be considered on-track toward proficiency are calculated based on the total increment between a student's base score and the proficiency standard in eighth grade (700). However, instead of setting growth targets by evenly dividing the total increment between the base score and 700, the increments are proportioned so that the growth increments follow a curve that matches the concave nature of the typical proficiency standards.

This means that, for example, instead of setting the growth increment from third to fourth grade to be a quarter (0.25) of the total necessary improvement to reach 700 by eighth grade, the growth increment for fourth grade is set to slightly more than a quarter (0.295) of the total necessary improvement to reach 700 by eighth grade. Thus, the annual increment that a student must attain in order to be classified as on-track to proficiency is calculated using grade-specific growth target multipliers. The multiplier for any grade "k" is simply a ratio of the difference between the current and previous year's proficiency standards, P_k and P_{k-1}, to the difference between the proficiency standard for eighth grade and the previous year, or,

$$\left(P_k - P_{k-1}\right) / \left(700 - P_{k-1}\right).$$

These multipliers are presented in Exhibit B.5.

Exhibit B.5 Arkansas Growth Target Multipliers for Grades 3 Through 7	
Grade	**Multiplier**
4	0.295
5	0.319
6	0.385
7	0.542

Another method of illustrating a specific growth increment is to use the following formula that calculates the minimum score necessary to be considered on-track toward proficiency. For any student in grade k, the minimum score necessary to be considered "on-track" is his or her previous score, $Y_{k-1,}$, plus the difference between 700 and his or her previous score, $700\text{-}Y_{k-1}$,

times the multipliers outlined above. Thus, if the current score, Y_k, is greater than or equal to this function, or,

$$Y_k \geq Y_{k-1} + \left(\left(\left(P_k - P_{k-1} \right) / \left(700 - P_{k-1} \right) \right) \times \left(700 - Y_{k-1} \right) \right),$$

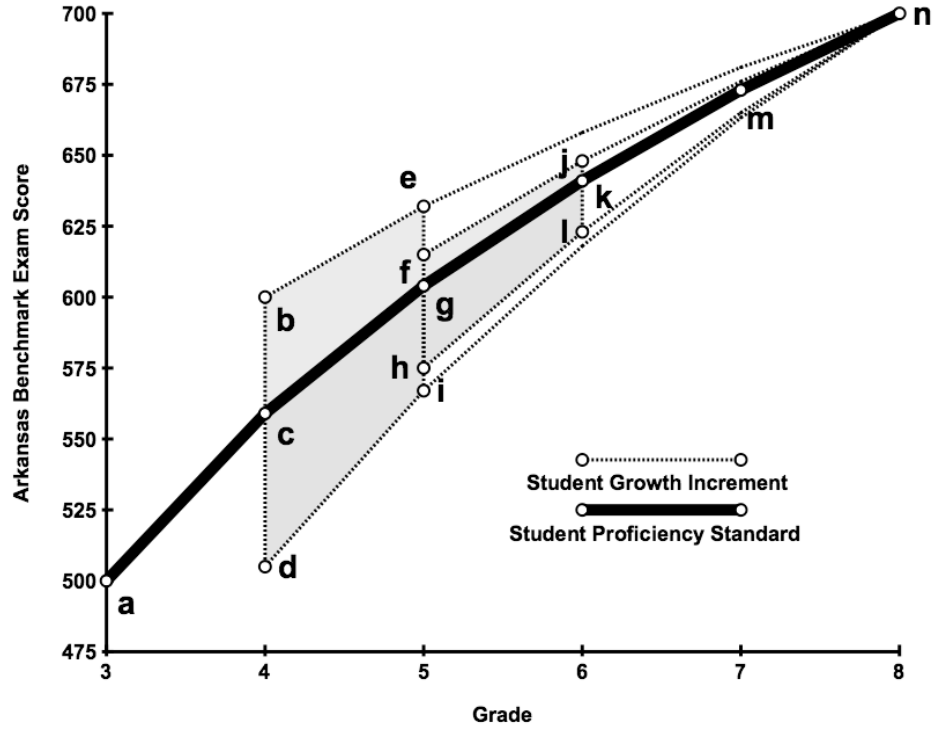
that student is considered to be on-track toward proficiency. From this formula, we see that this model resets the growth target every year rather than setting a series of annual targets based on the first below-proficient exam score. This formula can also determine minimum scores even for students who score above the proficiency threshold. Thus, students can be identified that are not improving toward the eighth grade standard of 700 even though they are currently proficient.

The illustration in exhibit B.3 below showcases each of these features. This figure shows the proficiency standards for each grade along the solid line connecting points a, c, g, k, m and n. The shaded polygons illustrate the necessary growth between years to be on-track toward proficiency for a set of hypothetical students.

Student Z is below proficient in the fourth grade, he scored 505 (point d) instead of the proficiency standard of 559 (point c). In applying the formulas above, Arkansas' growth model would set a new target score of 567 in the fifth grade (point i) to be considered on-track. If student Z scored exactly 567 in the fifth grade, then his growth targets would follow the dotted line from point i to point n. In addition, imagine that student Z in fact did better than scoring 567 in the fifth grade—let us say that he scored 575 (point h). Then, that student's growth target for sixth grade would be reset and instead would be 623 (point l). Note that because student Z did better than his growth target in the fifth grade, his new growth target for the sixth grade is slightly higher—point l is above the dotted line from point i to point n.

Another student, student X, scored well above the standard in fourth grade at 600 (point b). By applying the same formula, the Arkansas growth model would set a fifth-grade target score of 632 (point e), which is likewise well above the proficiency standard of 604 (point g). Next, imagine that student X scored lower than his target score in fifth grade—let us say he scored 615 (point f). Student X scored above proficient in fifth grade, but because is score is lower than his target score generated from fourth grade, he is not considered to be on-track toward proficiency. The Arkansas model does reset scores each year, and so we can see that his target for sixth grade is set at 648 (point j), which is closer to the proficiency standard of 641 (point k). Note that his sixth-grade target is below the dotted line from point e to point n. This shows that his new target score is reflecting that he did not score as high in the fifth grade as he did in the fourth grade. He would need to score at least 691 in the seventh grade (point k), which is more than the normal seventh-grade proficiency standard of 673 (point l).

Exhibit B.6
Illustration of Arkansas' Method for Determining Whether a Student Is On-Track Toward Proficiency



Label	a	b	c	d	e	f	g	h	i	j	k	l	m	n
Grade	3	4	4	4	5	5	5	5	5	6	6	6	7	8
Score	500	600	559	505	632	615	604	575	567	648	641	623	673	700

Delaware

Delaware's transition matrix model uses a "value table" method for AYP determinations that assigns points for students depending on the type and extent of changes between the performance levels (see Exhibit 6 in the report). Points in the value table increase with both the magnitude of change in level of proficiency, and the overall level of proficiency. As a result, a student scoring toward the bottom during his previous grade receives fewer points for moving up one level than he would if he moved up several levels. All students who surpass their grade level cutoff receive the maximum points regardless of how high they scored or whether their underlying scores actually declined from the prior year. From these points assigned to students, schools meet their AMO if the average number of points per student meets or exceeds a target number of points set by Delaware for particular subjects and grades.

Delaware's point system is designed to be an analogue to the AMO system used for status models by giving partial weight to students who are below proficient but still making progress. In fact, the target number of points that Delaware sets is simply the standard AMO that is rescaled to the point system.

For example, in the 2006–2007 school year the status model AMO for reading was 68 percent proficient. For that same year, the target number of points that a school's students needed to gain on average was 204. If we divide the target points by the maximum number of points a student can receive, we return to the AMO (204/300 = 0.68). To illustrate how a school meeting the AMO under the status model would perform under the growth model, imagine two schools with 100 students, school Z and school X (see exhibit B.7 below)

In the first school, school Z, exactly 68 out of 100 students (68 percent) were proficient in reading. That school meets the AMO of 68 percent exactly. Let us assume that in this school all students who were below proficient did not move up any categories from their previous grade. Then, under Delaware's growth model, that school would receive

$$68 \times 300 = 20400 \text{ points,}$$

out of a total possible

$$100 \times 300 = 30000 \text{ points.}$$

The average number of points per student would be

$$\frac{20400}{100} = 204,$$

which is exactly the target number of points a school needs to make their AMO under the growth model, and the number of achieved points, 20,400, is exactly 68 percent (the AMO) of the total possible 30,000 points. Thus school Z would make their AMO under both the status model and the growth model.

To illustrate how a school that would not meet its AMO under the status model, but would make its AMO under the growth model, consider another school X which also has 100 students, but only 30 students are proficient. Under the status model, only 30 percent of the students are proficient, which is well below the target AMO of 68 percent. Yet, the school does have many students who are improving. Under the growth model, the proficient students account for only

$$30 \times 300 = 9000 \text{ points,}$$

which is only 30 percent of the total possible 30,000 points. Remember, though, school X has students who are improving, including 30 students who moved from PL 1A to PL 1B for

$$30 \times 150 = 4500 \text{ points,}$$

and another 15 students moved from PL 1B to PL 2A for

$$15 \times 175 = 2625 \text{ points,}$$

and another 15 students moved from PL 1B to PL 2B for

$$15 \times 225 = 3375 \text{ points,}$$

and 5 students who did not improve, who stayed at PL1A for

$$5 \times 0 = 0 \text{ points.}$$

That school would total 20,500 points out of a possible 30,000 points. That would be 68.3 percent of the total possible points, greater than the AMO for reading that year. The average number of points per student would be 205, which is larger than the 204 points necessary to meet the AMO. Under the growth model, school X would meet the AMO.

This means that Delaware's status model is actually like a simple transition matrix in which all students who are proficient are given the maximum number of points, and all student who are below proficient are given no points at all. The growth model, then, gives students who improve a partial number of points. The table in Exhibit B.7 can be used to estimate the proportion of the total points each improvement is worth. Another way to think about Delaware's growth model is that a student who moves from the first level, PL 1A, to the second level, PL 1B, is given half (0.50) the weight of a student who is proficient when the percent proficient is calculated. This means that in Delaware's transition matrix system, two students who move from the first to the second level equal one student who is proficient.

Exhibit B.7
Example Comparison of Two Hypothetical Schools in Delaware to Illustrate Growth Model

				School Z		School X	
Year 1	Year 2	Points Per Student	Proportion of Total Possible Points	Number of Students Per Cell	Points Per Cell	Number of Students Per Cell	Points Per Cell
PL 1A	PL 1A	0	0.00	0	0	5	0
PL 1A	PL 1B	150	0.50	0	0	30	4,500
PL 1A	PL 2A	225	0.75	0	0	0	0
PL 1A	PL 2B	250	0.83	0	0	0	0
PL 1A	Proficient	300	1.00	0	0	0	0
PL 1B	PL 1A	0	0.00	0	0	0	0
PL 1B	PL 1B	0	0.00	0	0	0	0
PL 1B	PL 2A	175	0.58	0	0	15	2,625
PL 1B	PL 2B	225	0.75	0	0	15	3,375
PL 1B	Proficient	300	1.00	0	0	0	0
PL 2A	PL 1A	0	0.00	0	0	0	0
PL 2A	PL 1B	0	0.00	0	0	0	0
PL 2A	PL 2A	0	0.00	0	0	0	0
PL 2A	PL 2B	200	0.67	0	0	5	1,000
PL 2A	Proficient	300	1.00	0	0	0	0
PL 2B	PL 1A	0	0.00	0	0	0	0
PL 2B	PL 1B	0	0.00	0	0	0	0
PL 2B	PL 2A	0	0.00	0	0	0	0
PL 2B	PL 2B	0	0.00	32	0	0	0
PL 2B	Proficient	300	1.00	0	0	0	0
Proficient	PL 1A	0	0.00	0	0	0	0
Proficient	PL 1B	0	0.00	0	0	0	0
Proficient	PL 2A	0	0.00	0	0	0	0
Proficient	PL 2B	0	0.00	0	0	0	0
Proficient	Proficient	300	1.00	68	20,400	30	9,000
Total				100	20,400	100	20,500
Percent Proficient Year 2				68%		30%	
Average Points					204		205
Meet AMO by Status (68% proficient)?				Yes		No	
Meet AMO by Growth (204 Average Points)?				Yes		Yes	

Florida

Florida's growth model applies to students in grades 3 through 10 using the Developmental Scale Scores (DSS) from the Florida Comprehensive Assessment Tests (FCAT) for mathematics and reading. The state uses a trajectory model that bases growth targets on the score required for proficiency three years after the first year tested (usually, grade 3). To be counted as proficient, a student who was not proficient at the baseline must close the gap between his baseline score and the proficiency cutoff three grades later by one-third each year. This means that a student can only make proficiency by growth for two years because the gap must be fully closed in the third year, (i.e., a student must meet or exceed the minimum proficiency score in year 3). A student first enrolled in grade 9 must close the gap by half to be counted as "on-track." The cutoffs for grades 3 through 7 are presented in Exhibit B.8.

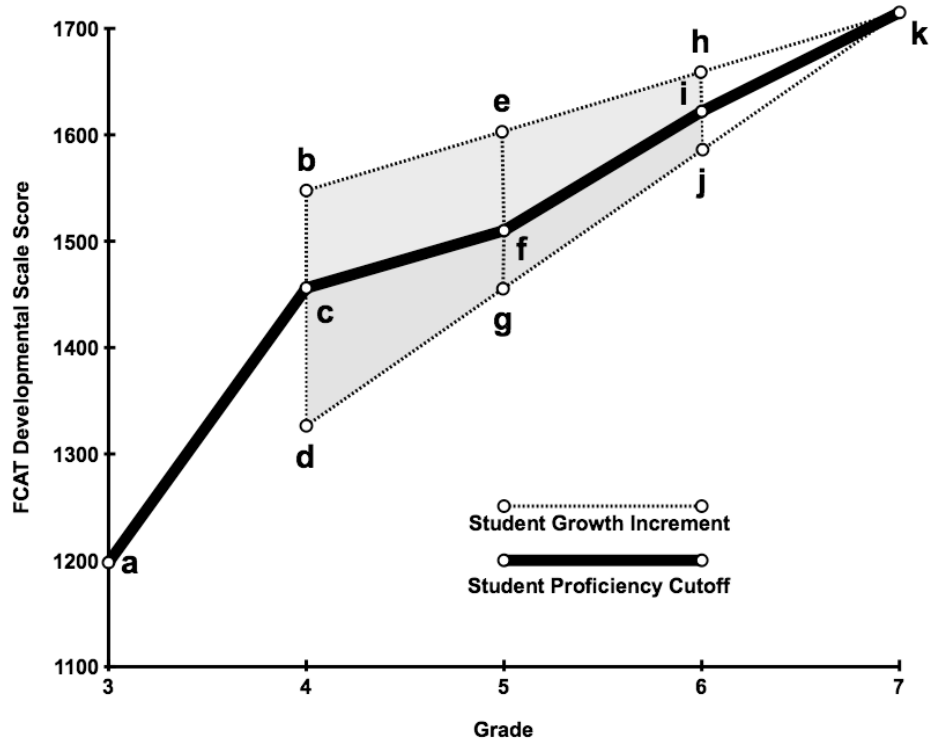
Exhibit B.8
Florida's Cutoff Developmental Scale Scores to Be Considered Proficient on the FCAT for Grades 3 Through 7

Grade	DSS Cutoffs for Reading	DSS Cutoffs for Math
3	1198	1269
4	1456	1444
5	1510	1632
6	1622	1692
7	1715	1786

The illustration in exhibit B.9 below showcases this process for reading. The bold line connecting points a, c, f, i, and k represent the FCAT DSS score cutoffs for reading for grades 3 through 7. The shaded polygons represent the growth increments necessary to be considered on-track toward proficiency. To illustrate the properties of Florida's growth model, let us revisit student Z and student X.

Student Z is below proficient. He scored 1326 in fourth grade (point d), which is below the grade-specific cutoff of 1456. Florida's growth model would then set growth targets for the following two years, fifth grade and sixth grade, by dividing the difference between his fourth-grade score and the seventh-grade cutoff into thirds. In other words, he would be expected to close the gap between his fourth-grade score and the seventh-grade cut-off by a third during the first year after he was not proficient (fifth grade). In turn, he would be expected to close the gap by two-thirds after two years of not being proficient (sixth grade). He would then be expected to be proficient after three years (seventh grade). Geometrically, this is drawing a straight line from point d at grade 4 to point k at grade 7. This line then sets growth targets at point g and point j for grades 5 and 6, respectively. Thus, the growth target for student Z in fifth grade is 1455 (point g) and 1586 in sixth grade (point j).

Exhibit B.9
Illustration of Florida's Method for Determining Whether a Student Is On-Track Toward Proficiency in Reading



Label	a	b	c	d	e	f	g	h	i	j	k
Grade	3	4	4	4	5	5	5	6	6	6	7
Score	1198	1548	1456	1326	1603	1510	1455	1659	1622	1586	1715

Unlike Arkansas' trajectory model, Florida's method is not intentionally designed as a method for tracking the progress of proficient students. Florida does not scale growth targets to match the pattern of cutoff scores. As a result, this method is very sensitive to the pattern of cutoffs set for each grade. This means that if this method were used to draw growth targets for proficient students, then the magnitude of the higher expectations would fluctuate each grade, depending on grade-by-grade standards. Furthermore, the problem of how far above the cutoff a growth target would be for any one grade would greatly depend on *which* grade was used for the base year.

For example, imagine that student X scored 1548 (point b) in the fourth grade, which is 92 points above the cutoff of 1456 (point c). Let us further assume that for some reason, that score was used at a baseline for future growth targets. As a result, that student's growth targets would be represented by the line from point b to point k. By following that line we see that student X would have to score 1603 points to be considered on-track (point e), which is 93 more points than the cutoff of 1510 in the fifth grade (point f). He would not be expected to score as many more points in sixth grade, however, because the cutoff scores increase by a large amount between the fifth and sixth grades. In the sixth grade, student X would have to score 1659 (point

h), or only 37 more points than the cutoff of 1622 (point i). This illustrates that students who score above proficient would be subject to inconsistent expectations because the pattern of cutoff scores in Florida is not linear, and the model is not designed to mimic the pattern of standards from grade to grade or reset expectations.

North Carolina

North Carolina uses a trajectory model to identify growth targets and both End-of-Grade (EOG) tests and End-of-Course (EOC) tests for assessing student progress. The state uses a Standardized Scale Approach (SSA) to growth which uses the normative distribution of student performance in the standard setting year of any test edition as a common basis to build a scale. State documents note that this approach is useful for measuring the growth in student performance from one year to the next and also adapts well to the changes in curriculum and subsequent changes in test editions.

The SSA system uses a time-locked modified z-scale termed a "change scale" or "c-scale." Thus, 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.

The 2005–06 school year was the standard setting year for the Mathematics EOG tests at grades 3–8 and the 2002–03 school year was the standard setting year for the Reading EOG tests in grades 3–8. NC performs an equating study to set the achievement level cut scores at the same time the c-scale is built.

The trajectory is built based on the student's performance either the previous year, or on the third- grade pretest, whichever is appropriate to the grade in which the student first enters the state. Therefore, the following table illustrates the basis for prediction, the targeted test for proficiency, the years of trajectory, and the percent of difference between baseline performance and proficiency expected by the trajectory based on the year the student is first enrolled in the state in a tested grade.

Exhibit B.10 Grades and Tests Used for Trajectory Growth in North Carolina and the Percent of Difference Expected to Be Closed Per Year

Grade of First Enrollment	Test Used as the Basis for Prediction	Test Used as Target for Proficiency	Years to Proficiency	Percent of Difference Closed Per Step	Steps to Proficiency
3	3rd grade pretest	6th grade EOG	4	25%	4
4	4th grade EOG	7th grade EOG	4	33%	3
5	5th grade EOG	8th grade EOG	4	33%	3
6	6th grade EOG	Algebra I or English I EOC	4	33%	3
7	7th grade EOG	Algebra I or English I EOC	4	50%	2
8	8th grade EOG	Algebra I or English I EOC	3	100%	1

The trajectories are built individually by student and separately for reading or mathematics. Therefore, a student will have a trajectory based on their baseline mathematics score and the proficiency cut score for mathematics separate from reading. In the upper grades, Algebra I is the AYP assessment for 10th-grade students and is the trajectory target for math while English I is the trajectory target for reading or language arts.

The following table displays the performance expected of students to be counted as on trajectory for inclusion in the proposed method of comparing school performance to AMO targets.

For a student who enters in third grade and has a grade 3 pretest:

Year in State-Tested Grade	Decrease From Baseline Assessment in Performance Discrepancy
1	25% of Original Gap
2	50% of Original Gap
3	75% of Original Gap
4 or more	Student Must Be Proficient

For a student who enters in fourth, fifth, or sixth grade:

Year in State-Tested Grade	Decrease From Baseline Assessment in Performance Discrepancy
1	Baseline, Not On Trajectory
2	33% of Original Gap
3	66% of Original Gap
4 or more	Student Must Be Proficient

Therefore, if a subgroup has met its 95 percent participation target but has not met its proficiency target, and the subgroup has met its other academic indicator, the process of incorporating the growth measure would be:

1) First identify if the student has been in membership the full academic year and is both tested and not proficient.
2) These three conditions being met, the number of years the student has been in the state will be determined using the historic files from the state's accountability system.
3) If the student has been in the state (in a tested grade) for four years or more, the student will remain non-proficient for comparison to the annual measurable objectives (AMO). If the student has been in the state public schools three years or less, the correct baseline score will be located (using the table above).
4) The student's performance on the baseline assessment in the subject of interest will be converted to the c-scale.
5) Based on the student's baseline score and proficiency in the target year, a difference will be calculated.
6) The decrease in the difference will be compared against Table 4 above based on the number of years in the tested grades in North Carolina.

7) If the student's performance on the current assessment is equal to or better than the minimum from the previous step, include the student in the percent proficient calculation to compare against the state's AMOs.

To illustrate, assume a student enters North Carolina in the fourth grade. The student scores below proficient in the current school year in reading. This child's known test scores are listed below.

Grade	3 EOG	4 EOG	5 EOG
Developmental Score	Not in NC	229	241
C-scale score		-2.68	-1.98

Because the student's first full year in the state is the fourth-grade year, the student will need to be on trajectory to be proficient by the end of the seventh grade and thus on the seventh-grade EOG for reading. The developmental score for seventh grade reading equivalent to proficient is 252. The associated c-scale score is -1.00.

Because the student was not in the state for the third-grade test, the fourth-grade EOG score will be used as the baseline. The difference between the baseline and proficient on the seventh grade test in terms of c-scale scores is 1.68 (difference between 2.68 and 1.00). For the current year (fifth grade, the second year in the state), the student must perform well enough on the test to have 33 percent less difference between the c-scale score for proficiency and his baseline (fourth- grade EOG) c-scale score (divide 1.68 by 3 = 0.56).

For this to be true, the child would need to score at least -2.12 (difference between 2.68 and 0.56). The child's actual c-scale score is -1.98 which means the child met the standard to be deemed on trajectory for the current year and thus will be included in the percent of students on trajectory or proficient for comparison to the AMO for the school as a whole and any subgroups to which the child may belong.

Tennessee

Adapted from "Evaluation of the 2005–06 Growth Model Pilot Program," U.S. Department of Education, 2009.

Tennessee uses a projection model to assess student growth. The state uses a student's history of test scores in an equation to project or predict that student's future score. To complete this process, previous cohorts of student scores are used in generating a prediction equation that can be applied to the current cohort of students. For example, last year a cohort of sixth-graders in Tennessee were tested on the state reading exam. These same sixth-graders also had scores from the state reading exam for grades 3, 4, and 5. The scores on the sixth-grade reading test are placed in a matrix called **Y**. The reading scores for grades 3–5 are placed in a matrix called **X**. All the reading scores from grades 3–6 are combined into a design matrix called **XY**. The matrices are used in a statistical procedure to generate a covariance matrix called **C** with submatrices $\mathbf{C_{XX}}$ and $\mathbf{C_{XY}}$ ($\mathbf{C_{YX}} = \mathbf{C_{XY}}^{\mathbf{T}}$) and $\mathbf{C_{YY}}$. These submatrices are used for various statistical functions but primarily in calculations for $\mathbf{b} = \mathbf{C_{XX}}^{-1}\mathbf{C_{YX}}$ to generate the regression

coefficients b_1, b_2, ... b_N. For example, the projected score is computed using variations of the following equation:

$$\text{Projected Score} = M_Y + b_1(X_1 - M_1) + b_2(X_2 - M_2) + \ldots + = M_Y + \mathbf{x}_i^T\mathbf{b}$$

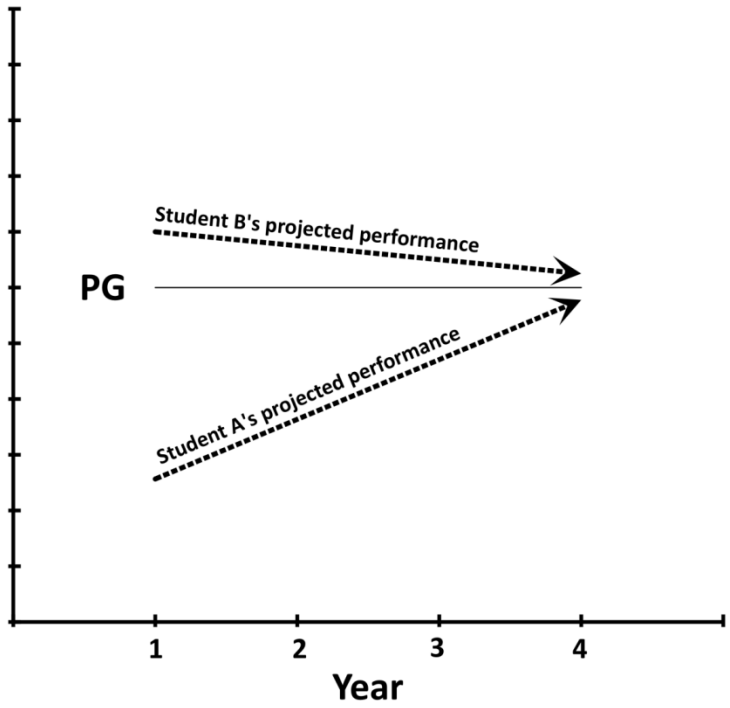
where M_Y, M_1, etc. are estimated mean scores for the "future test score" or response variable (Y). The previous test scores can also be referred to as the "predictor variables." To complete projected scores in the equation you make the following substitutions:

M_Y = estimated mean score on test
b_1, b_2, ... b_N = regression coefficients used to predict performance
X_1, X_2, ... X_N = previous reading scores
M_1, M_2, ... M_N = average school reading scores

The Tennessee model includes a statewide "average schooling effect," which is obtained by calculating the mean scores for each grade of a particular school and then averaging those means over all schools within the state. It is intended to account for the fact that a current school has no control over the effectiveness of the schools that their students will attend in the future (thus potentially affecting the student's growth). The average schooling effect assumes that each student will have the "average schooling experience" of all Tennessee schools.

Tennessee's model projects scores for all students, estimating each student's performance in reading and math in three years. Each student's projection is based upon his or her available test scores. For example, student A has reading scores for 2003, 2004, and 2005 whereas student B has reading scores for 2003 and 2005. In both cases, projected scores are computed using the equation described above using student-specific values.

Exhibit B.11
Illustration of Tennessee's Method for Determining Whether a Student Is On-Track Toward Proficiency in Reading



In Exhibit B.11, student A is currently below proficient (year 1) but projected to be proficient in year 4. In Tennessee's growth model, student A would be considered proficient in year 1 for AYP determinations and in each of the succeeding years if he or she continues to be on this trajectory to proficient.

Student B is missing data for 2004. Tennessee addresses the issue of missing test scores by using the regression coefficients from ($b = C_{XX}^{-1}C_{YX}$) and constants in the equation described above to fill in the missing test scores. Thus, all relevant student data are included in projecting score for a student. Using the available data for student B, a growth trajectory is developed that shows a projected decline in performance, though the student is projected to remain above proficient. Because student B is projected to remain above proficient by year 4, student B is proficient for the current year AYP determinations. If student B's projection indicated he or she would fall below proficient, student B would be considered non-proficient in the current year AYP determinations.

The Department of Education's mission is to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access.

www.ed.gov