

# CHAPTER 7: ONLINE TESTING

## Introduction

The Texas Education Agency is pursuing initiatives to incorporate online testing into most 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 district personnel, and enhanced security. In general, the movement toward 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.

Several online testing initiatives have 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) exam aligned with the TAKS mathematics tests at grades 9, 10, and exit level was developed exclusively in online form. A pilot study of online testing of TAKS grade 8, including mathematics, reading, and social studies, was conducted in 2004. In 2005 the live TAKS grade 8 tests in these subject areas were offered online to a voluntary sample of districts, and a study of the comparability of the online and paper versions of the tests was conducted. An online version of the TAKS grade 8 science field test was also administered. Finally, in June 2005 an additional online testing opportunity was offered for exit level retesters. Each exit level retest was offered online and in paper form, and a study was conducted to examine the comparability of these two modes of administration.

Online testing continued throughout the 2005–2006 school year, with another opportunity for retesters to take the TAKS exit level tests online in December 2005. Comparability studies were conducted again in spring 2006 for TAKS grade 8 reading, mathematics, and social studies. In addition, comparability studies were expanded to include TAKS grade 8 science as well as TAKS grade 9 reading and mathematics. An online pilot administration of the RPTE II field test for grade 2 students was administered during a spring testing window as a first step in transitioning to a fully online RPTE II assessment. Finally, the TAKS exit level retest administration conducted in July 2006, allowed districts to test students in both online and paper modes.

In this chapter, data from several of the online initiatives outlined above are presented. The first section briefly describes the online delivery system while the second section describes the online Algebra I End-of-Course exam and performance standards. The third section describes the TAKS online testing initiatives at grade 8, grade 9, and exit level. Included in this section is a description of technical methodology used to assess the comparability of online and paper test scores, and the results of the TAKS grade 8, grade 9, and July exit level retest online comparability studies. Information about the spring 2006 grade 2 online RPTE II feasibility pilot is found in Appendix 9.

## **The Online Test Delivery System**

All 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 results of tests and has been developed to meet the stringent requirements of the Texas Student Assessment Program and protect the integrity of test items and student data.

Several key elements have been included in the system's design so that it meets the needs of the state's programs. The system was designed to take advantage of existing hardware and software that are already installed in schools. Access is controlled through user IDs and passwords. All transmissions are encrypted, and no test questions or responses are stored on the local workstation when testing concludes. Once a testing session has started, the software locks down the workstation to prevent items from being copied, printed or e-mailed and to prohibit the use of unauthorized applications. Students can access formula charts, calculators, or other required aids, as determined for each test. When an item includes a reading passage or other stimulus, the passage or stimulus appears on the screen together with the item or it is displayed in a separate window. The system also allows test administrators to control which tests will be administered when and which students will be in each testing session. While the test is in progress, a student's current status can be monitored from the test administrator's workstation.

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 Exam**

The Algebra I End-of-Course (EOC) exam was revised in 2005 to be better aligned to the TAKS grades 9, 10, and exit level mathematics tests. More specifically, a new test blueprint was developed containing the same five algebra objectives that are on the TAKS grades 9, 10, and exit level test blueprints.

The new version of the Algebra I EOC exam is only available online and was administered for the first time in spring 2005 and then again in both fall 2005 and spring 2006. Performance standards were set in November 2005. For more information on the Algebra I EOC exam standards, see Chapter 12: Standards.

## **TAKS: Grade 8, Grade 9, and Exit Level**

The TAKS grade 8 reading, mathematics, science, and social studies tests and the TAKS grade 9 mathematics tests were administered as live online tests during a week-long window preceding the scheduled spring 2006 paper administrations. TAKS grade 9 reading was administered on the same day as the paper test. The TAKS exit level retests in English language arts, mathematics, science, and social studies were each administered as live online tests in July on the same day as the paper tests. For each administration, the online and paper versions

of the tests included identical items. Since the online tests 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 advantage or disadvantage 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, grade 9, 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 the context of the grade 8, grade 9, and exit level tests, the evaluation was between the online and paper modes. In some equating designs (for example, linear or equipercentile equating), standard errors of equating can be calculated using 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 (see Kolen & Brennan, 2004, pp. 232–235) is a useful procedure for calculating 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 isolate mode differences. 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 grade 9 online and paper tests are outlined below.

1. The paper version of each test is calibrated and equated to the reporting scale using standard procedures for all TAKS assessments. This results in a raw score to scale score conversion table for each paper test. In the case of the July retests, a pre-equated raw score to scale score conversion table was used.
2. A random sample of students is drawn with replacement from the online group of students. To estimate sampling error, the sample is the same size as the online group.

3. A sample of students is drawn from the paper group. Each student drawn from the paper administration of the statewide test is matched to a student in the online sample from step 2. The matching variables include gender, ethnicity, and prior or current year test scores.
4. The test items are calibrated separately for the online sample and the paper sample centering on people (that is, 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 frequently, typically 100 to 500 times. Note that these bootstrap replications incorporate the error in selecting the matched samples as well as the 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 represents the conditional bootstrap standard errors of the equating.
10. Raw score points for which the difference between the online and paper scale score conversions exceeds two standard errors of the equating indicate a significant mode effect (Dorans & Lawrence, 1990).

## **TAKS Grade 8**

The online versions of the TAKS grade 8 reading, mathematics, science, and social studies tests were administered April 10–14, 2006. This was the week prior to the statewide administration of the paper tests. Campuses could test in multiple subjects—23 participated in reading (1,080 students), 19 in mathematics (494 students), 42 in science (2,285 students), and 33 in social studies (1,427 students). In order to conduct the comparability study, students who participated in the online administration were matched on their previous or current reading and mathematics test scores to students who took the same test during the statewide paper administration; students were also matched on gender and ethnicity. Table 7 on the following page shows the demographic information for the online and paper testing groups.

**Table 7. Demographic Information of TAKS Grade 8 Online and Paper Groups**

Mode	Subject	Male	White	Hispanic	African American	Other Ethnicity
Online	Reading	46.48%	54.63%	40.93%	3.61%	0.83%
	Mathematics	47.98%	74.70%	23.08%	1.42%	0.81%
	Social Studies	47.02%	56.48%	37.77%	4.41%	1.33%
	Science	48.27%	52.47%	36.81%	9.28%	1.44%
Paper	Reading	49.04%	41.57%	42.47%	12.53%	3.43%
	Mathematics	49.02%	41.59%	42.49%	12.49%	3.43%
	Social Studies	49.04%	41.56%	42.52%	12.48%	3.44%
	Science	49.01%	41.52%	42.50%	12.54%	3.44%

The comparability analyses described in the previous section were independently performed by two Pearson Educational Measurement psychometricians. Tables 1 through 4 in Appendix 18 summarize the comparability analysis results for reading, mathematics, science, and social studies.

For reading (Table 1 in Appendix 18), the comparability analyses indicated that the online version of the test was more difficult than the paper version, although differences in equating score conversions were not statistically significant at many score points. At the upper raw score points (36 and above), the results indicated that scale score conversions differed by more than two standard errors of the linking. In terms of the raw score conversions, the differences were between one-half and two-thirds of a raw score point. These results indicated that the scores on the online version of the test were not comparable to scores on the paper version. As a result, a separate score conversion table was used based on the online scale scores calculated in Table 1.

The mathematics results (Table 2 in Appendix 18) indicate that large differences occurred throughout the scale. Differences in raw score conversions exceeded two points over much of the score range. All differences in scale score conversions exceeded two standard errors of the linking. A separate score conversion table was used based on the online scale scores calculated in Table 2.

The results for science (Table 3 in Appendix 18) show that differences in the raw score conversions were one-half point or less throughout the score scale. In terms of scale score conversions, the differences were between 6 and 12 points over most of the scale. A separate score conversion table was used based on the online scale scores calculated in Table 3.

For social studies (Table 4 in Appendix 18), the results show large differences throughout the scale. Differences in raw score conversions exceeded one-half of a raw score point over nearly the entire score range. All differences in scale score conversions exceeded two standard errors of the linking. A separate score conversion table was used based on the online scale scores calculated in Table 4.

## TAKS Grade 9

The online version of the TAKS grade 9 reading test was offered on the February 21, 2006, paper administration date only. The TAKS grade 9 mathematics test was offered online in a testing window (April 10–14) one week before the April 20, 2006, paper administration date. Campuses could test in multiple subjects—81 participated in reading (2,882 students) and 50 participated in mathematics (3,158 students). In order to conduct the comparability study, students who participated in the online administration were matched on their previous reading and mathematics test scores, as well as on gender and ethnicity, to students who took the same test during the statewide paper administration. Table 8 shows the demographic information for the online and paper testing groups.

**Table 8. Demographic Information of TAKS Grade 9 Online and Paper Groups**

Mode	Subject	Male	White	Hispanic	African American	Other Ethnicity
Online	Reading	48.47%	54.34%	31.54%	12.46%	1.67%
	Mathematics	49.37%	55.45%	31.95%	10.35%	2.25%
Paper	Reading	49.21%	41.29%	41.43%	13.76%	3.52%
	Mathematics	49.05%	43.56%	39.90%	13.19%	3.36%

The comparability analyses described in the previous section were independently performed by two PEM psychometricians. Tables 5 and 6 in Appendix 18 summarize the comparability analysis results for grade 9 reading and mathematics.

The results for reading (Table 5 in Appendix 18) indicate that the online version of the test was more difficult than the paper version. Differences in scale scores between online and paper versions ranged from 10 to 15 scale score points throughout most of the scale. These differences also exceeded two standard errors of the linking throughout most of the scale. Raw score differences exceeded half a raw score point at portions of the scale. A separate score conversion table was used based on the online scale scores calculated in Table 5. Note that on Table 5 the scale scores associated with the online raw scores of 40, 41, and 42 were the same values as those for the paper test. Since no students in the online sample attained a score point of 3 on any of the open-ended items, none attained a raw score of 40, 41, or 42 on the overall test. Thus the online scale scores computed by the comparability method at these raw score points are less stable. Providing the same scale scores at the top three raw score points for the online and paper tests avoids reporting large differences in scale scores between the two modes in an area of the scale in which no student data were present for the online condition.

For mathematics (Table 6 in Appendix 18), results also indicate that the online version of the test was more difficult than the paper version. Differences in scale scores between online and paper versions ranged from 25 to 30 scale score points throughout most of the scale. These differences also exceeded two standard errors of the linking throughout the entire scale. Raw score differences exceeded one raw score point across most of the scale. A separate score conversion table was used based on the online scale scores calculated in Table 6.

# TAKS Exit Level

The online versions of the TAKS exit level July retest in English language arts (ELA), mathematics, science, and social studies were administered on July 11, 12, 13, and 14, respectively. Each was administered on the same day as the statewide administration of the paper tests. Campuses could test in multiple subjects—114 participated in ELA (964 students), 135 in mathematics (2,146 students), 146 in science (2,856 students), and 103 in social studies (587 students). In order to conduct the comparability study, students who participated in the online administration were matched on their previous ELA, mathematics, science, and social studies test scores to students who took the same test during the statewide paper administration, as well as being matched on gender and ethnicity. The demographic information for the online and paper testing groups is found in Table 9.

**Table 9:  
Demographic Information of TAKS July Exit Level Online and Paper Samples**

Mode	Subject	Male	White	Hispanic	African American	Other Ethnicity
Online	ELA	59.02%	33.20%	46.78%	17.01%	3.01%
	Mathematics	39.16%	29.26%	46.27%	22.89%	1.57%
	Science	34.59%	25.84%	50.25%	22.09%	1.82%
	Social Studies	34.24%	17.38%	59.11%	21.64%	1.87%
Paper	ELA	56.39%	25.10%	54.04%	18.49%	2.38%
	Mathematics	39.65%	24.83%	48.26%	25.36%	1.55%
	Science	36.00%	21.32%	53.09%	23.55%	2.04%
	Social Studies	35.03%	16.82%	61.73%	19.24%	2.21%

The comparability analyses were independently performed by two PEM psychometricians. Tables 7 through 10 in Appendix 18 summarize the comparability analysis results for ELA, mathematics, science, and social studies.

The results for ELA (Table 7 of Appendix 18) indicate that the online version was comparable to the paper version. The scale score differences did not exceed two standard errors of the linking for two sections of the raw score scale. In the lower end, differences favored the paper group, and at the upper end, the differences favored the online group. In the middle of the raw score range, no mode effects were found. The same score conversion table was used for paper and online.

The results for mathematics (Table 8 of Appendix 18) indicate that the online version was more difficult than the paper version. The differences in the raw score conversions were between one-half and three-fourths of a raw score point. Across the entire raw score range, the scale score differences exceeded two standard errors of the linking. A separate score conversion table was used based on the online scale scores calculated in Table 8.

The results for science (Table 9 of Appendix 18) indicate that the online version of the test was comparable to the paper version. Differences in raw score conversions were less than half a raw score point throughout the entire score range. Differences in scale score conversions were less than 5 points over most of the score range. The same score conversion table was used for paper and online.

The results for social studies (Table 10 of Appendix 18) indicate that the online version of the test was more difficult than the paper version. Differences in raw score conversions exceeded a half of a raw score point over nearly the entire score range. All of the differences in scale score conversions exceeded two standard errors of the linking. A separate score conversion table was used based on the online scale scores calculated in Table 10.

## 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 was sent to district testing coordinators shortly after the testing period ended containing a link to an online survey. The survey questions and tallies of responses for TAKS grade 8, TAKS grade 9, Algebra I end-of-course, and July TAKS exit level retests are in Appendix 5.

Student responses to online testing were generally positive. Most students agreed that the questions were easy to view and work through on the computer. 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 the 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 described several online testing initiatives undertaken by the Texas Education Agency during the 2005–2006 school year. These included the development and online administration of the Algebra I End-of-Course exam with new performance standards. Additionally, studies of online and paper test comparability were conducted at TAKS grade 8 and grade 9.

The comparability analyses indicated mode-of-administration effects for several tests at TAKS grade 8, grade 9, and exit level. In all cases where a mode effect was found, lower performance on the online versions of the tests justified the use of separate raw to scale score conversions for students testing online. The fairly consistent mode-of-administration effects suggest the need for additional comparability research. This research will include more in-depth analyses to investigate causes of mode differences as TAKS online testing continues into the future.