

# CHAPTER 7: ONLINE TESTING

## Introduction

The Texas Education Agency is pursuing initiatives to incorporate online testing into all of its assessment programs—the Texas Assessment of Knowledge and Skills (TAKS), the Reading Proficiency Tests in English (RPTE), and alternative assessments for students with special needs. The motivations for moving to online assessments include greater flexibility in administration, reduced administration burdens on school district personnel, and the development of a more innovative student assessment program. In general, the movement towards electronic testing in K–12 assessment programs is picking up momentum as schools increase their technology capabilities and students become more and more comfortable using the computer for a variety of educational tasks.

Some online testing has been implemented in Texas since fall 2002, when an end-of-course examination in Algebra I was made available online and districts were given the option of using this test in either online or paper format. In spring 2005, a new Algebra I end-of-course (EOC) form that is more aligned with the TAKS mathematics tests at Grades 9, 10, and exit level, was developed exclusively as an online form. A pilot study of online testing at TAKS Grade 8, including mathematics, reading, and social studies, was conducted in 2004. In 2005, the live Grade 8 TAKS tests in these subject areas were offered online to a voluntary sample of school districts, and a study of the comparability of the online and paper versions of the tests was conducted. An online version of the Grade 8 TAKS science field test was also administered. Last, in 2005, an additional testing opportunity was offered for exit level retesters. Each of the exit level tests was offered online and in paper form and a study was conducted to examine the comparability of the tests in the two modes of administration.

In this chapter, data from the various online initiatives outlined above are presented. The first section briefly describes the online delivery system. The second section describes the online Algebra I end-of-course test and summarizes test results from the spring 2005 administration. The third section describes the TAKS online testing initiatives at Grade 8 and exit level. Included in this section is a description of technical methodology used to assess the comparability of online and paper test scores, the results of the TAKS Grade 8 online comparability study, and the results of the online TAKS exit level comparability study. The next section presents an analysis of the item level differences between the two modes, and the final section describes the results of surveys about the TAKS Grade 8 and exit level online tests. These surveys were completed by students and test administrators.

## The Online Test Delivery System

All of the online testing described in this chapter was delivered using Pearson Educational Measurement's eMeasurement system. This system provides a comprehensive set of secure online tools for authoring, delivering, and reporting tests and has been developed to meet the timely delivery and stringent security requirements of the Texas assessment program.

Several key elements have been included in the system's design so that it meets the needs of the state's programs. A logon ID and password is required to enter the system. No item images or response data are stored on the hard drives of the PCs being used for testing. Once a testing session has started, the testing software locks various aspects of the system so that items cannot be copied or printed. Students can access formula sheets, calculators, or other required aids, as determined for each test. When an item includes a reading passage or other stimulus, the item appears on the screen together with the passage or stimulus. The system also includes utilities for test administrators that allow them to control which tests will be administered when and which students will be in each testing session.

Further information about the eMeasurement system can be found at <http://etest.pearson.com>, including an overview of the system, information on delivery and reporting, and a list of frequently asked questions.

# Algebra I End-of-Course

## Overview

The Algebra I end-of-course (EOC) test was revised in 2005 to be better aligned to the TAKS Grades 9, 10, and exit level mathematics tests. A new test blueprint was developed containing the same five algebra objectives that are on the TAKS Grades 9, 10, and exit level TAKS test blueprints. The new Algebra I EOC test blueprint is shown below.

**Table 7. ALGEBRA I END-OF-COURSE TEST BLUEPRINT**

<b>Objectives</b>	<b>Number of Items</b>
Objective 1: Functional Relationships	8
Objective 2: Properties and Attributes of Functions	11
Objective 3: Linear Functions	11
Objective 4: Linear Equations and Inequalities	11
Objective 5: Quadratic and Other Nonlinear Functions	9
<b>Total number of items</b>	<b>50</b>

The new version of the Algebra I EOC test is only available online and was administered for the first time in 2005 during the April 25–June 24, 2005, testing window. No performance standards have been set. A standard setting is planned for November 2005. For this reason, a scale score system does not currently exist for the Algebra I EOC test, and scores are reported on the raw score scale. Descriptive statistics and the raw score distribution for the spring administration can be found in Appendix 10.

## Relationship to TAKS

A number of students who took the TAKS Grade 9 or Grade 10 mathematics test also took the online Algebra I EOC test. A study was conducted to examine the relationship between performance on the Algebra I test and that on the TAKS tests in Grades 9 and 10 to help educators better understand the usefulness of the Algebra I scores. The Pearson correlations between the Algebra I test and the Grades 9 and 10 TAKS mathematics tests were 0.71 and 0.55, respectively.

## TAKS: Grade 8 and Exit Level Retests

The Grade 8 TAKS tests—mathematics, reading, and social studies—were administered as live online tests during a window the week before the scheduled spring 2005 paper tests. The Grade 8 TAKS science field test was also administered online during the same testing window. The exit level tests—ELA, mathematics, science, and social studies—were administered online and in paper form in June 2005 to examinees retesting in any subject area in which they had

not previously met the passing standard. With the exception of the Grade 8 TAKS science field test, the online and paper versions of the tests were identical. Since the online test were live and counted in the same manner as the results for students who took the paper versions, it was necessary to conduct research studies to ensure that the paper and online results were comparable and did not provide an advantage or disadvantage to students who tested in either mode. Surveys were also administered to students and test administrators to gather information about the online testing experience.

## Comparability Analyses

The issue of comparability between online and paper tests has several facets. When the same test is administered in both delivery modes, studies should be conducted to determine whether the use of a single score table is warranted. If mode effects are detected, it may be necessary to use a separate score table for each mode of delivery.

The approach used to assess comparability for the TAKS Grade 8 and exit level tests was a variation of one outlined by Dorans and Lawrence (1990). Their approach was designed to check the statistical equivalence of nearly identical test forms and did so by evaluating differences in the raw score-to-scale score conversion tables. In this context, the evaluation was between the online and paper modes. In some equating designs (e.g., linear or equipercentile equating), standard errors of equating can be calculated through known formulas. For tests where equating is done using the Rasch model, formulas for calculating standard errors of equating are not available. The bootstrap method (c.f., Kolen & Brennan, 2004, p. 232–235) is a useful procedure to calculate standard errors of equating using the relevant Rasch model. These standard errors can then be used to evaluate an equating between the online group and a paper group.

To accurately examine the comparability of the paper and online versions of a test, the groups of students taking the test in the two modes must be assumed comparable on the skill being measured by the test. If the two groups are not equivalent on the skill being measured it is not possible to separate mode form differences due to the skill levels of the groups. There are two ways to achieve group equivalence: one is to randomly assign students to either the paper or online testing condition; the other is to match each student in the online condition to a student in the paper condition on the basis of relevant variables such as previous test scores.

The steps used to examine the comparability of the TAKS Grade 8 and exit level online and paper test are outlined below.

1. The paper version of the test is calibrated and equated to the reporting scale using standard procedures for the paper test. This results in the raw score to scale score table for the paper test.
2. A random sample of students is drawn with replacement from the online group of students. The sample is the same size as the online group.

3. A sample of students is drawn from the paper group. This may be done in one of two ways:
  - a. If students were randomly assigned to a testing condition, then randomly select, with replacement, a sample of students from the paper group.
  - b. If students are being matched, then match each student in the online sample from step 2 to a student from the statewide administration of the paper test. The matching variables are prior year test scores.
4. The test items are calibrated separately for the online sample and the paper sample centering on people (i.e., the mean ability in each group is set to zero).
5. The theta estimate for each raw score in the online group is used to obtain an estimated raw score using the item parameters from the calibration of the paper test group. These are the equated raw scores for the online group on the scale of the paper test.
6. The equated raw scores for the online group are transformed to scale scores using the paper raw scores, corresponding scale scores from step 1, and linear interpolation. These are the scale scores for the online group on the scale of the paper test.
7. Steps 2 through 6 are repeated a large number of times, typically 500 or 1,000. Note that these bootstrap replications incorporate the error in selecting the matched samples as well as equating error.
8. The average of the equated scale scores at each raw score for the online group over the replications comprises the online scale score table.
9. The standard deviation of online scale score conversions at each raw score represented the conditional bootstrap standard errors of the equating.
10. Raw score points for which the difference between the online and paper scale score conversions exceed two standard errors of the equating indicate a significant mode effect.

## TAKS Grade 8

The online versions of the TAKS Grade 8 mathematics, reading, science, and social studies field tests were administered during the period April 12–15, 2005. This was the week prior to the statewide administration of the paper tests. A total of 109 campuses participated in the study—20 in reading (1,842 students), 16 in mathematics (1,273 students), 15 in social studies (1,451 students), and 58 in science (6,173 students). In order to conduct the comparability study, students who participated in the online administration were matched on their seventh-grade reading and mathematics test scores to students who took the same test during the statewide paper administration. The comparability analyses described in the previous section were independently performed by two PEM psychometricians. Tables 1 through 3 in Appendix 10 summarize the comparability analysis results for mathematics, reading, and social studies.

For mathematics (Table 1), the comparability analyses indicated that the online version of the test was slightly more difficult than the paper version, although differences in equating score conversions were not statistically significant at most score points. In terms of the raw score conversions, the differences were less than one-half of a point. In terms of scaled score conversions, the differences were less than five points over most of the scale. At the upper raw score points (41 and above), the results indicated that scale score conversions differed by more than two standard errors of the linking. However, the online raw score equivalents at these score levels were actually closer to the paper raw scores than they were at other parts of the score scale where scale score differences did not exceed two standard errors of the linking (i.e., raw scores between 27 and 40). In addition, the "Commended" level corresponded to the same raw score (44) based on either the paper or online scale score conversions. Given that the online and paper performance differences were not statistically significant at most score points, a decision was made to use the paper score conversion table for the students testing online.

For reading (Table 2 in Appendix 10), large differences between online and paper test performance occurred throughout the scale, indicating that the online version of the test was more difficult than the paper version. Differences in raw score conversions exceeded one and a half points over much of the score range. Differences in scale score conversions were over 20 points over most of the score range. All of the differences in scale score conversions exceeded two standard errors of the linking. These results indicated that the scores on the online version of the Grade 8 reading test were not comparable to the scores based on the paper version. As a result, a separate score conversion table was used for Grade 8 TAKS reading, based on the computer test scale scores calculated in Table 2.

For social studies (Table 3) slight differences in both raw score and scale score conversions occurred, indicating that the online version of the test was slightly more difficult than the paper version. The raw score differences were less than one-half of a point, and the scale score differences were less than six points. None of the scale score differences exceeded two standard errors of the linking. Given that the online and paper performance differences were not statistically significant at all score points, a decision was made to use the paper score conversion table for the students taking the Grade 8 social studies test online.

## **Exit Level**

In June 2005, selected campuses across the state participated in a comparability study for the exit level TAKS (English language arts, mathematics, science, and social studies retest). Students in the participating schools who failed to meet the passing standard on one or more of the spring 2005 exit level TAKS tests were eligible to participate in the paper/online study. In addition, participating campuses were permitted to nominate a limited number of students who would enter Grade 11 in fall 2005 to participate in this study.

Participating students were randomly assigned to take either an online or paper version of each subject-area TAKS test. All results for students taking this administration of the exit level

TAKS tests served as official scores, and statistical analyses (equating) were used to ensure that participating students were not advantaged or disadvantaged by testing online. There were 1,002 students who took the mathematics test, 695 who took the ELA test online, 1,014 who took the science test online, and 373 who took the social studies test online.

The results of the comparability analyses are shown in tables 4 through 7 in Appendix 10. The tables indicate four separate performance standards for each test. There is one standard for Commended Performance and three standards for Met Standard—2 SEM, 1 SEM, and Panel Recommended (PR).

For mathematics and science (Tables 4 and 7), the online versions of the tests were more difficult than the paper versions. For social studies (Table 6), the online version of the test was slightly less difficult than the paper version. For English language arts (Table 5), the online version of the test was more difficult than the paper version over most of the score range.

For mathematics (Table 4), differences in raw score conversions were between one-half and three-quarters of a point (in favor of the paper group) for the middle 50 percent of the score range. Differences in scale score conversions exceeded 10 points and were statistically significant in the upper 25 percent of the score range.

For English language arts (Table 5), differences occurred throughout the scale. In the lower 85 percent of the score range, differences in raw score conversions exceeded one point (in favor of the paper group) over much of the score range. Differences in the scale scores were between 10 and 20 points over most of this score range. In the upper 15 percent of the score range, differences in both the raw score and scale scores favored the online group. Although these differences (especially for the scale scores) appear quite dramatic, they are at least in part attributable to the impact of weighting the essay prompt by 4. While none of the scale score differences (in any part of the score range) exceeded two standard errors of the linking, the differences were very close to doing so across a large portion of the score range. Additionally, the standard deviations of the equivalent scale scores over the 500 replications were quite large.

For social studies (Table 6), small differences in both raw score and scale score conversions also occurred. The raw score differences were less than half a point (in favor of the online group) and the scale score differences were less than seven points for most of the score range. None of the scale score differences exceeded two standard errors of the linking.

For science (Table 7), small differences in both raw score and scale score conversions occurred. The raw score differences were less than half a point (in favor of the paper group) and the scale score differences were less than seven points for most of the score range. None of the scale score differences exceeded two standard errors of the linking.

Based on the results of the comparability analyses, separate June 2005 score conversion tables were used for students taking mathematics or English language arts online. These score conversion tables are contained in Appendix 10. For science and social studies, the results indicated that using the paper conversion table for online students was justified.

## Item-Level Analyses

One advantage of the comparability designs implemented in spring 2005 is that the same items were administered online and in paper. This provided data for direct item-level comparisons across testing modes. The purpose of these analyses was to identify any patterns or similarities among items which showed a mode effect and to provide additional information which might shed light on our overall test level comparability findings.

The analyses consisted of examining the following statistics.

1. **Comparison of item p-values.** Item p-values were computed separately for the online and paper conditions and compared using an independent samples t-test. An alpha level of 0.05 with a Bonferroni correction for Type I error was used to evaluate all results. Additionally, differences in item p-value by mode were examined in relationship to sequential position of the item on the test.
2. **Comparison of response distributions.** Response distributions for each item were computed separately for the online and paper conditions and compared using a chi-square test. An alpha level of 0.05 with a Bonferroni correction for Type I error was used to evaluate all results.
3. **Differential Item Functioning (DIF) by testing mode.** DIF analyses by testing mode were conducted using the DIF command in WINSTEPS. Using theta rather than raw scores as the conditioning variable, WINSTEPS produces a t-statistic that is equivalent to the Mantel-Haenszel significance test. An alpha level of 0.05 with a Bonferroni correction for Type I error was used to evaluate all results. WINSTEPS computes the Rasch item difficulty for the item of interest separately for the focal and reference groups (holding all other item difficulties and all theta estimates fixed at their values from the main calibration). Next, the item difficulty values for the focal and reference groups are subtracted from each other and divided by the joint standard error to produce the t-statistic.
4. **Comparison of performance on objective scores.** Objective scores were computed separately for the online and paper conditions and compared using an independent samples t-test. An alpha level of 0.05 with a Bonferroni correction for Type I error was used to evaluate all results.

The full results from all item level analyses for the TAKS Grade 8 and exit level tests are provided in Appendix 11. Items are highlighted in bold where significant mode effects were found. A summary of the results is provided here.

Looking across grades, the results seem to suggest a pattern of item by mode effects for mathematics items involving graphing and geometric manipulations. In these cases, a common strategy employed by students to draw on the item to solve the problem was not readily available in the online mode. Research by Greenwood, Cole, McBride, Morrison, Cowan, & Lee (2000) suggested that items were harder on the computer when they required spatial or

gross motor skills. Additionally, research from the National Assessment of Educational Progress project Math Online (MOL) (NAEP, 2001) hypothesizes that items that require students to make a scale drawing or work with geometric shapes are likely to show mode effects.

The findings for TAKS Grade 8 reading largely reinforce the results of the overall test comparability study with significant differences favoring paper occurring at both the item and objective levels. Findings for Grade 11 TAKS English language arts suggest that the majority of multiple-choice items on this test did not exhibit mode effects. However, there was a significant difference in the scores on the essay between the paper and online groups. Approximately 10 percent more students received a score of 2 or more in the paper group than in the online group. The difference in findings across grades with respect to the multiple-choice items could reflect differences in the presentation of reading passages between the two online administrations. For TAKS Grade 8, reading passages were presented in the upper pane of a horizontally split screen and questions were presented in the lower pane. The examinees could adjust the relative size of the panes as they preferred. For TAKS Grade 11, the reading stimulus materials (consisting of two passages and a graphical image or picture) were first presented one after the other, and the test taker was required to consider these materials before they answered questions. Each question was presented independently, and if the examinees wished to refer back to the reading stimulus materials, they could open them in separate exhibit windows. Further research is needed to examine this hypothesis in more detail.

The results from science and social studies are more difficult to interpret. The TAKS Grade 11 analyses showed no item or objective level mode effects for either subject area. This is consistent with the results of the overall test comparability study. At TAKS Grade 8, however, differences were seen in both subject areas, but were not clearly interpretable. For social studies, a large number of items were flagged as significant according to one or more of the three statistical tests. However, these items were split approximately 50/50 in terms of favoring paper or favoring online testers.

## Survey Results

Each student who took an online test was administered a survey about their online testing experience immediately after the completion of the test. The survey automatically appeared as a separate section of the test following the last test question. Additionally, an e-mail sent to test administrators shortly after the testing period ended contained a link to an online survey. The survey questions and tallies of responses are contained in Appendix 11.

Student responses to online testing were generally positive. Most students agreed that the test directions were easy to understand and questions were easy to work through. The majority of students rated themselves as having average to above average computer skills and reported that they had used computers in their coursework and outside of school. Test

administrators were also positive about online testing. The majority of test administrators responding to this survey reported that they had sufficient computers available for online testing and that test security was easy to maintain. They agreed that having fewer materials to manage was an advantage of online testing over paper testing.

## Summary

This chapter describes several online testing initiatives that were undertaken by the Texas Education Agency during the 2004–2005 school year. These included the development and online administration of a new Algebra I end-of-course test as well as studies of online and paper test comparability at TAKS Grade 8 and exit level.

In general, these online testing initiatives were well received by the participating schools and students. Surveys indicated that the online testing initiatives at TAKS Grade 8 and exit level were well received by students and testing administrators. The comparability analyses indicated some mode-of-administration effects at both TAKS Grade 8 and exit level. In TAKS Grade 8 reading, students testing online scored significantly lower than comparability groups of students testing on paper, resulting in the use of a separate raw-to-scale score conversion table for the online students. For TAKS Grade 8 mathematics and social studies, online test results did not differ significantly from the paper results and the observed differences were judged to be small enough to justify the use of a common score conversion table for both testing modes. For exit level TAKS, lower performance on the online versions of the mathematics and English language arts tests justified the use of separate raw-to-scale score conversions for students testing online in these tests. However, for science and social studies, no significant mode effects were found and a common raw-to-scale score conversion table was used for students testing both online and by paper in these content areas.

The mixed comparability results at TAKS Grade 8 and exit level suggest the need for additional comparability research. This research will include more in-depth analyses to investigate causes of mode differences and additional comparability studies as online testing with TAKS continues in the future.

# CHAPTER 8: STUDENT SUCCESS INITIATIVE (SSI)

## Background

The Student Success Initiative includes three initiatives that together provide a system of academic support to help ensure achievement on grade level in reading and mathematics so that every student can succeed throughout his or her school career. The SSI incorporates a grade advancement component adopted by the Texas Legislature in 1999, which is being implemented on a phase-in schedule. The law ties promotion to performance on state-mandated assessments in reading at Grade 3, reading and mathematics at Grade 5, and reading and mathematics at Grade 8 (beginning in 2007–2008). The law applies to students who take the TAKS tests in English, the TAKS tests in Spanish, or SDAA II.

Students must be allowed at least three opportunities to pass a test under the SSI grade advancement requirements. A student may advance to the next grade level only by passing these tests. If a student has not passed after three opportunities and a parent appeals, the student may be promoted by unanimous decision of a grade placement committee (GPC) consisting of the principal or the principal's designee, the student's parent or guardian, and the student's teacher for the subject area tested. The GPC may advance a student only if they determine that he or she is likely to perform at grade level after accelerated instruction.

The academic support provided under the Student Success Initiative takes many forms. Students who are identified as being at risk of failing an SSI assessment must be provided extra instruction prior to the first test administration. Students who do not pass an SSI assessment must be provided additional instruction before each subsequent testing opportunity. In addition, a copy of a TAKS study guide is provided by the state to any student who does not meet the standard on a TAKS assessment. Funding and instructional support for the Student Success Initiative are also provided through the Texas Reading Initiative and Texas Mathematics Initiative.

## The Grade Placement Committee

The *Grade Placement Committee Manual for Grade Advancement Requirements of the Student Success Initiative* provides detailed information about the grade advancement and accelerated instruction requirements of the law. The manual is designed to guide GPCs through the process of evaluating student performance, determining accelerated instruction, and, when applicable, making promotion decisions. The manual also provides districts with detailed instructions for implementing all the requirements of the SSI and includes time lines and flow charts as well as sample forms and letters to aid districts with parental notification and documentation of activities. The grade placement committee manual and other information related to SSI requirements are located on the Student Success Initiative section of the Student Assessment Division website. This section also includes links to other agency websites, such as the Reading Initiative and Mathematics Initiative.

## Students Receiving Special Education Services

The Student Success Initiative regulations also apply to students who are receiving special education services, whether they take TAKS or SDAA II. The grade advancement requirements differ depending on the assessment in which the student participates. The admission, review, and dismissal (ARD) committee makes all assessment decisions for students served by special education. The ARD committee also serves as the GPC for students receiving special education services who fail to perform satisfactorily on any SSI assessment. ARD committee members are responsible for making all decisions about accelerated instruction and grade advancement.

In 2005 the SSI grade advancement requirements applied to all enrolled Grade 3 students receiving special education services who took the Grade 3 TAKS reading test or an SDAA II reading test and to all enrolled Grade 5 students receiving special education services who took the Grade 5 TAKS reading or mathematics test or an SDAA II test in reading or mathematics. Because there is only one test administration for each SDAA II instructional level and content area annually, districts may administer TEKS-based formal or informal assessments to students for whom SDAA II is deemed appropriate so that those students may have multiple testing opportunities to meet SSI requirements.

The *ARD Committee Decision-Making Process for the Texas Assessment Program* training manual and the grade placement committee manual include sections on SSI requirements for students who receive special education services. Both manuals are available on the Student Assessment Division website.

## Alternate Assessment Option

State law allows districts to administer an alternate assessment approved by TEA in lieu of TAKS to students who have failed an SSI assessment on the previous two testing opportunities. The grade placement committee is responsible for deciding whether a student will take an alternate test. However, the local school district must first determine whether this option will be made available to students in the district.

The Student Assessment Division allowed districts to administer the alternate test on June 29, 2005. The Iowa Test of Basic Skills (ITBS) Level 9 was the only alternate assessment approved for use in 2004 and 2005. Using data from a study conducted by Riverside Publishing in 2003 that links TAKS and ITBS, the Student Assessment Division determined a comparable scale score passing standard for the ITBS that reflects the 2005 TAKS Grade 3 reading scale score passing standard.