

July 2006

NO CHILD LEFT BEHIND ACT

States Face
Challenges Measuring
Academic Growth
That Education's
Initiatives May Help
Address



GAO
Accountability · Integrity · Reliability

Highlights

Highlights of [GAO-06-661](#), a report to congressional requesters

Why GAO Did This Study

The No Child Left Behind Act (NCLBA) requires that states improve academic performance so that all students reach proficiency in reading and math by 2014 and that achievement gaps close among student groups. States set annual proficiency targets using an approach known as a status model, which calculates test scores 1 year at a time. Some states have interest in using growth models that measure changes in test scores over time to determine if schools are meeting proficiency targets.

To determine the extent that growth models were consistent with NCLBA's goals, GAO assessed (1) the extent that states have used growth models to measure academic achievement, (2) the extent that growth models can measure progress in achieving key NCLBA goals, and (3) the challenges states may face in using growth models to meet adequate yearly progress (AYP) requirements and how the Department of Education (Education) is assisting the states. To obtain this information, we conducted a national survey and site visits to 4 states.

While growth models are typically defined as tracking the same students over time, GAO used a definition that also included tracking schools and groups of students. In comments, Education said that this definition could be confusing. GAO used this definition of growth to reflect the variety of approaches states were taking.

www.gao.gov/cgi-bin/getrpt?GAO-06-661.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Marnie S. Shaul (202) 512-7215 or shaulm@gao.gov.

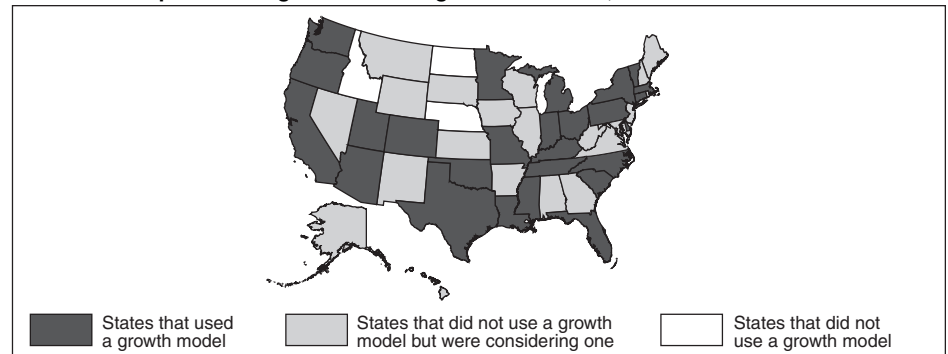
NO CHILD LEFT BEHIND ACT

States Face Challenges Measuring Academic Growth That Education's Initiatives May Help Address

What GAO Found

Twenty-six states were using growth models, and another 22 were considering or in the process of implementing growth models, as of March 2006. States were using or considering growth models in addition to status models to measure academic performance and for other purposes. Seventeen states were using growth models prior to NCLBA. Most states using growth models measured progress for schools and for student groups, and 7 also measured growth for individual students. States used growth models to target resources for students that need extra help or award teachers bonuses based on their school's performance.

States That Reported Using or Considering Growth Models, as of March 2006



Source: GAO analysis.

Certain growth models can measure progress in achieving key NCLBA goals. If states were allowed to use these models to determine AYP, they might reduce the number of lower-performing schools identified for improvement while allowing states to concentrate federal dollars in the lowest-performing schools. Massachusetts sets growth targets for schools and their student groups and allows them to make AYP if they meet these targets, even if they do not achieve state-wide goals. Some lower-performing schools may meet early growth targets but not improve quickly enough for all students to be proficient by 2014. If these schools make AYP by showing growth, their students may not benefit from improvement actions provided for in the law.

States face challenges measuring academic growth—such as creating data and assessment systems to support growth models—that Education's initiatives may help address. The ability of states to use growth models to make AYP determinations depends on the complexity of the model they choose and the extent that their existing data systems meet requirements of their model. Education initiated data grants to support state efforts to track individual test scores over time. Education also started a pilot project for up to 10 states to use growth models that met the department's specific criteria to determine AYP. Education chose North Carolina and Tennessee out of 20 states that applied. With its pilot project, Education may gain valuable information on whether growth models overstate progress or appropriately credit improving schools.

Contents

| | | |
|-----------------------------|--|-----------|
| Letter | | 1 |
| | Results in Brief | 3 |
| | Background | 5 |
| | Nearly All States Were Using or Considering Growth Models to Supplement Other Measures of Academic Performance | 12 |
| | Certain Growth Models Provide Useful Information on the Extent That Schools Are Achieving Key NCLBA Goals but May Overlook Some Low-Performing Schools if Used for AYP | 19 |
| | States Face Challenges In Measuring Year-to-Year Growth That Education’s Pilot Program and Data System Grants May Help Address | 27 |
| | Concluding Observations | 33 |
| | Agency Comments and Our Evaluation | 33 |
| Appendix I | Scope and Methodology | 35 |
| Appendix II | Selected Data from GAO’s Survey of States’ Use or Consideration of Growth Models | 38 |
| Appendix III | Comments from the Department of Education | 45 |
| Appendix IV | GAO Contact and Staff Acknowledgments | 47 |
| Related GAO Products | | 48 |

Tables

| | |
|--|----|
| Table 1: Types of Growth Models and States Using Them, as of March 2006 | 14 |
| Table 2: How a Status Model and Certain Growth Models Measure Progress in Achieving Key NCLBA Goals | 20 |
| Table 3: Criteria for Participating in Education’s Growth Model Pilot Project | 30 |
| Table 4: States Submitting a Proposal for Growth Model Pilot or Awarded a Grant for a Longitudinal Data System | 32 |
| Table 5: Grades of Growth Model Reporting, by State | 39 |
| Table 6: Level of Growth Model Reporting, by State | 40 |
| Table 7: Measure of Achievement in Growth Models | 41 |
| Table 8: Characteristics of Assessments Used in Growth Models | 43 |

Figures

| | |
|---|----|
| Figure 1: Hypothetical Example of Annual Proficiency Targets Set under a Status Model | 7 |
| Figure 2: Hypothetical Example of Closing Achievement Gaps | 8 |
| Figure 3: Safe Harbor | 9 |
| Figure 4: Results from Two Hypothetical Schools, Shown with a Status Model and a Growth Model | 11 |
| Figure 5: Map of States That Reported Using or Considering Growth Models, as of March 2006 | 13 |
| Figure 6: Illustration of School-Level Growth | 15 |
| Figure 7: Example of Measurement of Above Average Growth for a Fourth-Grade Student under Tennessee’s Model | 16 |
| Figure 8: States’ Introduction of Growth Models | 17 |
| Figure 9: Targets for a Selected School in Massachusetts Compared to State Status Model Targets | 21 |
| Figure 10: Results for a Selected School in Massachusetts in Mathematics | 22 |
| Figure 11: Results for Selected Students in Mathematics from a School in Tennessee | 24 |

Abbreviations

| | |
|-------|----------------------------------|
| AYP | adequate yearly progress |
| NCLBA | No Child Left Behind Act of 2001 |

This is a work of the U.S. government and is not subject to copyright protection in the United States. It may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



United States Government Accountability Office
Washington, DC 20548

July 17, 2006

The Honorable Howard P. “Buck” McKeon
Chairman
The Honorable George Miller
Ranking Minority Member
Committee on Education and the Workforce
House of Representatives

The Honorable Michael N. Castle
Chairman
Subcommittee on Education Reform
Committee on Education and the Workforce
House of Representatives

The Honorable John A. Boehner
House of Representatives

The Honorable Sam Graves
House of Representatives

The nation’s economic prosperity and global competitiveness depend in large part on the effective education of the 48 million students who attend public schools. Congress passed the No Child Left Behind Act of 2001 (NCLBA) requiring states to steadily improve academic performance so that, at a minimum, all students are proficient, that is, able to read and do math at grade level. Among the law’s principal goals are that all students are proficient by 2014 and achievement gaps close between high- and low-performing students, especially those in designated groups such as economically disadvantaged students.

With these two key goals in mind, states were required to set challenging standards for both academic content and achievement and to determine whether schools made adequate yearly progress (AYP) toward meeting those standards. For a school to make AYP, it must meet or exceed the state’s annual proficiency targets and meet or demonstrate progress on a target on another measure—graduation rates in high school or attendance or other measures in elementary and middle schools. If schools do not meet these requirements, their students may be eligible to receive tutoring or transfer to another school.

States set proficiency targets using status models that calculate the percentage of students with test scores that meet or exceed these targets 1 year at a time. With status models, states or districts determine whether schools make AYP based on annual performance while generally not taking into account how much better or worse the school did as compared to the previous year. Thus, a school that is showing large increases in student achievement but has too few students at the proficient level would not likely make AYP. Further, status models do not account for the fact that student characteristics in a school may change from one year to the next and these changes can affect whether a school makes AYP.

Because of the limitations of status models, some states have expressed interest in determining AYP by using growth models that measure year-to-year progress in proficiency. “Growth models” is a term that refers to a variety of methods of tracking changes in proficiency levels or test scores over time. One type of model, known as an improvement model, measures year-to-year school growth, but does not account for the fact that different students constitute a school from one year to the next. Because of this, some researchers do not consider it to be a growth model. In this report, we included improvement models as a type of growth model in order to provide a broad assessment of options that may be available for states. Growth models vary in complexity, such as calculating annual progress in a school’s average test scores from year to year, estimating test score progress while accounting for factors such as student background, or projecting future scores based on current and prior years’ results. In 2005, the Department of Education (Education) started a pilot project to allow up to 10 states to use growth models to determine whether schools make AYP for the 2005-2006 school year. The pilot project requires that state growth models meet certain criteria, such as measuring progress toward universal proficiency by 2014.

In response to congressional interest in how growth models may be used to meet the law’s key goals and in anticipation of reauthorization of the Elementary and Secondary Education Act of 1965, we assessed (1) the extent that states have used growth models to measure academic achievement, (2) the extent that growth models can measure school progress in achieving key NCLBA goals, and (3) challenges states may face in using growth models to meet AYP requirements and how Education is assisting the states.

To address these objectives, we conducted a survey of all states, the District of Columbia, and Puerto Rico to determine whether they were using growth models as of March 2006. We received responses from all

except Puerto Rico, and in this report we will refer to the 51 respondents as states. We visited or conducted telephone interviews with state and local educational agency officials in 8 states that collectively use a variety of growth models to understand their use. We selected these states and schools based on expert recommendations and on variation in the types of models they used. To examine the extent that these models measure progress toward key goals of universal proficiency by 2014 and closing achievement gaps, we analyzed student-level data from selected schools in Massachusetts and Tennessee, states selected because of data availability. In both cases, GAO conducted an assessment of the reliability of these data and found the data to be sufficiently reliable for illustrating how growth models measure progress toward key goals of NCLBA. We conducted site visits to those states and selected school districts and schools to learn how their results were calculated. We also conducted site visits in California and North Carolina, two states with different types of growth models, to provide additional perspectives. To identify challenges to using growth models and Education's assistance to states, we interviewed Education officials, state education officials, and other experts, including members of Education's growth model working group. We also reviewed relevant federal and state laws, policies, and guidance as well as research on growth models. For more information on our methodological model, see appendix I. We conducted our work between June 2005 and May 2006 in accordance with generally accepted government auditing standards.

Results in Brief

In addition to using their status models to determine AYP, nearly all states were using or considering growth models for a variety of other purposes, as of March 2006. Twenty-six states were using growth models, and another 22 were either considering or in the process of doing so. Most states that used growth models did so for schools as a whole and for particular groups of students, and 7 also measured growth for individual students. For example, Massachusetts measures changes in schools' and groups' average test scores but not in individual students' scores, while Tennessee sets different expectations for growth for each student based on the student's previous test scores. Seventeen of the states that used growth models had been doing so prior to passage of the NCLBA, while 9 began after the law's passage. States used their growth models for a variety of purposes, such as targeting resources for students that need extra help or awarding teachers bonus money based on their school's relative performance.

Certain growth models may also measure progress toward achieving key NCLBA goals. While growth models may allow states to recognize school progress by tracking student gains over time, if states were allowed to use growth models to determine AYP, they might reduce the number of lower-performing schools eligible for federally required assistance while allowing federal dollars to be concentrated in other lower-performing schools that do not make AYP. We found that certain growth models, like status models, are capable of tracking progress toward the goals of universal proficiency by 2014 and closing achievement gaps. For example, Massachusetts uses its model to set targets based on the growth that it expects from schools and their student groups. Schools can make AYP if they reach these targets, even if they fall short of reaching the statewide proficiency targets set with the state's status model. Tennessee designed a model that projects students' test scores and whether they will be proficient in the future. If 79 percent of a school's students are predicted to be proficient in 3 years, the school would reach the state's 79 percent proficiency target for the current school year. It may be advantageous for schools if their states use growth models to determine AYP, since doing so could recognize improvements schools are making. According to some district officials, growth models could help their schools not be identified for improvement, enabling them to use their own interventions instead of being required to implement school transfer programs or work with state-approved supplemental educational service providers. However, if states allow schools to make AYP by using growth models, they may overlook some lower-performing schools. For example, one school in Massachusetts that served a high-minority, low-income population missed the state English/Language Arts status model proficiency targets overall and for each of its student groups. One group, students with disabilities, scored 44.3 points even though the proficiency target was 75.6. Yet the school was able to make AYP because the school and all of its groups met or exceeded their academic growth targets, including the group of students with disabilities which improved by 6.3 points. These schools would need to increase student proficiency at a faster rate than schools making AYP under a status model.

States face technical challenges such as creating data and assessment systems to support growth models, and Education has taken steps to help states. The ability of states to use growth models to determine AYP depends on the complexity of the model they choose and the degree to which their existing data systems can be used to meet additional demands the model may require. Complex growth models may require states to uniquely identify all students and may call for data on student background, teacher assignments, and courses. Moreover, states that use growth

models for the first time need at least 2 years of data before they could attain usable results. Ensuring that these models are valid and reliable also presents a unique set of challenges, as does hiring technical experts to implement them and training local administrators and teachers to interpret and present results. In 2005, Education started a pilot project to allow up to 10 states to use growth models that met the department's criteria along with their status models, to determine whether schools make AYP. In May 2006, Education approved 2 states—North Carolina and Tennessee—to participate in the pilot and to make AYP determinations for the 2005-2006 school year based on their proposed growth models. Education will also consider proposals from other states for the 2006-2007 school year. Recognizing that NCLBA requires increasingly detailed data and analyses, Education also has initiated a separate grant program to help states develop data systems that can measure, track, and analyze student test scores from grade to grade and year to year.

By proceeding with a pilot project with clear goals and criteria and by requiring states to compare results from their growth model with status model results, Education is poised to gain valuable information on whether or not growth models are overstating progress or whether they appropriately give credit to fast-improving schools. In comments on a draft of this report, Education expressed concern that the use of a broader definition of growth models would be confusing. GAO used this definition in order to reflect the variety of approaches states have been taking to measure growth in academic performance.

Background

The No Child Left Behind Act of 2001¹ increased the federal government's role in kindergarten-12th grade education by setting two key goals:

- to reach universal proficiency so that all students score at the proficient level of achievement—as defined by the states—by 2014, and
- to close achievement gaps between high- and low-performing students, especially those in designated groups: students who are economically disadvantaged, are members of major racial or ethnic groups, have learning disabilities, or have limited English proficiency.

¹ Pub. L. No. 107-110 (Jan. 8, 2002).

With these two key goals in mind, NCLBA requires states to set challenging academic content and achievement standards in reading or language arts and mathematics² to determine whether school districts and schools make AYP toward meeting these standards.³

Education has responsibility for general oversight of the NCLBA. As part of this oversight, Education is responsible for reviewing and approving state plans for meeting AYP requirements. As we have reported, it approved all states' plans—fully or conditionally—by June 2003.⁴ It also reviews state systems of standards and assessments to ensure they are aligned with the law's requirements. As of April 2006, Education had approved these systems for Delaware, South Carolina, and Tennessee and was in the process of reviewing them in other states.

Status Models

States measure AYP using a status model that determines whether or not schools and students in designated groups meet proficiency targets on state tests 1 year at a time. To make AYP, schools must

- show that the percentage of students scoring at the proficient level or higher meets the state proficiency target for the school as a whole and for designated student groups,
- test 95 percent of all students and those in designated groups, and
- meet goals for an additional academic indicator (which can be chosen by each individual state for elementary and middle schools but must be the state-defined graduation rate in high schools).

States generally used data from the 2001-2002 school year to set the initial percentage of students that needed to be proficient for a school to make AYP, known as a starting point, as prescribed in the NCLBA and Education's guidance. Using these initial percentages, states then set annual proficiency targets that increase up to 100 percent by 2014. For example, for schools in a state with a starting point of 28 percent to achieve 100 percent by 2014, the percentage of students who scored at or

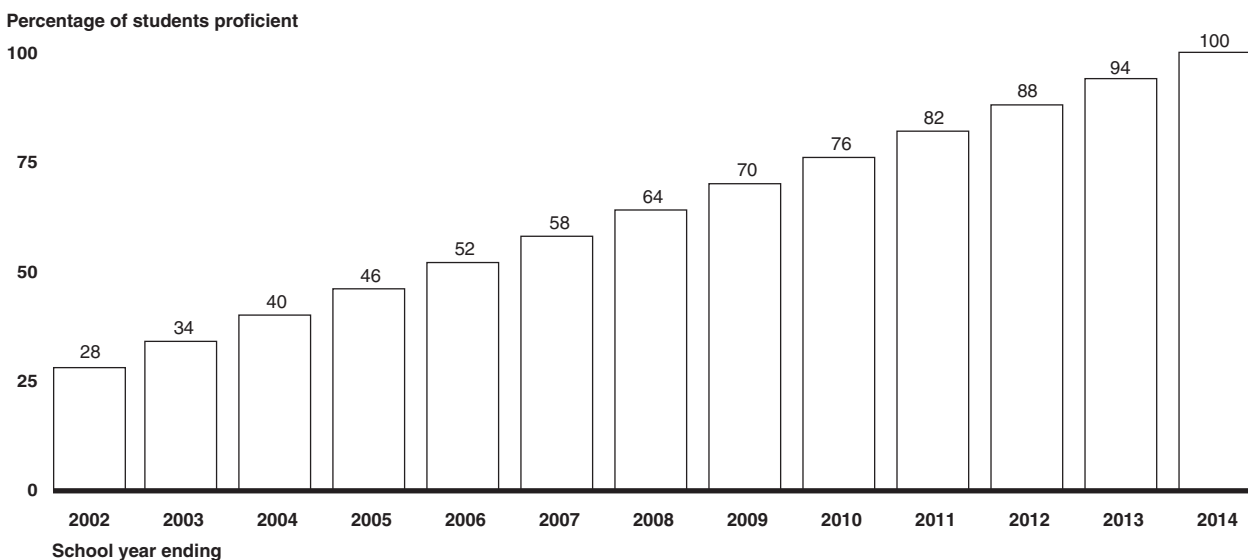
² The law also requires content standards to be developed for science beginning in the 2005-2006 school year and science tests to be implemented in the 2007-2008 school year.

³ States determine whether schools and school districts make AYP or not. For this report, we will discuss AYP determinations in the context of schools.

⁴ GAO, *No Child Left Behind Act: Improvements Needed in Education's Process for Tracking States' Implementation of Key Provisions*, [GAO-04-734](#), (Washington, D.C.: Sept. 30, 2004).

above proficient on the state test would have to increase by 6 percentage points each year, as shown in figure 1.⁵ Setting targets for increasing proficiency through 2014 does not ensure that schools will raise student performance to these levels. Instead, the targets provide a goal, and schools that do not reach the goal will generally not make AYP.

Figure 1: Hypothetical Example of Annual Proficiency Targets Set under a Status Model



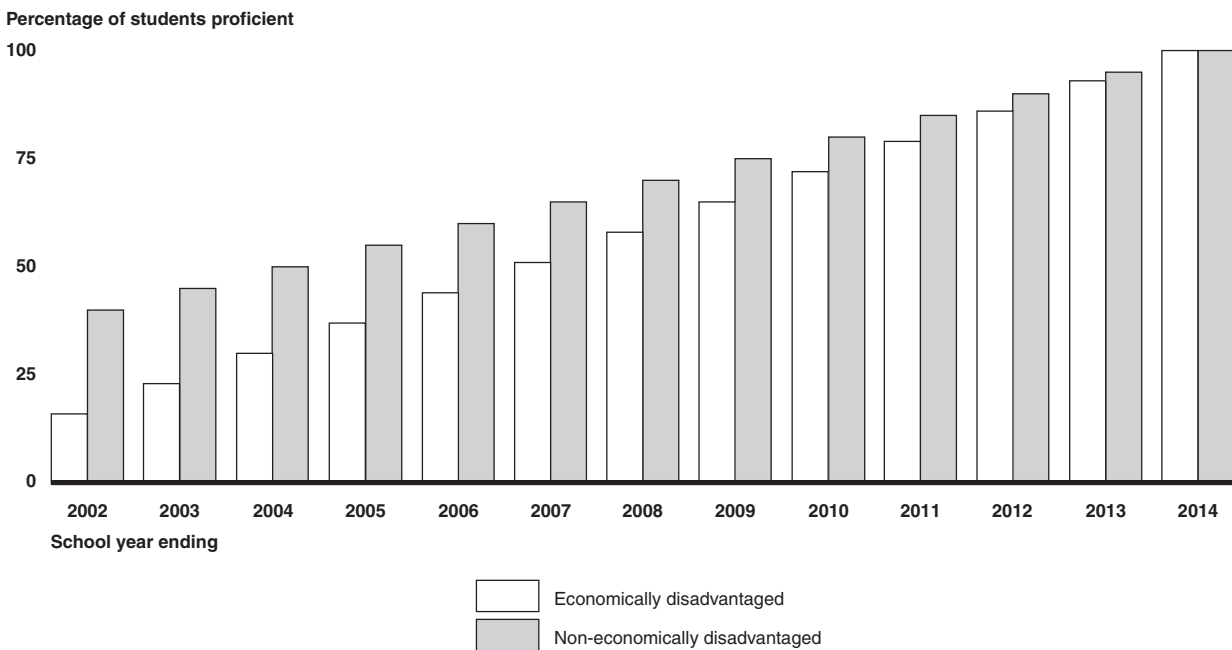
Source: GAO.

School districts with schools receiving federal funds under Title I Part A that do not make AYP for 2 or more years in a row must take action to assist students, such as offering students the opportunity to transfer to other schools or providing additional educational services like tutoring. School districts with schools that meet these criteria must set aside an amount equal to 20 percent of their Title I funds to provide these services and spend up to that amount depending on how much demand exists for these services to be provided. These schools, in consultation with their districts, are also required to implement a plan to improve their students' achievement.

⁵ States were able to map out different paths to universal proficiency, so long as there were increases at least once every 3 years and those increases led to 100 percent proficiency by 2014. See [GAO-04-734](#), pages 16-18.

The law indicates that states are expected to close achievement gaps, but does not specify annual targets to measure progress toward doing so. States thus have flexibility in the rate at which they close these gaps. To determine the extent that achievement gaps are closing, states measure the difference in the percentage of students in designated student groups and their peers that reach proficiency. Using a hypothetical example, figure 2 shows how closing achievement gaps between economically disadvantaged students and their peers would be reported.

Figure 2: Hypothetical Example of Closing Achievement Gaps



Source: GAO.

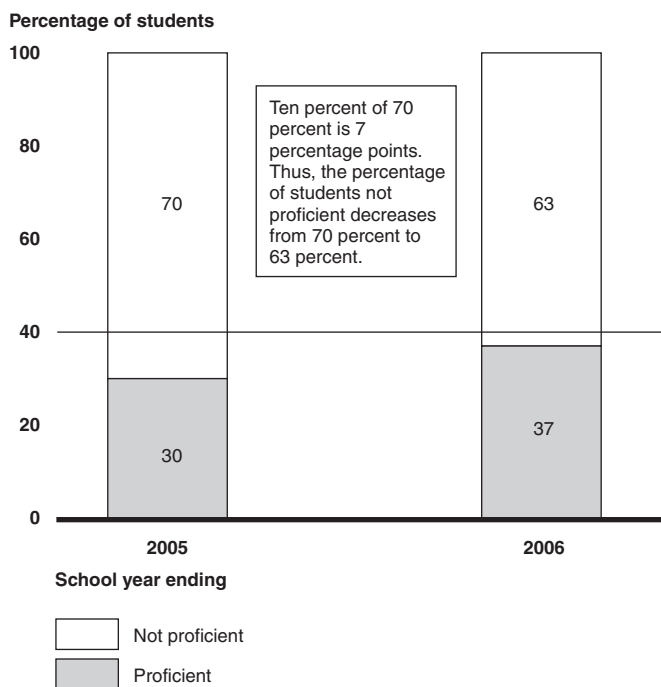
In this example, 40 percent of the school’s non-economically disadvantaged students were proficient compared with only 16 percent of disadvantaged students in 2002, a gap of 24 percentage points. To close the gap, the percentage of students in the economically disadvantaged group that reaches proficiency would have to increase at a faster rate than that of their peers. By 2014, the gap is eliminated, with both groups at 100 percent proficient.

Safe Harbor

If a school misses its status model target, the law also provides a way for it to make AYP if it significantly increases the proficiency rates of student groups that do not meet the proficiency target. The law includes a

provision, known as safe harbor, which allows a school to make AYP by reducing the percentage of students in designated student groups that were not proficient by 10 percent, so long as it also shows progress on another academic indicator. Safe harbor measures academic performance similar to certain growth models, according to one education researcher. For example, in a state with a status model target of 40 percent proficient, a school could make AYP under safe harbor if 63 percent of a student group were not proficient compared to 70 percent in the previous year. See figure 3.

Figure 3: Safe Harbor



Source: GAO.

Growth Models

In contrast to status models that measure the percentage of students at or above proficiency in a school 1 year at a time, growth models measure change in achievement or proficiency over time. Some of these models show changes in achievement for schools and student groups using students' average scores. Other models provide more detailed information on how individual students progress over time. Growth models can enable school officials to monitor the year-to-year changes in performance of students across many levels of achievement, including those who may be

well below or well above proficiency. They may also be used to predict test scores in future years based on current and prior performance.

While definitions of growth models vary, for this report, GAO defines a growth model as a model that measures changes in proficiency levels or test scores of a student, group, grade, school, or district for 2 or more years. Some definitions restrict the use of the term “growth models” to refer only to those models that measure changes for the same students over time.⁶ GAO included models in this report that track different groups of students in order to provide a broad assessment of options that may be available to states. Growth models can be designed to measure successive groups of students (for example, students in the third grade class in 2006 with students in the third grade class in 2005) or track a cohort of students over time (for example, students in the fourth grade in 2006 with the same students in the third grade in 2005).⁷

School-level growth models track changes in the percentage of students that reach proficiency or their achievement scores over time. For example, the charts in figure 4 show how two hypothetical schools measure their proficiency with a status model and with a measure of progress over time.

⁶ Council of Chief State School Officers, *Policymakers' Guide to Growth Models for School Accountability: How Do Accountability Models Differ?* Washington, D.C.: Oct. 2005.

⁷ Robert L. Linn “Accountability Models,” in Susan H. Fuhrman and Richard F. Elmore, eds. *Redesigning Accountability Systems for Education*, pp. 73-95. New York: Teachers College Press, 2004.

Figure 4: Results from Two Hypothetical Schools, Shown with a Status Model and a Growth Model

| School | Status model Percentage proficient in 2 school years (Annual proficiency target = 40 percent) | | Growth model Change in percentage proficient |
|--------------------------|---|-----------------|---|
| | 2004–2005 | 2005–2006 | |
| Washington Middle School | <p>AYP: Yes</p> | <p>AYP: Yes</p> | <p>-5</p> |
| Adams Elementary School | <p>AYP: No</p> | <p>AYP: No</p> | <p>5</p> |

Source: GAO.

In the case of Washington Middle School, a growth model shows a decline in performance, while a status model indicates that the school exceeded the state proficiency target of 40 percent. This school was able to make AYP even though its proficiency rate decreased. In contrast, the use of a growth model with Adams Elementary School shows that the school improved its performance, but its status model results indicate that the school did not meet the 40 percent proficiency target. That school did not make AYP, even though its proficiency rate increased. Thus, the type of model used could lead to different perspectives on how schools are performing.

Individual-level growth models track changes in proficiency or achievement for individual students over time. For example, individual student growth can be measured by comparing the difference between a student's test scores in 2 consecutive years. A student may score 300 on a test in one year and 325 on the test in the next year, resulting in an increase of 25 points.⁸ These scores could then be averaged to measure school-level results as in the previous example. Individual student growth can also be measured over more than 2 years to identify longer-term trends in performance. Additionally, growth can be projected into the future to predict when a student may reach proficiency, and that information may be used to target interventions to students who would otherwise continue to perform below standard.

Nearly All States Were Using or Considering Growth Models to Supplement Other Measures of Academic Performance

Nearly all states were using or considering growth models to track performance, as of March 2006. Although NCLBA requires states to use status models to determine whether schools make AYP, the 26 states with growth models reported using them for state purposes such as identifying schools in need of extra assistance. Seventeen of these states had growth models in place prior to NCLBA.

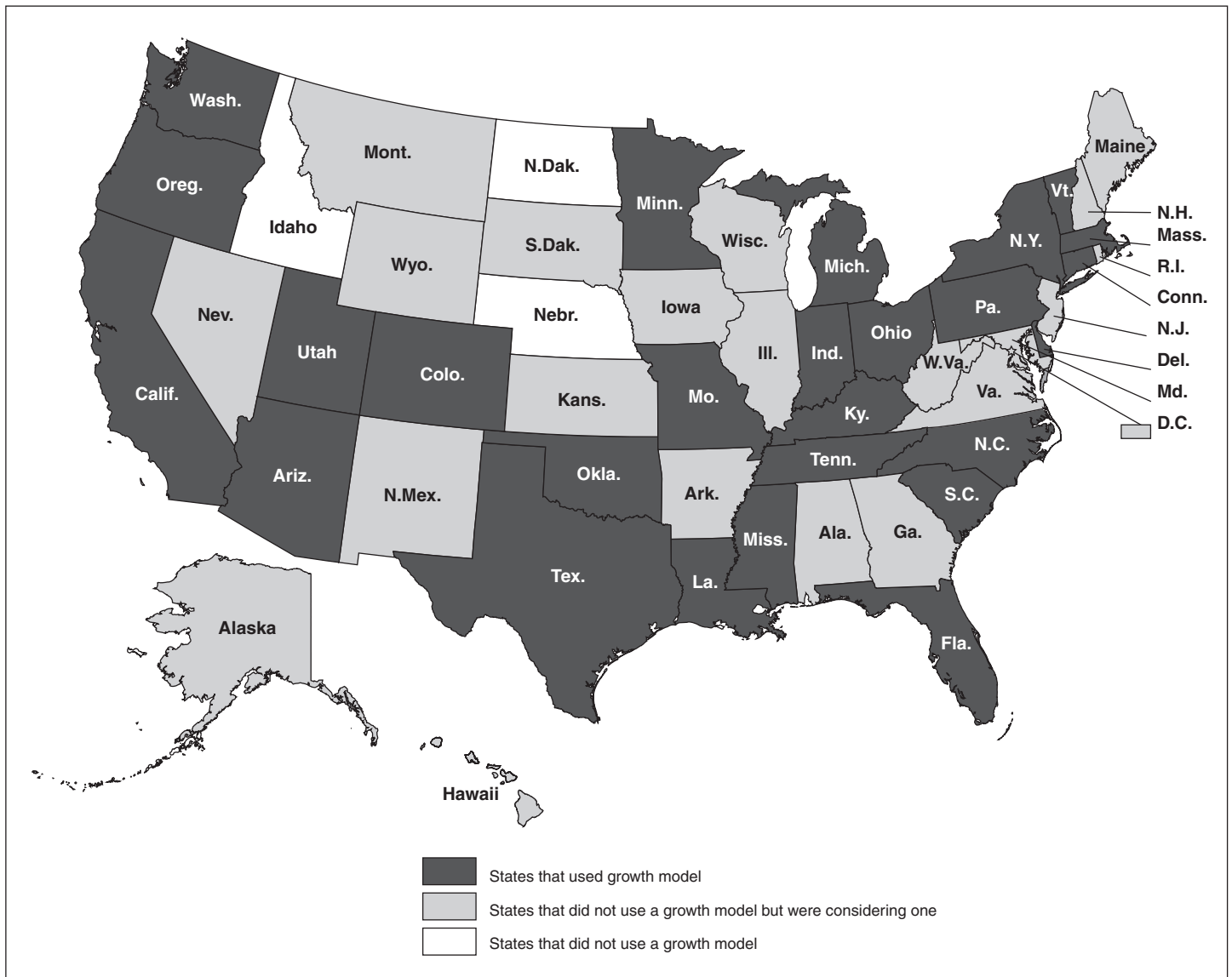
Half of the States Were Using Growth Models, and Most Remaining States Were Considering Them

Twenty-six states reported using growth models in addition to using their status models to track the performance of schools, designated student groups, or individual students, in our survey as of March 2006 (see figure 5). Additionally, nearly all states are considering the use of growth models:

- 20 of 26 states that used one growth model were also considering or in the process of implementing another growth model, and
- 22 of 25 states that did not use growth models were considering or in the process of implementing them to provide more detailed information about school, group, or student performance.

⁸ Since students are expected to know more over time, such growth may or may not be what is expected of students.

Figure 5: Map of States That Reported Using or Considering Growth Models, as of March 2006



Source: GAO survey.

Of the 26 states using growth models, 19 states reported measuring changes for schools and student groups, while 7 states reported measuring changes for schools, student groups, and individuals, as seen in table 1.

Table 1: Types of Growth Models and States Using Them, as of March 2006

| Measures growth of schools and groups | Measures growth of schools, groups, and individual students |
|---|--|
| Compares the change in scores or proficiency levels of schools or groups of students over time. | Compares the change in scores or proficiency levels of schools, groups of students, and individual students over time. |
| Data requirements, such as measuring proficiency rates for schools or groups, are similar to those for status models. | Data requirements, such as tracking the proficiency levels or test scores for individual students, are typically more involved than those for status models. |
| Arizona California Colorado Connecticut Delaware Indiana Kentucky Louisiana Massachusetts Michigan Minnesota Missouri New York Ohio Oklahoma Oregon Pennsylvania Vermont Washington | Florida Mississippi North Carolina South Carolina Tennessee Texas Utah |

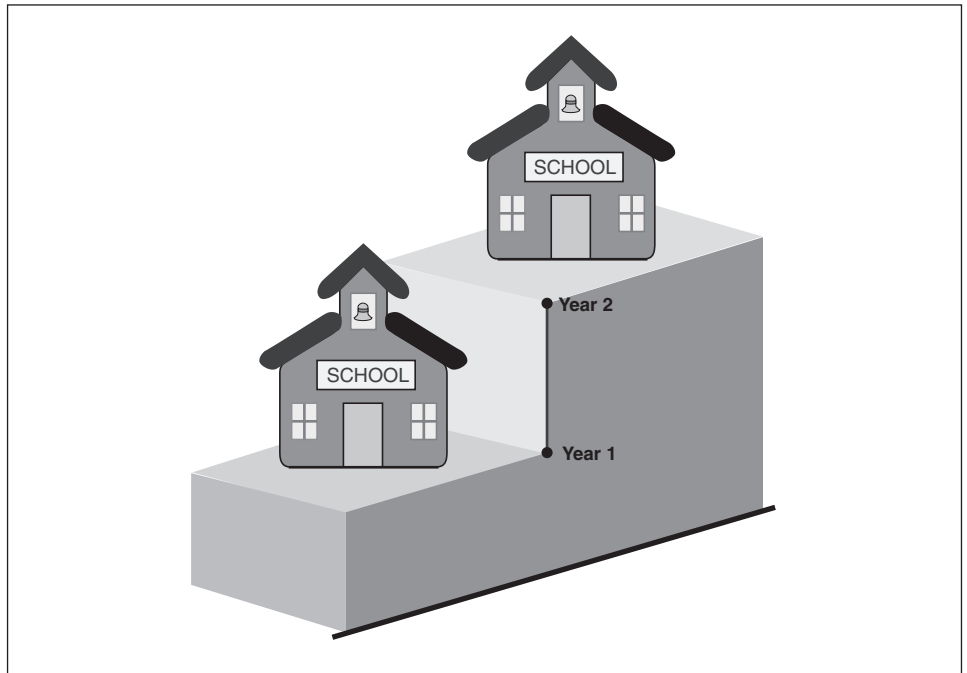
Source: GAO survey.

For example, Massachusetts’ model measures growth for the school as a whole and for designated student groups. The state awards points to schools in 25-point increments for each student,⁹ depending on how students scored on the state test. Schools earn 100 points for each student who reaches proficiency, but fewer points for students below proficiency. The state averages the sum of those points to award a final score to schools. Growth in Massachusetts is calculated by taking the difference in the average points that a school earns between 2 years, as illustrated in figure 6. Because growth in Massachusetts is based on the number of students in schools and groups scoring at various levels of proficiency, the

⁹ Students with disabilities are generally included in these calculations. The state is allowed to give different tests to students with significant cognitive impairments and to count them differently for calculating points awarded to schools.

data needed to calculate growth are similar to those used for calculating status model results.

Figure 6: Illustration of School-Level Growth

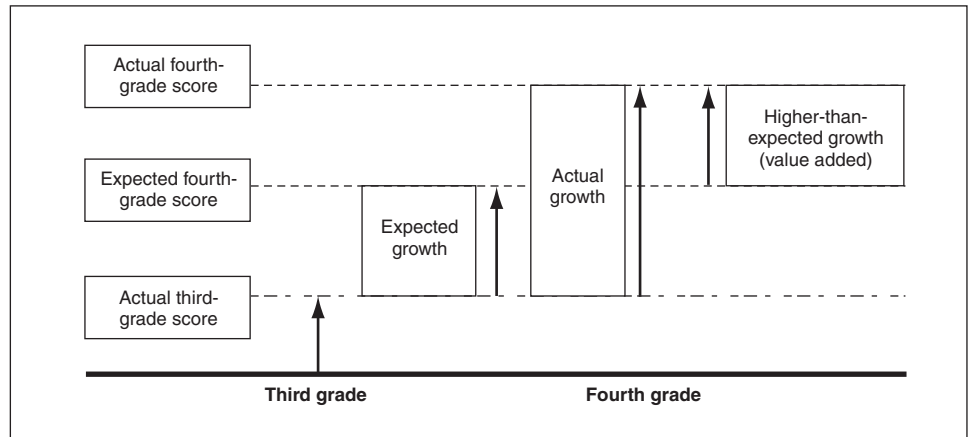


Sources: GAO and Art Explosion.

Other models measure growth and set goals for each individual student as well as for student groups and the school as a whole. For example, Tennessee reported using a growth model for state accountability purposes. Its model, known as a value-added model, sets different goals for each student based on the students' previous test scores. The goal is the score that a student would be statistically expected to receive, and any difference between a student's expected and actual score is considered that student's amount of yearly growth,¹⁰ as shown in figure 7.

¹⁰ Tennessee's growth model described here is not used to make AYP determinations under NCLBA. However, Tennessee is planning to use another growth model to make AYP determinations for the 2005-2006 school year.

Figure 7: Example of Measurement of Above Average Growth for a Fourth-Grade Student under Tennessee’s Model



Source: GAO illustration based on information provided by the State of Tennessee.

Tennessee’s model, known as the Tennessee Value-Added Assessment System, determines growth for designated student groups and for the school as a whole by using information on the relationships between all students’ test scores to estimate future scores for students and schools based on past performance over time. The model can therefore determine whether schools are below, at, or above their level of expected performance. The model also estimates the unique contribution that the teacher and school make to each student’s growth in test scores over time.¹¹ The state then uses that amount of growth, the unique contribution of the school, and other information to grade schools with an A, B, C, D, or F, which is considered a reflection of the extent to which the school is meeting its requirements for student learning.

Another approach that measures individual student growth is Florida’s approach that calculates, according to a state official, “student learning gains.” Unlike Tennessee’s model, which estimates future scores of individual students, Florida’s model measures individual growth from one year to the next by comparing the student’s present performance with performance in prior years. For example, a school can get credit for student growth if a student moved from a lower level of proficiency to a

¹¹ The state calculates the unique contribution of schools and teachers by using a multivariate, longitudinal statistical method where results are estimated using data specific for students within each classroom or school.

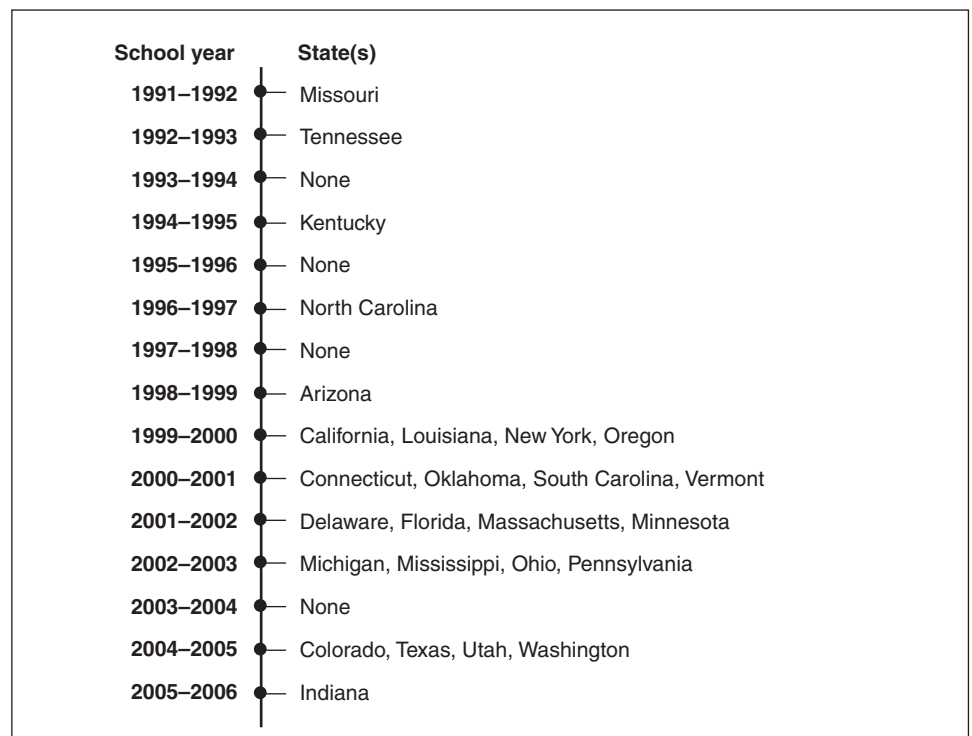
higher one or if a student's test scores increased substantially—more than 1 year's worth of growth—even if the student did not move into a higher proficiency level.

Models that calculate results for individual students require more data than are typically used for calculating status model results. In particular, states must use a data system that is capable of tracking the performance of individual students over time.

States Were Using Growth Models for State, Rather than Federal, Purposes

Seventeen of the 26 states using growth models reported that their models were in place before the passage of the NCLBA during the 2001-2002 school year, and the remaining 9 states implemented them after the law was passed, as shown in figure 8.

Figure 8: States' Introduction of Growth Models



Source: GAO survey.

Once NCLBA was enacted, states were required to develop plans to show how they would meet federal requirements for accountability as measured

by whether their schools made AYP. Education approved these plans, but generally did not permit states to include growth models. According to Education officials, since NCLBA requires that states make AYP determinations on the basis of the percentage of students who are proficient at one point in time—rather than the increase or decrease in that percentage over time—growth models were considered inconsistent with the goals of the act. For example, California began using its model, called the Academic Performance Index, in the 1999-2000 school year to set yearly growth targets for schools. These targets were based on combined test scores for reading/language arts, mathematics, and other subjects. However, according to officials at the California Department of Education, California’s model, developed prior to NCLBA, was not designed to explicitly achieve the law’s key goals of universal proficiency by 2014 or closing achievement gaps. Further, a California Department of Education official explained that because the model did not report scores from reading, math, and other subjects separately, California was not approved to make AYP determinations using its model.¹² In contrast, Massachusetts’ growth model was in place prior to NCLBA passage and then was adapted to align explicitly with the law’s key goals. Education approved Massachusetts’ AYP plan, allowing the state to use both its status model and growth model to determine AYP.

Instead of using growth models to make AYP determinations, states used them for other purposes, such as rewarding effective teachers and designing intervention plans for struggling schools. For example, North Carolina used its model as a basis to decide whether teachers receive bonus money. Tennessee used its value-added model to provide information about which teachers are most effective with which student groups. In addition to predicting students’ expected scores on state tests, Tennessee’s model was used to predict scores on college admissions tests, which is helpful for students who want to pursue higher education. In addition, California used its model to identify schools eligible for a voluntary improvement program.

The type of growth model used has implications for how results may be applied. California’s model provides information about the performance of its schools, enabling the state to distinguish higher-performing from lower-performing schools. However, the model does not provide information about individual teachers or students. In contrast, Tennessee’s model does

¹² The state uses its growth model as the additional indicator specified in NCLBA.

provide information about specific teachers and students, allowing the state to make inferences about how effective its teachers are. While California may use its results for interventions in schools, Tennessee may use its results to target interventions to individual students.

Certain Growth Models Provide Useful Information on the Extent That Schools Are Achieving Key NCLBA Goals but May Overlook Some Low-Performing Schools if Used for AYP

Certain growth models measure the extent that schools and students are achieving key NCLBA goals. While the use of growth models may allow states to recognize gains schools are making toward the law's goals, it may also put students in some lower-performing schools at risk for not receiving additional federal assistance. While states developed growth models for purposes other than NCLBA, states such as Massachusetts and Tennessee have adjusted their state models to use them to meet NCLBA goals. The Massachusetts model has been used to make AYP determinations as part of the state's accountability plan in place since 2003. This model is approved by Education in part because it complies with the key goal of universal proficiency by 2014. Tennessee submitted a new model to Education for the growth model pilot project that differs from the value-added model we describe earlier. The value-added model, developed several years prior to NCLBA, gives schools credit for students who exceeded their growth expectations. The new model gives schools credit for students projected to reach proficiency within 3 years in order to comply with the key NCLBA goal of showing that students are on track to reach proficiency by 2014.

Certain Growth Models Measure Progress in Achieving Key NCLBA Goals of Universal Proficiency by 2014 and Closing Achievement Gaps

Like status models, certain growth models can measure progress in achieving key NCLBA goals of reaching universal proficiency by 2014 and closing achievement gaps. Our analysis of how models in Massachusetts and Tennessee can track progress toward the law's two key goals is shown in table 2.

Table 2: How a Status Model and Certain Growth Models Measure Progress in Achieving Key NCLBA Goals

| | Status model | Growth models | |
|-------------------------------|---|---|--|
| | | Massachusetts (school-level and group-level) | Tennessee ^a (student-level) |
| Universal proficiency by 2014 | Sets same annual proficiency target for all schools in the state | Sets biennial growth targets for each school/group in the state | Sets same annual proficiency target for all schools in the state |
| | State proficiency targets increase incrementally to 100% by 2014 | School/group growth targets increase incrementally to 100% proficiency by 2014; increments may be different by school/group | State proficiency targets increase incrementally to 100% by 2014 Projects future test scores to determine if students may be proficient |
| | School makes AYP if it reaches the state proficiency target | School makes AYP if it reaches the state proficiency target or its own growth model targets | School makes AYP if it reaches the state proficiency target based on students projected to be proficient in the future |
| Closing achievement gaps | State proficiency target applies to each student group in all schools | Each student group in a school has its own growth target | State proficiency target applies to each student group in all schools |
| | School makes AYP if each student group reaches the state proficiency target | School makes AYP if each student group reaches the state proficiency target or its own growth model target | School makes AYP if each student group reaches the state proficiency target based on students projected to be proficient in the future |

Source: GAO analysis of NCLBA and of information provided by the states of Massachusetts and Tennessee.

Note: Additional requirements for schools to make AYP are described in the background section of this report. Massachusetts refers to proficiency targets as performance targets and growth targets as improvement targets.

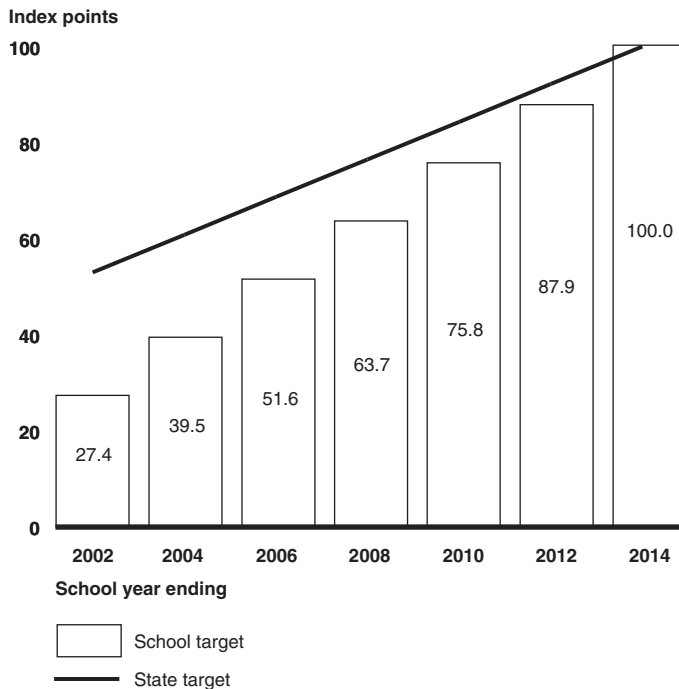
^aThe information presented in this table reflects the model Tennessee proposed to use as part of Education's growth model pilot project, as opposed to the value-added model it uses for state purposes we described earlier in this report. The information is based on the March 2006 revision of the proposal the state initially made in February 2006.

Our analysis of data from selected schools in those states demonstrates how these models measure progress toward the key goals. One school in Massachusetts had a baseline score of 27.4 points in math.¹³ Its growth target for the following 2-year cycle was 12.1, requiring it to reach 39.5 points by 2004. In comparison, the state's target using its status model was 60.8 points in 2004. The growth target was set at 12.1 because, if the

¹³ Massachusetts scores each school with an index system that awards points based on how many students are at different proficiency levels instead of using the percentage of students that are proficient. For example, a school gets 0 points for each student at the lowest of five proficiency levels and 100 points for each fully proficient student, with increments of 25 points for students in the levels in between.

school's points increased this much in each of the state's six cycles, the school would have 100 points by 2014. In so doing, it would reach universal proficiency in that year, as is seen in figure 9.

Figure 9: Targets for a Selected School in Massachusetts Compared to State Status Model Targets

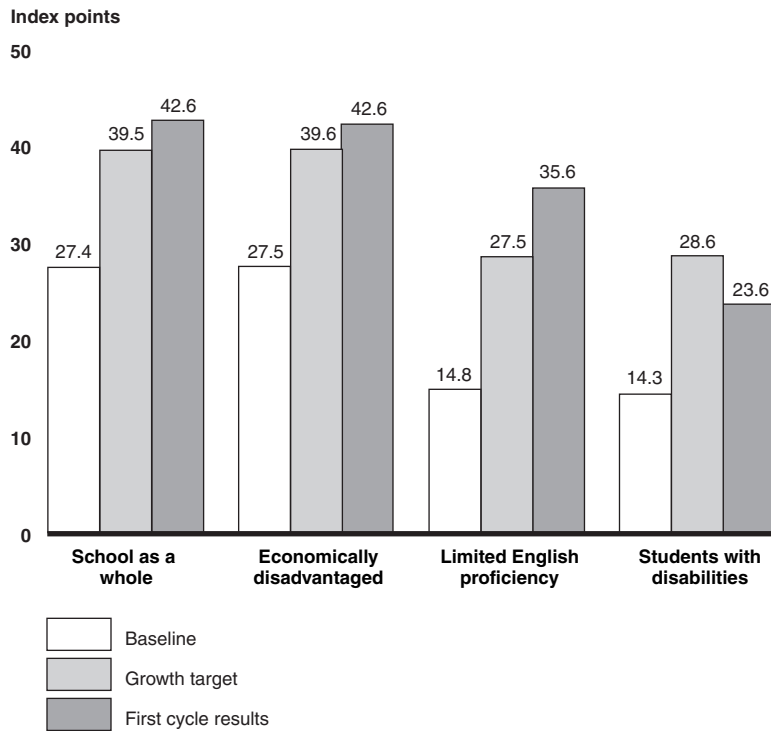


Source: GAO analysis of data provided by Massachusetts Department of Education; *Commonwealth of Massachusetts Consolidated State Application Accountability Workbook*, June 29, 2005.

In fact, the school scored 42.6 in 2004, thus exceeding its target of 39.5. The school also showed significant gains for several designated student groups that were measured against their own targets. However, the school did not make AYP because gains for one student group were not sufficient. This group—students with disabilities—showed gains of 9.3 points resulting in a final score of 23.6 points, short of its growth target of 28.6.¹⁴ Figure 10 compares this school's baseline, target, and first cycle results for the school as a whole and for selected student groups.

¹⁴ According to a state officials, when calculating results using the percentage of students that were not proficient for safe harbor calculations (as described in NCLBA), the state found that the results from this student group were not high enough for the school to make AYP.

Figure 10: Results for a Selected School in Massachusetts in Mathematics



Source: GAO analysis of data provided by the Massachusetts Department of Education.

Massachusetts has designed a model that can measure progress toward the key goal of NCLBA by setting growth targets for each year until all students are proficient in 2014. Schools like the one mentioned above can get credit for improving student proficiency even if, in the short term, the requisite number of students have yet to reach the current status model proficiency targets. The model also measures whether achievement gaps are closing by setting targets for designated student groups, similar to how it sets targets for schools as a whole. Schools that increase proficiency too slowly—that is, do not meet status or growth targets—will not make AYP.

Tennessee developed a different model that also measures progress toward the NCLBA goals of universal proficiency and closing achievement gaps. Tennessee created a new version of the model it had been using for state purposes to better align with NCLBA.¹⁵ Referred to as a projection

¹⁵ Tennessee continues to use its original model to rate schools based in part on the unique contributions—or the value added—of the school to student achievement.

model, this approach projects individual student's test scores into the future to determine when they may reach the state's status model proficiency targets in the future.¹⁶ This model was accepted as part of Education's pilot project, allowing the state to use it to make AYP determinations in the 2005-2006 school year.¹⁷

In order to make AYP under this proposal, a school could reach the state's status model targets by counting as proficient those students who are predicted to be proficient in the future. The state projects scores for elementary and middle school students 3 years into the future to determine if they are on track to reach proficiency, as follows:

- fourth grade students projected to reach proficiency by seventh grade,
- fifth grade students projected to reach proficiency by eighth grade, and
- sixth, seventh, and eighth grade students projected to reach proficiency on the state's high school proficiency test.

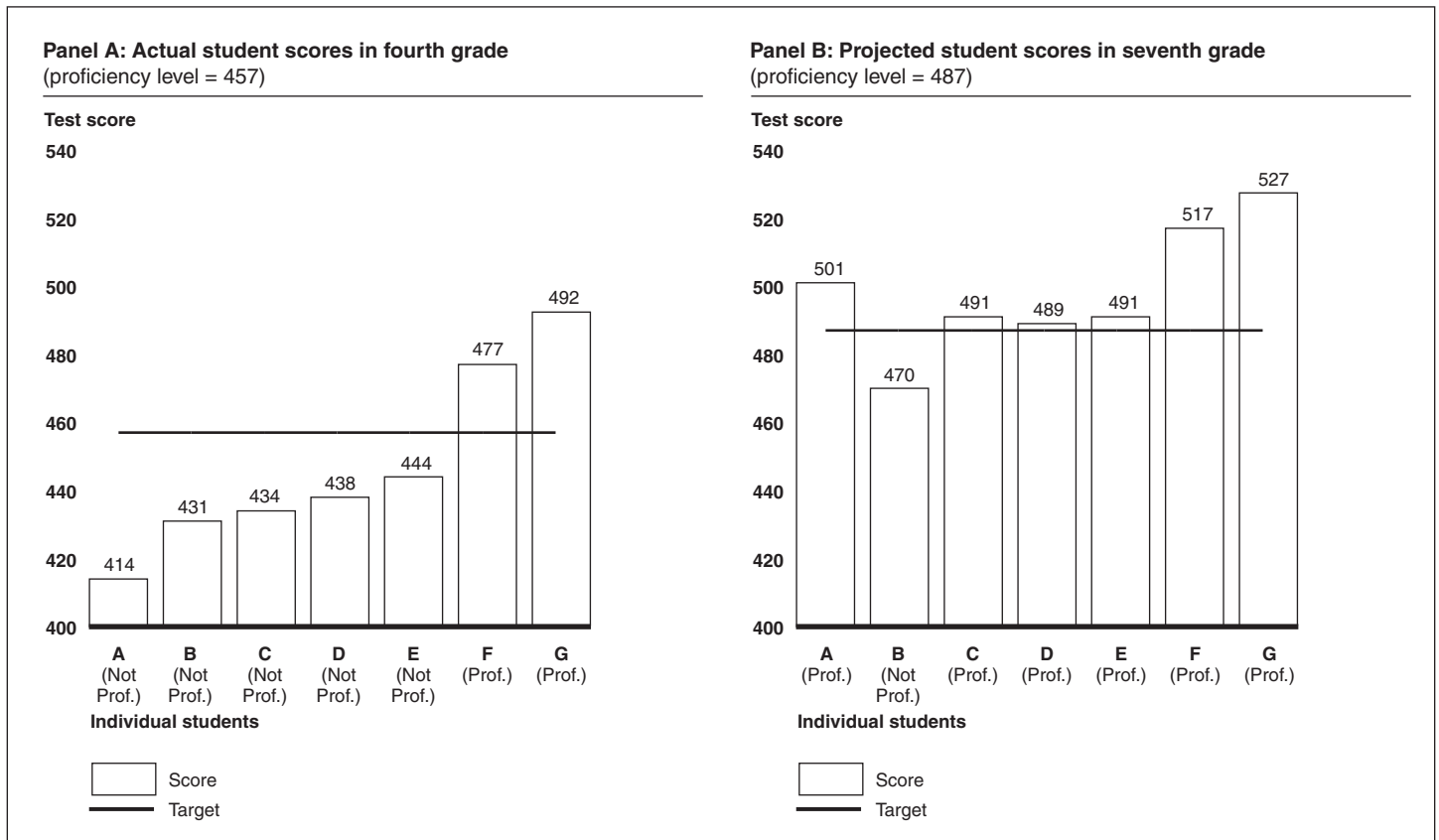
These projections are based on prior test data and are not based on student characteristics. Also, the projections are based on the assumption that the student will attend middle or high schools with average performance (an assumption known as average schooling experience), and allow the student's current school to count them as proficient in the current year if they are projected to be proficient in the future.

Tennessee estimated that of its 1,341 elementary and middle schools, 47 schools that did not make AYP using its status model would be able to make AYP under its proposed model that gives schools credit for students projected to be proficient in the future. At our request, Tennessee provided analyses for students in several schools that would make AYP under the proposed model. To demonstrate how the model works, we selected students from a school and compared their actual results in fourth grade (Panel A) with their projected results for seventh grade (Panel B) (see figure 11).

¹⁶ While Tennessee's model estimates future performance, other models are able to measure growth without these projections. For example, as mentioned earlier in the report, Florida uses a model that calculates results for individual students by comparing performance in the current year with performance in prior years.

¹⁷ Tennessee submitted this model to Education to use as part of the Growth Model Pilot Project in March 2006 on the basis of feedback received regarding its original submission made in February 2006.

Figure 11: Results for Selected Students in Mathematics from a School in Tennessee



Source: Tennessee Department of Education.

Note: The same students are presented in both panels (for example, student A in panel A is the same student as student A in panel B). While these data reflect the scores of individual students, Tennessee provided data to GAO in such a way that student privacy and confidentiality were ensured. Data are illustrative and are not meant to be a statistical representation of the distribution of students in this school.

Some students who were not proficient based on their scores in 2004-2005 were projected to be proficient by the time they reach later grades. For example, student A did not score at the proficient level in fourth grade but was projected to score at the proficient level in seventh grade. The state has proposed to determine whether schools make AYP by using the percentage of students who are projected to be proficient (like student A) in the future, instead of the percentage of students presently proficient. For example, if 79 percent of an elementary school's students are projected to be proficient on future math tests, the school will make AYP

for the state's 79 percent target in the 2005-2006 school year, regardless of the percentage of students in that school that are currently proficient.

Tennessee's proposed model can also measure achievement gaps. Under NCLBA, a school makes AYP if all student groups meet the state proficiency target. For example, a school could have a 20 percentage point gap for one group if, for example, 59 percent of students with limited English proficiency were proficient compared to 79 percent of their peers. While results based on projections may show that achievement gaps are closing, gaps would actually be closed only if the projections were realized.

Giving Credit for Growth May Overlook Some Low- Performing Schools in the Near Term

Using these models to measure progress, states could recognize improvement by allowing some schools to make AYP even though the schools may have relatively low achieving students. These schools may have a long way to go before reaching 100 percent proficiency and will need to increase student proficiency at a faster rate than schools making AYP under a status model. If a school that receives funds under Title I is unable to sustain this rate of progress, it may have difficulty reaching universal proficiency by 2014. In addition, if a school that did not meet status model targets but made AYP by meeting growth model targets, its students may not qualify for additional assistance provided for by NCLBA. Schools that receive Title I funds and that do not make AYP for 2 consecutive years are identified for improvement. According to some school district officials, it may be helpful not to be identified for improvement because they can devise their own interventions instead of implementing school transfer programs or working with state-approved supplemental educational service providers. While delaying these interventions may disadvantage students in some Title I schools, reducing the number of schools identified for improvement could allow for greater concentration of dollars in the lowest-performing schools.

In Massachusetts, of the 134 schools in the two districts we analyzed, 23 of the 59 schools that made AYP did so based on the state's growth model even though they did not reach the state's status model proficiency rate targets in 2003-2004.¹⁸ The state had its growth model approved by

¹⁸ Another 11 schools also met the growth target, but these 11 schools made AYP under NCLBA's safe harbor provision, by which they reduced by 10 percent the percentage of students that had not reached proficiency.

Education as part of its accountability plan and therefore was able to determine that these 23 schools made AYP. One of these schools served a high-minority, low-income population and missed the state proficiency target in English/Language Arts of 75.6 points for the school as a whole and for each of its student groups. For example, one student group, students with disabilities, scored 44.3 points, missing the target by 31.3. However, this school made AYP, because the school as a whole and each of its student groups had shown enough improvement to meet their growth targets—including the group of students with disabilities that improved by 6.8 points.

In Tennessee—of the 1,341 schools for which the state made AYP determinations in the 2004-2005 school year—47 of the 353 schools (13.3 percent) that had not made AYP would do so if the state’s proposed projection model were applied.¹⁹ However, some of these schools have many other indicators of needing assistance. For example, one school that would be allowed to make AYP under the proposed model was located in a high-poverty, inner-city neighborhood. That school receives Title I funding, as two-thirds of its students are classified as economically disadvantaged. The school was already receiving services required under NCLBA to help its students. If it makes AYP 2 years in a row, these services may no longer be required.

Additionally, estimates of future proficiency often rely on certain assumptions. In the case of Tennessee’s proposed model, a key assumption is that students would receive an average schooling experience in the years between when the data were measured and when the final projection is made. According to Tennessee officials, an average schooling experience is defined as one in which a student receives instruction in a school whose performance is the average of all schools in the state. To the extent that a student attends a school with performance that is significantly different from average, actual performance is likely to deviate from the estimates, rendering those estimates relatively less reliable. Moreover, by allowing a school to count students’ future proficiency in the current year, the Tennessee proposal may only be

¹⁹ Under the state’s current model, of the 1,341 schools assessed, 988 made AYP and 353 did not. When the proposed projection model was applied, 1,035 schools would have made AYP and 306 schools would have not. Thus, 47 schools that did not make AYP in the 2004-2005 school year would have been designated as making AYP if the state’s proposed projection model were in place in that school year.

delaying a school's inability to meet status model targets and forestalling needed assistance.

States Face Challenges In Measuring Year-to-Year Growth That Education's Pilot Program and Data System Grants May Help Address

States face challenges in implementing growth models that Education's initiatives may help address. Challenges states face include the extent that states' data and assessment systems will support the models, whether the models can generate valid and reliable results, and states' expertise to use, manage and communicate results about growth. These challenges are generally similar to those faced by states in implementing status models but are accentuated, because growth models measure progress over multiple years and thus require more data and systems designed to track data over time. Education's growth model pilot program and data system grants may make it possible for more states to meet AYP requirements using a growth model, but greater usage largely depends upon improving states' data and assessment systems.

Requirements of Growth Models Pose Challenges for Implementation

One challenge states face in using growth models is the ability to collect comparable data over at least 2 years, a minimum requirement for any growth model. States must ensure that test results are comparable from one year to the next and possibly from one grade to the next, both of which are especially challenging when test questions and formats change. Depending on the type of model, states may incorporate scores from 2, 3, or even more prior years. Officials from 13 states that were implementing or considering the use of growth models told us that they need to consider their state's ability to make comparisons from one year to the next before their model could be operational. Other states that are implementing new data systems or assessments may have to wait a few years before they have enough data to assess progress from one year to the next. For example, one of those state officials said that his state will need at least 3 years of test data in order to set realistic multiyear growth targets for its proposed growth model. Some states currently using growth models, such as Florida and Ohio, have been collecting and comparing student data for several years.

A significant challenge to implementing growth models that use student-level data is the capacity to collect these data across time and schools. This capacity often requires a statewide system to assign unique numbers to identify individual students. At least 37 states have systems with unique numbers as of April 2006, according to officials with the Data Quality Campaign (a nonprofit organization that helps states improve data

quality). Developing and implementing these systems is a complicated process that includes assigning numbers, setting up the system in all schools and districts, and correctly matching individual student data over time, among other steps. For example, school staff must have students' unique numbers when students change schools. However, Education officials have cited cases of school staff assigning a new number for a student instead of locating the student's original number. Additionally, peer reviewers for Education's growth model pilot project cited concerns about the ability of 3 states to correctly match student data from year to year. Some states have contracted with outside organizations to assist them in establishing these systems. In addition, one model provides a "teacher effect score" as an estimate of the impact that individual teachers have on individual students' academic achievement, thus requiring even more information.

Ensuring data are free from errors is important for calculations using status models and growth models. Doing so is more important when using growth models, because errors in multiple years can accumulate, leading to unreliable results. Fourteen state officials cited concerns about the design and reliability of growth models in areas ensuring data accuracy and measuring progress.

States also need greater research and analysis expertise to use growth models as well as support for people who need to manage and communicate the model's results. For example, Tennessee officials told us that they have contracted with a software company for several years because of the complexity of the model and its underlying data system. Florida has a contract with a local university to assist it with assessing data accuracy, including unique student identifiers required for its model. In addition, states will incur training costs as they inform teachers, administrators, media, legislators, and the general public about the additional complexities that occur when using growth models. For example, administrators in one district in North Carolina told us that personnel issues are their main concerns with using growth models. Their district lacks enough specialists who can explain the state's growth model to all principals and teachers in need of guidance and additional training. In an effort to address their limited capacity, district officials told us they have been collaborating with neighboring districts to share training resources regarding the state's growth model.

Education Has Two Programs to Help States Develop Growth Models

In November 2005, Education announced a pilot project for states to submit proposals for using a growth model—one that meets criteria established by the department—along with their status model, to determine AYP. Education officials told us that the department is conducting its pilot project under authority provided in the law that, upon request from a state, allows the Secretary to waive certain requirements in the NCLBA.²⁰ While the NCLBA does not specify the use of growth models for making AYP determinations, the department started the pilot in part to gain information on how these models might help schools achieve the law’s key goals. According to Education officials, 7 states had already requested to use growth models for AYP determinations before the department invited states to submit growth model proposals.

For the growth model pilot project, each state had to demonstrate how its growth model proposal met Education’s criteria, referred to as “core principles” outlined in its November 2005 announcement. While many of these criteria are consistent with the legal requirements of status models, tracking student progress over time and having an assessment system with tests that are comparable over time are new (see table 3).

²⁰ 20 U.S.C. 7861.

Table 3: Criteria for Participating in Education’s Growth Model Pilot Project

| | |
|---|--|
| Key NCLBA goals | <ul style="list-style-type: none">• ensure all students are proficient by 2014• set annual goals to ensure achievement gaps are closing for all students |
| All students meet the same standard | <ul style="list-style-type: none">• establish high expectations for low-achieving students while not setting expectations based on student demographic characteristics or schools |
| Content areas | <ul style="list-style-type: none">• produce separate accountability decisions in reading and math |
| Participation | <ul style="list-style-type: none">• ensure all students in tested grades are included in assessments/accountability system• schools/districts must be held accountable for results of subgroups• model applied statewide must include all schools/districts |
| Assessment system | <ul style="list-style-type: none">• include annual assessments in grades 3-8 and once in high school in both reading and mathematics• must have been operational for more than a year• must receive approval through NCLBA peer review process• must also produce comparable results from grade to grade/year to year |
| Data system | <ul style="list-style-type: none">• track student progress |
| Participation rates and additional academic indicator | <ul style="list-style-type: none">• include student participation rates in assessment system and student achievement on an additional academic indicator |

Source: Education.

Twenty states submitted proposals to Education by the February 17, 2006 deadline. Education reviewed proposals from the 14 states that planned to make AYP determinations for the 2005-2006 school year and forwarded 8 of them for peer review. In May 2006, Education approved North Carolina and Tennessee to use their proposed growth models to make AYP determinations for the 2005-2006 school year. Education noted that those states met all of the department’s criteria, such as reaching the key NCLBA goals of universal proficiency and closing achievement gaps. Additionally, Education and peer reviewers noted that those states had many years of experience with data systems that support calculating results using growth models. The 6 states whose proposals had received peer review were invited to resubmit proposals in September 2006. Other states that had submitted proposals for the 2006-2007 school year, and those that had not previously submitted proposals were invited to do so by November 1, 2006, for potential implementation in the 2006-2007 school year.

While Tennessee received unconditional approval to implement its proposed growth model, peer reviewers noted they were concerned that Tennessee’s use of “average school experience” is likely to result in inaccurate projections, especially for disadvantaged students. This is because many students attend schools in districts that are struggling, and the schools they are likely to attend 3 years out could provide them with a school experience that is markedly below average. For this reason,

Education requested that the state, after it implements the model, provide data to compare actual results with its projections.

North Carolina received approval as long as its system of standards and assessments was approved by July 1, 2006. Reviewers of the state's proposal noted that the state proposed to average student results for calculating growth, instead of examining growth results of all students, in direct violation of Education's criteria. According to Education, the state changed its original approach so that growth would account for all students and would not use averages.

Six states had proposals that were peer-reviewed but not approved. The department cited a variety of reasons for not approving these proposals, including that they did not lead to universal proficiency by 2014, applied growth calculations to nonproficient students only (instead of all students), used a margin of error on individual test results that would likely lead to students' being counted as proficient when in fact they were not, and proposed annually resetting growth targets. Education is allowing these states to resubmit their proposals for review later in 2006. If approved then, they can use growth models to make AYP determinations in the 2006-2007 school year.

Approved states must report to Education the number of schools that made AYP on the basis of their status and growth models. Education expects to share the results with other states, Congress, and the public after it assesses the effects of the pilot.

In addition to the growth model pilot project, Education announced in April 2005 a competition for grants for the design and implementation of statewide longitudinal data systems. While independent of the pilot project, states with a longitudinal data system—one that gathers data on the same student from year to year—will be better positioned to implement a growth model than they would have been without it. Many states applied to participate in the growth model pilot project or received a grant (see table 4).

Table 4: States Submitting a Proposal for Growth Model Pilot or Awarded a Grant for a Longitudinal Data System

| State | Submitted a proposal for the growth model pilot project | Awarded a grant to assist in developing a statewide longitudinal data system |
|----------------|---|--|
| Alaska | ✓ | ✓ |
| Arizona | ✓ | |
| Arkansas | ✓ | ✓ |
| California | | ✓ |
| Colorado | ✓ | |
| Connecticut | | ✓ |
| Delaware | ✓ | |
| Florida | ✓ | ✓ |
| Hawaii | ✓ | |
| Indiana | ✓ | |
| Iowa | ✓ | |
| Kentucky | | ✓ |
| Maryland | ✓ | ✓ |
| Michigan | | ✓ |
| Minnesota | | ✓ |
| Nevada | ✓ | |
| New Hampshire | ✓ | |
| North Carolina | ✓ | |
| Ohio | ✓ | ✓ |
| Oregon | ✓ | |
| Pennsylvania | ✓ | ✓ |
| South Carolina | ✓ | ✓ |
| South Dakota | ✓ | |
| Tennessee | ✓ | ✓ |
| Utah | ✓ | |
| Wisconsin | | ✓ |

Source: GAO analysis of data provided by Education.

Note: States not listed neither applied to participate in the growth model pilot program nor received a grant to develop a longitudinal data system.

Longitudinal data systems link data, such as test scores and enrollment patterns, of individual students over time. Education intended the grants to help states generate and use accurate and timely data to meet reporting requirements, support decision making, and aid education research,

among other purposes. Education received applications from 45 states for the 3-year grants, and in November 2005, Education awarded a total of \$52.8 million in grants to 14 states. States receiving grants must submit annual and final reports on the status of the development and the implementation of these systems. Education plans to disseminate lessons learned and solutions developed by states that received grants.

Concluding Observations

While status models provide a snapshot of academic performance, growth models can provide states with more detailed information on how schools' and students' performance has changed from year to year. Growth models can recognize schools whose students are making significant gains on state tests but are still not proficient and may provide incentives for schools with mostly proficient students to make greater improvements. Educators can use the growth models of individual students to tailor interventions to the needs of particular students or groups. In this respect, models that measure individual students' growth provide the most in-depth and useful information, yet most of the models currently in use are not designed to do this.

Through its approval of Massachusetts' model and the growth model pilot program, Education is proceeding prudently in its effort to allow states to use growth models to meet NCLBA requirements. Education is allowing only states with the most advanced models that can measure progress toward NCLBA goals to use the models to determine AYP. If schools are allowed to make AYP by getting credit for growth, some lower-performing schools will make AYP and the opportunity for school improvements the federal law prescribes to help students may be missed. However, if schools that show the most growth but do not meet status model targets are permitted to make AYP, states could target Title I school improvement on their lowest-performing schools.

By proceeding with a pilot project with clear goals and criteria and by requiring states to compare results from their growth model with status model results, Education is poised to gain valuable information on whether or not growth models are overstating progress or whether they appropriately give credit to fast-improving schools.

Agency Comments and Our Evaluation

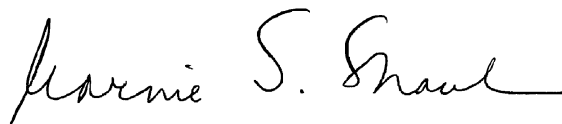
We obtained written comments on a draft of this report from the Department of Education. Education's comments are reproduced in appendix III. Education also provided additional technical comments, which have been included in the report as appropriate.

Education commented that it appreciates our concluding observation that the department “is poised to gain valuable information on whether or not growth models are overstating progress or whether they appropriately give credit to fast-improving schools.” Education expressed concern that the definition of growth models used in the report may confuse readers because it is very broad and includes models that compare changes in scores or proficiency levels of schools or groups of students. To inform its pilot project, Education used research that defines the term “growth model” to refer to models that track the growth of individual students.

For the purposes of this report, we defined growth models to include models that track growth of schools, groups of students, and individual students over time. While we acknowledge that some research exists to define growth models as tracking the same students over time, other research exists to show that there are different ways of classifying models that states use or could potentially use. As such, the definition used in this report reflects the variety of approaches states are taking to measure academic progress.

As agreed with your staff, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days after its issue date. At that time, we will send copies of this report to the Secretary of Education and other interested parties. We will also make copies available to others upon request. In addition, the report will be made available at no charge on GAO’s Web site at <http://www.gao.gov>.

Please contact me at (202) 512-7215 if you or your staff have any questions about this report. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. Key contributors are listed in appendix IV.



Marnie S. Shaul
Director, Education, Workforce,
and Income Security Issues

Appendix I: Scope and Methodology

To address the objectives of this study, we used a variety of methodological models. We interviewed experts in the field of measuring academic achievement as well as state, district, and school officials. We also reviewed documentation from states' Web sites, and examined published studies that detailed characteristics and policy issues of states' models. We conducted a series of interviews in selected states with officials who had a variety of experiences and viewpoints on growth models. In four of those states—California, Massachusetts, North Carolina, and Tennessee—we interviewed officials at the state, district, and school levels so that we could obtain a variety of perspectives on growth models. We selected those states for in-depth interviews based on diverse characteristics of their respective models, all of which were in place prior to the No Child Left Behind Act of 2001 (NCLBA).

To address the first objective, we surveyed state education agencies in the 50 states, the District of Columbia, and Puerto Rico and reviewed documentation from states' accountability workbooks. States reported to us whether they were using or considering the use of a growth model to measure academic achievement. The surveys were conducted using self-administered electronic questionnaires sent in an e-mail to all 52 states beginning January 13, 2006. We closed the survey on March 16, 2006, after the 51st respondent had replied. Puerto Rico did not complete the survey.

The survey asked respondents to indicate, first, whether the state was currently using a growth model. GAO classified school-level models, like improvement models, as growth models for the purposes of this report. Some restrict the use of the term “growth models” to refer only to those that measure changes for the same group of students or individual students over time (see, for example, Council of Chief State School Officers, *Policymakers' Guide to Growth Models for School Accountability: How do Accountability Models Differ?* Washington, D.C.: Oct. 2005). GAO included school-level models in this study to provide a broader assessment of options that may be available to states. If the state was using a growth model, we asked about its characteristics, whether the state was considering use of an additional model, whether the state planned to apply to Education's growth model pilot program, and how the results from its model were used. If the state was not using a growth model, we asked whether it was considering doing so. We also asked about characteristics of the model under consideration and about key issues that must be addressed in order for it to be implemented. In some cases, we asked additional questions in e-mails and in phone interviews. The other methods we used to learn about states' models included

reviewing documentation from states' Web sites and examining published studies that detailed characteristics and policy issues of states' models.

To address the second objective, we analyzed data from selected schools from two states, Massachusetts and Tennessee. These states were chosen based on a variety of factors, including expert recommendation, their use of different growth models, geographic diversity, and data availability. Within these states, we selected schools that were in urban, suburban, and rural areas.

- For Massachusetts, for one urban district and one suburban district, we selected the median school (as measured by the schools' index values) among schools that had shown growth but had not made adequate yearly progress in the 2004-2005 school year.¹
- For Tennessee, for one urban district and one rural district, we selected schools that were used in the state's growth model pilot project proposal.

State officials from Massachusetts provided individual student data to GAO from the two selected school districts. GAO reviewed the state's adequate yearly progress and growth model calculations and replicated school-level index and calculations from student and statewide data. State officials from Tennessee provided analyses its contractor had performed, also using individual student data. In both cases, GAO conducted an assessment of the reliability of these data and found the data to be sufficiently reliable for illustrating how growth models measure progress toward key goals of NCLBA. These assessments included electronic testing of data fields and interviews with state officials and in Tennessee's case, the contractor as well. These interviews consisted of questions regarding the history of the data system, system audits and security, and possible threats to the systems, among other topics. GAO's assessments also included reviews of documentation regarding the data systems.

To address the third objective, we used data from the survey and information provided to us by Education and state officials. We reviewed documentation related to Education's growth model pilot project and proposals submitted by several states. We interviewed Education and state

¹ When selecting the median school, we also selected schools ranked closely to it to ensure that we had selected elementary, middle, and high schools. For example, if the median school was an elementary school, we selected the middle and high schools that were ranked most closely to the elementary school.

officials about the pilot project, including criteria for selection and processes for review and approval. We conducted our work between June 2005 and May 2006 in accordance with generally accepted government auditing standards.

Appendix II: Selected Data from GAO's Survey of States' Use or Consideration of Growth Models

The tables below provide specific information on characteristics of states' growth models (as of the 2005-2006 school year), as reported on the survey. This information includes the grades in which growth models were reported, the level at which growth models were reported, the measures of achievement used to determine growth in test scores, and the characteristics of the assessments used to compare students' test scores.

Growth Model Reporting by Grade

States using growth models varied as to whether or not they used test scores from consecutive grades. Seventeen states reported using growth models in consecutive grades, while 9 states reported using them in nonconsecutive grades. For example, Tennessee uses test scores from grades 4 through 12, while Vermont uses grades 5, 8, and 10. Whether states used test scores from consecutive grades may depend on the type of model they used. The states that reported measuring individual student growth used test scores in consecutive grades (for example, grades 3 through 12 or 4 through 10). In contrast, the 19 states that use school-level information in their growth model calculations varied in the combination of grades they used in their models:

- 11 of those 19 states used growth models in three or more consecutive grades, and
- 8 used a variety of grade combinations.

For each state with a growth model, table 5 lists the grades in which the state reports school growth and indicates whether the model measures individual student growth.

**Appendix II: Selected Data from GAO's
Survey of States' Use or Consideration of
Growth Models**

Table 5: Grades of Growth Model Reporting, by State

| | Grades reported | Measured individual student growth |
|----------------|------------------------|---|
| Arizona | 3 - 8 | |
| California | 3 - 11 | |
| Colorado | 3 - 10 | |
| Connecticut | 4, 6, 8, 10 | |
| Delaware | 3 - 10 | |
| Florida | 4 - 10 | x |
| Indiana | 4 - 10 | |
| Kentucky | 3 - 12 | |
| Louisiana | 3 - 11 | |
| Massachusetts | 3, 4, 6-8, 10 | |
| Michigan | 4, 7, 8, 11 | |
| Minnesota | 3, 5, 7, 10, 11 | |
| Mississippi | 3 - 8 | x |
| Missouri | 3, 7, 8, 10, 11 | |
| New York | 4, 8, 9-12 | |
| North Carolina | 3 - 12 | x |
| Ohio | 3 - 8, 10 | |
| Oklahoma | 3 - 8, 10 | |
| Oregon | 3, 5, 8, 10 | |
| Pennsylvania | 3, 5, 8, 11 | |
| South Carolina | 4 - 8 | x |
| Tennessee | 4 - 12 | x |
| Texas | 4 - 11 | x |
| Utah | 3 - 12 | x |
| Vermont | 5, 8, 10 | |
| Washington | 4, 7, 10 | |

Source: GAO survey.

**Appendix II: Selected Data from GAO's
Survey of States' Use or Consideration of
Growth Models**

**Level of Reporting Growth
Model Results**

States with growth models reported results for schools but varied in terms of reporting results at other levels, such as the individual student or school district. Table 6 lists the different levels at which states with growth models reported results.

Table 6: Level of Growth Model Reporting, by State

| | Level at which growth model results are reported | | | | | | | | |
|----------------|--|-----------|---------|---------------------------|----------|--------|----------|-------|-------|
| | Student | Classroom | Teacher | Cohort/ grade level | Subgroup | School | District | State | Other |
| Arizona | | | | x | | x | | x | x |
| California | | | | | x | x | x | x | |
| Colorado | | | | | | x | | | |
| Connecticut | | | | x | x | x | x | x | |
| Delaware | | | | | | x | x | x | |
| Florida | x | | | | | x | x | | |
| Indiana | | | | x | x | x | x | x | |
| Kentucky | | | | x | x | x | x | x | |
| Louisiana | | | | | x | x | x | x | |
| Massachusetts | | | | x | x | x | x | x | |
| Michigan | | | | x | | x | | | |
| Minnesota | | | | x | x | x | x | x | |
| Mississippi | | | | | | x | | | |
| Missouri | x | x | x | x | x | x | x | x | |
| New York | | | | | x | x | x | | |
| North Carolina | | | | | | x | | | |
| Ohio | | | | | x | x | x | x | |
| Oklahoma | | | | x | x | x | x | x | |
| Oregon | | | | | | x | | | |
| Pennsylvania | | | | | x | x | x | x | |
| South Carolina | | | | | | x | x | | |
| Tennessee | x | x | x | x | x | x | x | x | |
| Texas | | | | | | x | x | x | |
| Utah | x | x | | x | x | x | x | x | |
| Vermont | | | | | x | x | x | | |
| Washington | | | | | | x | x | | |

Source: GAO survey.

Measure of Achievement in
Growth Models

The measure of achievement in growth models indicates the methods states use to compare individual and group scores to determine the amount of growth. Table 7 outlines the measures that each state with a growth model used to determine how growth is reported.

Table 7: Measure of Achievement in Growth Models

| | Measure of achievement in growth models | | | | |
|----------------|---|--------------------|-------------|----------------------------------|-------|
| | Scaled scores | Proficiency levels | Index units | Measure of relative distribution | Other |
| Arizona | | | | x | |
| California | | | x | | |
| Colorado | x | | | | |
| Connecticut | x | x | | | |
| Delaware | x | x | | | |
| Florida | | x | | | x |
| Indiana | x | x | | | |
| Kentucky | x | x | x | x | |
| Louisiana | | | x | | |
| Massachusetts | | | x | | |
| Michigan | | x | | | |
| Minnesota | x | x | x | x | |
| Mississippi | x | | | | |
| Missouri | x | | | | |
| New York | | x | x | | |
| North Carolina | | | | x | |
| Ohio | | | x | | |
| Oklahoma | | x | x | | |
| Oregon | | | x | | |
| Pennsylvania | | x | x | | |
| South Carolina | | | x | | |
| Tennessee | x | x | | x | x |
| Texas | x | | | | |
| Utah | x | x | | | |
| Vermont | | | x | | |
| Washington | | x | | | |

Source: GAO survey.

**Appendix II: Selected Data from GAO's
Survey of States' Use or Consideration of
Growth Models**

Notes: Scaled scores: Raw scores that have been transformed to scores with common attributes (i.e., a certain mean or standard deviation).

Proficiency levels: Descriptions of a student competency in a particular subject area, usually defined as ordered categories on a continuum, often labeled from "basic" to "advanced."

Index units: A tool for measuring school, district, and state level performance that is calculated by combining test results from multiple grades into a composite score. This score is then compared to a goal, from which a performance rating is generally given.

Measure of relative distribution: A means of indicating the percentage of scores in a distribution that are lower than a particular obtained score. The remaining scores are at the same level or higher.

**Appendix II: Selected Data from GAO's
Survey of States' Use or Consideration of
Growth Models**

**Characteristics of
Assessments Used in
Growth Models**

Growth models rely on data from state proficiency tests and measure growth with a variety of characteristics as shown in table 8.

Table 8: Characteristics of Assessments Used in Growth Models

| | Assessment mechanisms in growth models | | | | | Other |
|----------------|--|--------------------------|--|--|----------------------------------|-------|
| | Adjacent grades | Vertically linked scales | Vertically moderated content standards | Nationally norm-referenced assessments | Criterion-referenced assessments | |
| Arizona | x | | | x | | |
| California | | | | x | x | |
| Colorado | x | x | x | | x | |
| Connecticut | | | | | x | |
| Delaware | | | x | | x | |
| Florida | x | x | x | | x | |
| Indiana | x | x | | | x | |
| Kentucky | | | | x | x | |
| Louisiana | | | | | x | x |
| Massachusetts | x | | | | x | |
| Michigan | | | | | x | |
| Minnesota | | | | | x | |
| Mississippi | x | x | | | x | |
| Missouri | | | | x | x | |
| New York | x | x | x | | x | |
| North Carolina | x | x | x | | x | |
| Ohio | x | | x | | x | |
| Oklahoma | x | | x | | x | |
| Oregon | | | | | x | x |
| Pennsylvania | | | x | | x | |
| South Carolina | x | | x | | x | |
| Tennessee | x | x | x | x | x | |
| Texas | x | | | | x | |
| Utah | x | | | | x | |
| Vermont | x | | x | | x | |
| Washington | | | | | x | |

Source: GAO survey.

**Appendix II: Selected Data from GAO's
Survey of States' Use or Consideration of
Growth Models**

Notes: Adjacent grades: Grades that are consecutive, as opposed to nonconsecutive. This term is used in reference to tests administered in, for example, grades 3-5 or 9-12.

Vertically linked scales: Scales that link similar content across grades in a way that allows test scores to be directly comparable.

Vertically moderated content standards: Standards that equate content across grades and focus on categories of performance within a content area, not on changes in scaled scores.

Nationally norm-referenced assessments: Assessments that compare a student's score to those of a nationally representative group of students.

Criterion-referenced assessments: Assessments that relate a student's score on an assessment to a specific body of knowledge, such as the state's content standards, rather than to other students' scores.

Appendix III: Comments from the Department of Education



UNITED STATES DEPARTMENT OF EDUCATION

OFFICE OF ELEMENTARY AND SECONDARY EDUCATION

THE ASSISTANT SECRETARY

Ms. Marnie S. Shaul
Director, Education, Workforce
and Income Security Issues
Government Accountability Office
441 G Street, NW
Washington, DC 20548

JUL 10 2006

Dear Ms. Shaul:

I am writing in response to your request for comments on the Government Accountability Office (GAO) draft report (GAO-06-661), dated June 2006, and entitled "No Child Left Behind Act: States Face Challenges Measuring Academic Growth That Education's Initiatives May Help Address." I appreciate the opportunity to comment on the draft report and provide insight on our actions regarding the use of growth in making accountability determinations.

The national goal of universal academic proficiency embodied in the No Child Left Behind Act of 2001 (NCLB) created an unprecedented emphasis on the achievement of every student. Each State responded to NCLB by developing and implementing an accountability plan that holds schools, districts and States accountable for all students reaching grade-level proficiency in reading and mathematics by 2013-2014. These accountability plans are primarily based on schools and districts increasing the percentage of students proficient in reading and mathematics as set forth in statute, but States have asked for an additional approach to accountability that makes individual student academic growth a principal factor in determining whether a school or district is meeting its academic goals.

After seeking the input of accountability experts throughout the nation to identify how using measures of academic growth may be a viable method of strengthening accountability while maintaining the focus on academic proficiency for every child, Secretary Margaret Spellings announced a pilot project to allow up to ten States to use a growth-based accountability plan that aligns with the goals of NCLB. The two-fold purpose in allowing the growth-based accountability model pilot project was (1) to evaluate how growth models might help schools meet the NCLB proficiency goals and close the achievement gap, and (2) to provide the Department and States useful information to improve the quality of data, assessment, and accountability systems.

The Department used a peer review process to consider and select State growth model proposals for participation in the pilot project. The peer review served to ensure that the selected States have technically sound growth models, a high probability of success in incorporating measures of student growth into accountability systems, and a model consistent with NCLB. Further, as the Secretary has announced that she will consider new proposals to use growth models for the coming school year, the peer review

400 MARYLAND AVE., S.W., WASHINGTON, DC 20202
www.ed.gov

Our mission is to ensure equal access to education and to promote educational excellence throughout the nation.

provided useful feedback that States can use as they revise or develop growth proposals for consideration this fall.

Peer reviewers evaluated State proposals based on seven core principles the Secretary identified to serve as the foundation for States in proposing a growth-based accountability model. These seven core principles ultimately define what the Department considers to be a growth model. It is important to note that while many States responded to GAO's survey as using a growth model, the Department has approved only two to use a growth model for purposes of calculating adequate yearly progress (AYP) under NCLB this year. Only a few other States use some components of growth in making AYP determinations.

The Department is now developing a plan to evaluate rigorously the approved growth models and will share the results with States. The data provided by the pilot project will provide evidence on whether growth models can be fair and reliable methods to measure student improvement and to hold schools accountable for results.

In closing, while we appreciate GAO's concluding observation that the Department "...is poised to gain valuable information on whether or not growth models are overstating progress or whether they appropriately give credit to fast-improving schools," we do have major concerns that the draft report includes text that may confuse readers as to what growth models are. Research used to inform the Department's growth model pilot project only uses the term "growth model" to refer to models that track the growth of individual students. In contrast, the GAO report relies on a very broad definition of growth model, which includes models that compare changes in scores or proficiency levels of schools or groups of students. This results in the report identifying more States as using growth than research would support. This concern and other significant technical comments are enclosed. The Department will continue to support States in their efforts to improve data quality and accountability. Thank you again for the opportunity to comment.

Sincerely,



Henry L. Johnson

Enclosure

Appendix IV: GAO Contact and Staff Acknowledgments

GAO Contact

Marnie S. Shaul, (202) 512-7215, shaulm@gao.gov

Staff Acknowledgments

Blake Ainsworth (Assistant Director), Jason Palmer (Analyst-in-Charge), and Dan Alspaugh (Analyst-in-Charge) managed the assignment. Karen Febey, Shannon Groff, and Robert Miller made significant contributions to this report, in all aspects of the work. Kathy Larin, Harriet Ganson, Lise Levie, Beth Morrison, and Rachael Valliere provided analytic assistance. Luann Moy provided support with the survey. Anna Maria Ortiz and Beverly Ross provided analytic assistance with measuring school results related to key NCLBA goals. Jim Rebbe provided legal support and Mimi Nguyen developed the report's graphics.

Related GAO Products

No Child Left Behind Act: Improved Accessibility to Education's Information Could Help States Further Implement Teacher Qualification Requirements. [GAO-06-25](#). Washington, D.C.: Nov. 21, 2005.

No Child Left Behind Act: Education Could Do More to Help States Better Define Graduation Rates and Improve Knowledge about Intervention Strategies. [GAO-05-879](#). Washington, D.C.: Sept. 20, 2005.

No Child Left Behind Act: Most Students with Disabilities Participated in Statewide Assessments, but Inclusion Options Could Be Improved. [GAO-05-618](#). Washington, D.C.: July 20, 2005.

Charter Schools: To Enhance Education's Monitoring and Research, More Charter School-Level Data Are Needed. [GAO-05-5](#). Washington, D.C.: Jan. 12, 2005.

No Child Left Behind Act: Education Needs to Provide Additional Technical Assistance and Conduct Implementation Studies for School Choice Provision. [GAO-05-7](#). Washington, D.C.: Dec. 10, 2004.

No Child Left Behind Act: Improvements Needed in Education's Process for Tracking States' Implementation of Key Provisions. [GAO-04-734](#). Washington, D.C.: Sept. 30, 2004.

No Child Left Behind Act: Additional Assistance and Research on Effective Strategies Would Help Small Rural Districts. [GAO-04-909](#). Washington, D.C.: Sept. 23, 2004.

Special Education: Additional Assistance and Better Coordination Needed among Education Offices to Help States Meet the NCLBA Teacher Requirements. [GAO-04-659](#). Washington, D.C.: July 15, 2004.

Student Mentoring Programs: Education's Monitoring and Information Sharing Could Be Improved. [GAO-04-581](#). Washington, D.C.: June 25, 2004.

No Child Left Behind Act: More Information Would Help States Determine Which Teachers Are Highly Qualified. [GAO-03-631](#). Washington, D.C.: July 17, 2003.

Title I: Characteristics of Tests Will Influence Expenses; Information Sharing May Help States Realize Efficiencies. [GAO-03-389](#). Washington, D.C.: May 8, 2003.

GAO's Mission

The Government Accountability Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday, GAO posts newly released reports, testimony, and correspondence on its Web site. To have GAO e-mail you a list of newly posted products every afternoon, go to www.gao.gov and select "Subscribe to Updates."

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are \$2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. Government Accountability Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm

E-mail: fraudnet@gao.gov

Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Gloria Jarmon, Managing Director, JarmonG@gao.gov (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, D.C. 20548

Public Affairs

Paul Anderson, Managing Director, AndersonP1@gao.gov (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, D.C. 20548