

# Chapter 12: Student Growth Measures

## Overview

The goal of the Texas Student Assessment Program is to measure and support student progress toward achieving academic success. Its primary purpose is to provide an accurate measure of student achievement in reading, writing, mathematics, social studies, and science. Test performance results are used as a gauge for institutional accountability. Traditionally, the primary emphasis of the assessment program has been on a yearly snapshot of student and institutional performance, as opposed to student and institutional performance across time. Less emphasis has been placed on the assessment program's measures of student growth across testing years. With interest in the measurement of student achievement across time gaining momentum at the state and federal level, likely emphasis on the measurement of growth likely will increase in the future. The purpose of this chapter is to describe where Texas has been, where it is currently, and where it is going with respect to the measurement of student achievement growth.

## Historical Perspective on Growth Measurement in Texas

The following section provides a brief history of the measurement of student progress across time, or growth, in Texas.

### Texas Learning Index

The Texas Learning Index (TLI) was established in spring 1994 to indicate the performance necessary to be likely to succeed on the Texas Assessment of Academic Skills (TAAS) exit level test and to allow comparisons of student learning progress between grades. The TLI was developed to better meet the needs of school districts and students with respect to longitudinal comparability. The TLI metric was considered to have two essential characteristics. First, it provided an index of student progress toward the goal of passing the exit level TAAS test, which students were required to pass in order to be eligible to receive a Texas public high school diploma. Second, the metric permitted comparisons between administrations and between grades for use in the accountability system.

Before the TLI was developed, a vertical scale score system for the TAAS examinations was considered. The vertical scale would have placed results for grades 3–8 and the exit level test on the same scale. At that time, many concerns were expressed regarding a vertical scaling system. First, placing both grade 3 and exit level students on the same scale could lead to misinterpretations because of the large differences in the content of the test items across these grades. Second, a vertical scale implies a linear and well-defined curriculum across grades when such a well-ordered curriculum may not have been in place when the vertical scale was first considered. The TLI was developed as an alternative to the vertical scale due in part to these concerns.

To construct the TLI, the difficulties of the grade 3–8 examinations were aligned with the relative difficulty of the exit level exam. Each subject-area test was designed and constructed at each grade so that answering approximately 70 percent of the items correctly would yield a TLI score of 70. Because the TLI was not a vertical scale, comparisons in growth within a subject area across grades were only valid where growth referred to learning progress “typical” for all students between the grades. The TLI defined typical progress as maintaining the same position across testing years relative to one’s peers. For example, a TLI score of 70 for two years in a row did not mean that the student did not show growth between two consecutive grades. The student’s performance was interpreted as being in about the same position in the second year relative to his or her peers as in the first year, which was considered to be acceptable.

By tracking student performance across testing years, the TLI provided a means for schools to demonstrate improvements in their instructional programs in cases where the passing standard had not yet been met, as well as in cases where the passing standard had been exceeded. Likewise, individual students were able to demonstrate improvement regardless of their most recent achievement relative to the passing standard. For more information on the TLI, see Chapter 10: Scaling of the *2001–2002 Technical Digest* at <http://www.tea.state.tx.us/student.assessment/resources/techdigest/>.

The TLI was specific to the TAAS test. The 76th Texas Legislature in 1999 mandated that TEA develop a new assessment program to replace TAAS. This resulted in the development of the current assessment program, the Texas Assessment of Knowledge and Skills (TAKS), which was first administered to students in the 2002–2003 school year. A more recent measure of growth, the Texas Growth Index (TGI), was developed to measure student progress across years on the TAKS test.

## Texas Growth Index

The TGI was initially implemented in the 2004–2005 school year. The TGI is an estimate of a student’s academic growth on the TAKS tests over two consecutive years and in two consecutive grades. This growth index is used in the state accountability system as the basis for issuance of the Gold Performance Acknowledgments for Comparable Improvement in reading/English language arts (ELA) and in mathematics, which recognize districts and campuses for high academic performance on various indicators, such as commended performance on TAKS, college admissions tests, and advance placement and international baccalaureate examinations. It is also used to calculate the TAKS Progress Indicator in reading/ELA, mathematics, science, and social studies under the alternative education accountability (AEA) procedures. It is intended for use at the campus level and district level and is not intended to be used at the individual student level, to compare classrooms within a campus, or to evaluate teachers.

The TGI uses a formula to describe expected student growth from one year ( $YR_1$ ) to the next ( $YR_2$ ) in a particular subject. The formula uses students’ scores on  $YR_1$  to define students’ expected  $YR_2$  scores. A student’s growth is defined as the student’s  $YR_2$  observed score minus the student’s  $YR_2$  expected score converted to standard deviation

units. If the student's TGI is positive, the student is considered to have exceeded expectations of growth, and the magnitude of the TGI indicates the number of standard deviation units of growth. If the student's adjusted TGI is below the expected score, the student is considered to have fallen short of expectations. For example, a  $-1.00$  adjusted TGI means that the student performed one standard deviation below what was expected. Similarly, a  $1.50$  adjusted TGI means that a student performed 1.5 standard deviations higher than expected.

The TGI formulas initially are applied to scores from individual students prior to aggregation of TGI values to form the TGI for the campus, district, or other larger unit. The TGI for a campus or district is the mean of the student-level TGI scores for all students in the campus or district. A TGI of zero means that the year-to-year change in average scale score is equal to the average predicted change as calculated in the 2003–2004 base comparison years. A positive TGI means the group demonstrated growth that is larger than the expected growth for that group. A negative TGI indicates the group grew less than expected. The TGI was designed for scores to be aggregated at the campus level or at levels larger in size, such as districts; it was not designed as an evaluation tool at the individual student or classroom level.

Once the average TGI of a target campus is determined, it is compared with the TGIs of 40 campuses that are demographically most similar to it. The 40 campuses are arranged from highest to lowest average TGI. If the target campus falls in the top quartile and all other eligibility criteria are met, it is awarded a Gold Performance Acknowledgment for Comparable Improvement. For a more detailed explanation of Gold Performance Acknowledgment, see Chapter 5: Gold Performance Acknowledgments in the *2007 Accountability Manual*, available at <http://www.tea.state.tx.us/perfreport/account/2007/manual/>.

In addition to its use in evaluating campuses for Gold Performance Acknowledgement, the TGI is also used to calculate the TAKS Progress Indicator, which is used in evaluating registered alternative education campuses (AECs). For an explanation of how TGI is used to evaluate AECs, see Chapter 10: AEA Base Indicators in the *2007 Accountability Manual* which can be accessed at <http://www.tea.state.tx.us/perfreport/account/2007/manual/>.

As mentioned above, the TGI was not designed as an evaluation tool at the individual student or classroom levels. In response to recent interest at the state and federal levels in the measurement and evaluation of growth at the individual student and classroom levels, Texas conducted a pilot project, *Measuring Annual Improvement in Student Achievement*, over the course of the 2007–2008 school year. The purpose of the pilot project, described below, was to evaluate the merits of two potential models for measuring growth using TAKS data at the individual student level.

# Measuring Annual Improvement in Student Achievement

## Background

Interest in using student progress or growth information for evaluating campuses/districts and teachers is growing. At a national level, a December 2007 letter by U. S. Secretary of Education Margaret Spellings announced that states would be able to submit proposals for growth-based accountability models. States interested in this program are required to submit proposals to the U.S. Department of Education (USDE) that satisfy seven requirements for using growth measures in the accountability systems. If approved, states could use growth information in their Adequate Yearly Progress (AYP) calculations.

In Texas, both [House Bill 1](#) and [Senate Bill 1031](#) require TEA to measure and report student achievement across time. The instruments used to measure student achievement across time (i.e., statewide standardized tests) must be tied to preparation for mastery of exit level graduation tests. To this end, Texas is required to implement a vertical scale for TAKS in grades 3–8 in reading and mathematics. The EOC assessments required under Senate Bill 1031 also must be developed in a way that allows the measurement of growth.

Many possible methods could be used to measure individual student annual improvement; however, two approaches are currently under consideration by TEA, as these approaches are well-matched to the data conditions in Texas and can be adapted to accommodate upcoming changes to the Texas Student Assessment Program (e.g., the move to a vertical scale in TAKS and EOC assessments). The two approaches are the Reaching the Standard (RTS) method and the SAS<sup>®</sup> EVAAS<sup>®</sup> mixed-model, longitudinal method.

## Reaching the Standard Model

The RTS model provides a yearly growth target for students in each subject and compares students' actual performance to students' target performance to determine if students have progressed adequately over the school year in that subject. Growth targets are set at least twice for all students: once for grades 3 to 8 and again for grades 8 to 11 (exit level). Growth targets are reset for any student who scores below standard in all previous years and then scores at or above Met Standard but below Commended Performance in the current year. Once a student has reached Met Standard but has not reached Commended Performance, that student is expected to maintain scores above Met Standard in subsequent years to be classified as meeting growth targets according to the RTS model. Similarly, growth targets are reset for any student who scores below Commended Performance on all previous tests and then scores at or above Commended Performance in the current year. Once a student reaches Commended Performance, that student is expected to continue scoring at or above Commended Performance level in subsequent years.

The RTS is based on three premises reflecting a focus on student performance in three performance categories:

1. Students who score below Met Standard in grades 3–7 are expected to reach Met Standard by grade 8. Students who score below Met Standard in grades 8–10 are expected to reach Met Standard by grade 11, or exit level.
2. Students who score at or above Met Standard but below Commended Performance in grades 3–7 are expected to maintain their scores above Met Standard each year until grade 8. Students who score at Met Standard but below Commended Performance in grades 8–10 are also expected to maintain their scores above Met Standard each year until grade 11.

As an example, suppose a student scores below Met Standard in grades 3–5 and then scores one standard deviation unit above Met Standard in grade 6. This student must score at least one standard deviation unit above Met Standard in grade 7 to meet the RTS growth target. If the student scores one-half of a standard deviation unit above Met Standard in grade 7, the student will have met the TAKS passing standard but not the RTS growth target.

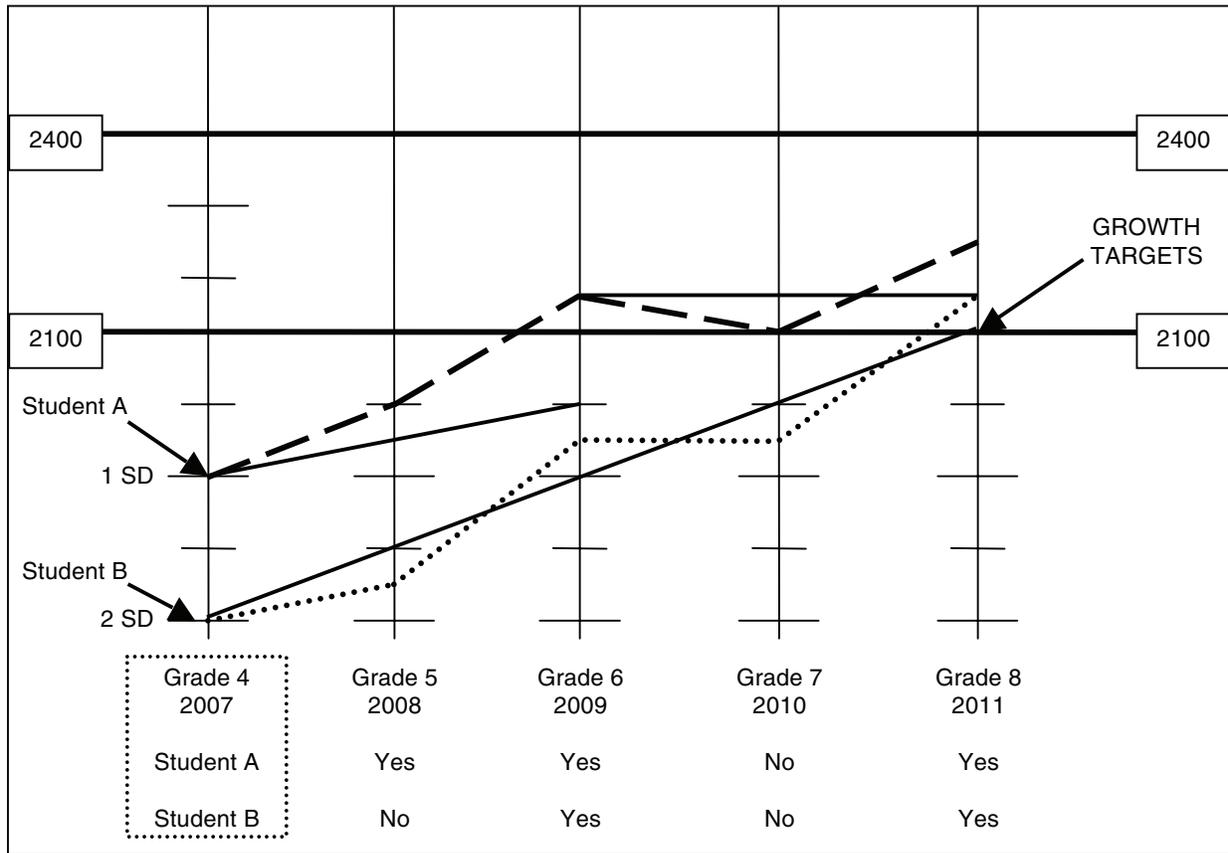
3. Students who score at or above Commended Performance in grades 3 to 7 are expected to score at or above Commended Performance each year until grade 8, and students who score at or above Commended Performance in grades 8–10 are expected to score at or above Commended Performance each year until grade 11. Students who score at or above Commended Performance are expected to continue scoring in that proficiency category. The amount the student scores above Commended Performance is not part of the RTS growth target (unlike RTS growth targets for students who reach Met Standard but not Commended Performance).

Figure 4 shows examples of how the RTS model works for two hypothetical students. As shown, student A was enrolled in grade 4 and scored 1 standard deviation unit below the Met Standard scale score of 2100. Student A had RTS growth targets set at one-fourth of a standard deviation unit closer to Met Standard each successive year until grade 8. Because this student scored one-half standard deviation below Met Standard in grade 5, Student A met the RTS growth expectations that year. Because student A scored above 2100 in grade 6, this student met her/his growth target in grade 6. Furthermore, Student A's growth targets were reset in grade 6, such that he/she was expected to maintain his/her score above 2100 in grades 7 and 8 in order to have met growth targets. The reset targets are indicated by the straight unbroken line. As shown in the graph, Student A met the TAKS standard in grade 7 but did not meet the RTS model target expectations, as this student did not maintain a score above 2100. In grade 8, Student A met growth expectations, as the scores exceeded the growth target.

Student B enrolled in grade 4 and scored 2 standard deviation units below Met Standard, which resulted in growth targets for this student set at one-half of a standard deviation

closer to passing each successive year until grade 8. In grade 5, Student B scored closer to the Met Standard score but did not score one-half standard deviation closer, so Student B did not meet RTS model growth expectations. Student B made enough progress to meet growth expectations in grade 6, but not in grade 7. However, Student B exceeded the passing standard in grade 8; meaning Student B met growth expectations in grade 8.

**Figure 4. RTS Growth Targets Compared with Actual Scores for Two Hypothetical Students**



## SAS Educational Value-Added Assessment System (EVAAS) Longitudinal Models

The EVAAS models estimate the number and percentage of students projected to be proficient in the future. The individual student's projected score is compared to the minimum score for proficiency (the cut score) in a future grade. For example, in the Texas growth pilot study students in grades 5–7 were projected to the English version of the grade 8 reading, mathematics, science, and social studies tests. Students in grades 8–10 were projected to the English version of the grade 11 reading, mathematics, science, and social studies tests. A student needed to have at least three historical scores to receive a projection.

Projections from grades 5–7, using the English test version, were made to grade 8 reading, mathematics, science, and social studies as follows:

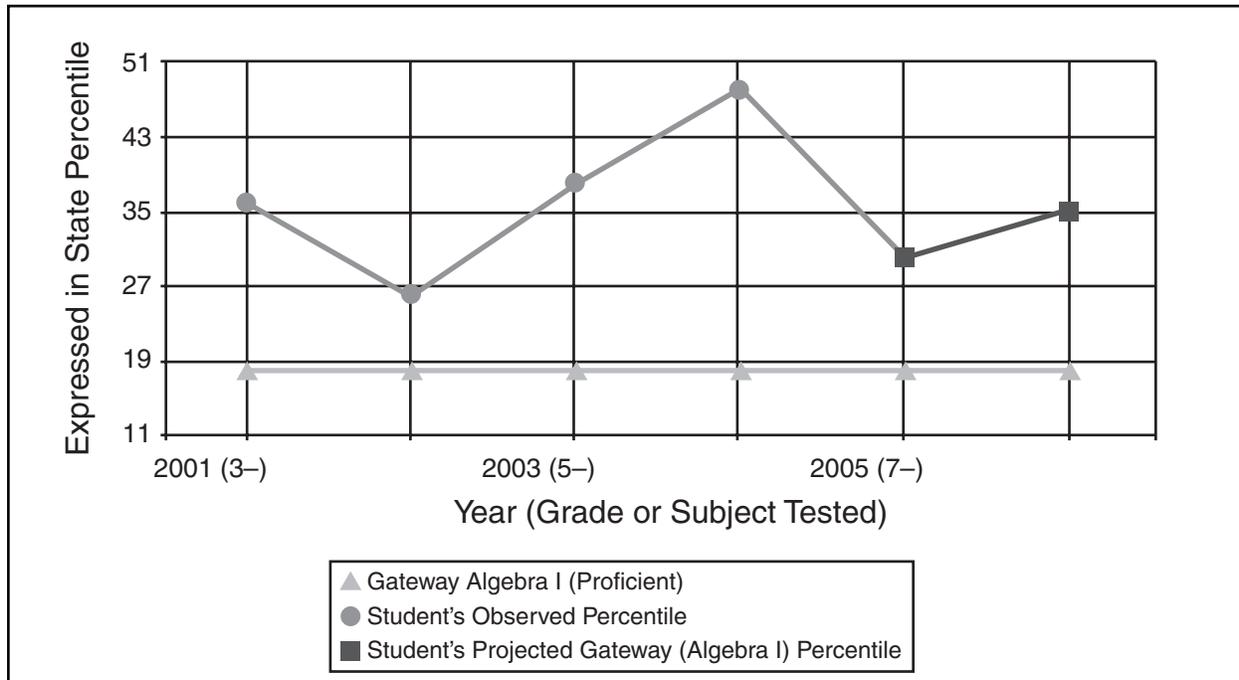
- Projections from grade 7 to grade 8 used grade 7 reading, writing, and mathematics; grade 6 reading and mathematics; grade 5 reading, mathematics, and science; and grade 4 reading, writing, and mathematics.
- Projections from grade 6 to grade 8 used grade 6 reading and mathematics; grade 5 reading, mathematics, and science; and grade 4 reading, writing, and mathematics.
- Projections from grade 5 to grade 8 used grade 5 reading, mathematics, and science; and grade 4 reading, writing, and mathematics.

Projections from grades 8–10, using the English test version, were made to grade 11 reading, mathematics, science, and social studies as follows:

- Projections from grade 10 to grade 11 used grade 10 reading, mathematics, science, and social studies; grade 9 reading and mathematics; grade 8 reading, mathematics, science, and social studies; and grade 7 reading, writing, and mathematics.
- Projections from grade 9 to grade 11 used grade 9 reading and mathematics; grade 8 reading, mathematics, science, and social studies; and grade 7 reading, writing, and mathematics.
- Projections from grade 8 to grade 11 used grade 8 reading, mathematics, science, and social studies; and grade 7 reading, writing, and mathematics.

Figure 5 demonstrates how the projection model is used for a student. The circles connected by lines depict the student's observed scores on a state assessment in mathematics in grades 3–7 from 2001 to 2005. Using the student's mathematics scores, the student's scores in all other subjects, and formulas that take all data from the state into account, the student was projected to score at the 35<sup>th</sup> percentile in the Algebra I assessment in grade 8 in 2006. This projection is higher than the proficiency standard for the algebra assessment, so the student in 2005 was considered to have met growth expectations.

**Figure 5. Report Demonstrating One Student’s Projected Algebra Performance Relative to the Proficiency Standard**



The EVAAS model uses multilevel regression methods to estimate the effects of campuses and districts on the academic growth of students. In some situations, the effects of teachers are also estimated. In this model, the amount of student growth attributable to campuses and districts (and teachers when the teacher effect is included) is considered the value that those campuses and districts add to students’ educational experience. This model uses “shrinkage estimates” for the effects of campuses/districts and teachers. In other words, campuses/districts and teachers will not be considered different from the average campuses/districts and teachers until sufficient data support the difference. When this value-added method is implemented, teacher effects can be calculated only if the teacher information is known for the students. In Texas, student score reports are not tracked by teacher. Instead, student score reports are tracked by header group, which can be, but is not always, grouped by the teacher. More information regarding the EVAAS value-added model can be found at <http://www.sas.com/govedu/edu/services/effectiveness.html>.

### **Pilot Study**

Statewide data from 2007, with historical data from 2004 to 2006, were used in the pilot study. Table 3 summarizes the cohorts for which student progress could be evaluated by either method in the pilot study. Cohort names include the language tested and grade of students in 2007. For each cohort, Table 3 summarizes the years for which data were available and the grades of each cohort of students in those years. Growth targets under each method are set or projected for reaching proficiency in grade 8 from grades 3–7 and for reaching proficiency in grade 11 from grades 8–10. Therefore, for the RTS model, data from some years for some cohorts were not used in the pilot study. For example,

grade 6 and 7 data were not used to evaluate growth for the English grade 9 cohort, or the cohort with students in grade 9 in 2007. However, these data were used with the EVAAS models.

**Table 3. Pilot Study Cohorts**

Cohort Name	Grade of Cohort in Study Years			
	2004	2005	2006	2007
English Grade 4			3	4
English Grade 5		3	4	5
English Grade 6	3	4	5	6
English Grade 7	4	5	6	7
English Grade 8	5	6	7	8
English Grade 9	6*	7*	8	9
English Grade 10	7*	8	9	10
English Grade 11	8	9	10	11
Spanish Grade 4			3	4
Spanish Grade 5		3	4	5
Spanish Grade 6	3	4	5	6

\*Data for these grades were used in EVAAS modeling but were not used in the RTS modeling.

In the pilot study, the following steps were taken:

1. Summary of Pilot Data—summary tables were developed describing the data that were used in the pilot study.
2. Practical and Psychometric Comparison—the two alternative methods were compared on several practical, political, and psychometric features:
  - growth calculations and definitions of growth expectations
  - transparency of calculations and results
  - responsiveness of the growth measures to instructional changes
  - experience implementing the models
  - growth for assessments on different scales
  - retained students
  - adding assessments
  - growth to other targets
  - data collection changes needed
  - method assumptions

- missing data
  - reliability of estimates and accuracy of projections
3. Pilot Growth Results—the methods were compared on the numbers and percentages of statewide students meeting growth expectations and the numbers and percentages of students meeting growth expectations in selected campuses. The numbers and percentages of campuses deemed to be highly effective based on the value-added analyses are also provided.
  4. Reporting Options—ways in which student growth and growth at aggregate levels can be reported for each method were reviewed. A description of reporting options, timing issues related to reporting growth information, and example reports were provided in this part of the study.

Extending Growth to EOC Assessments—a description is provided for how student growth can be calculated under each method as the EOC assessments required for SB 1031 are implemented. This section describes timing issues, scaling considerations, cumulative scoring considerations, and reporting considerations for implementing a growth measure with the EOC assessments.

## Results

The pilot study has been completed, and the results are currently under review by the TEA. The decision regarding the most appropriate model for measuring growth in Texas public education is forthcoming. The new growth model is expected to be implemented in the 2008–2009 school year.