Results of July 2008 TAKS Exit Level Retest Online Comparability Study Analyses

11/10/08

Summary

The Texas Assessment of Knowledge and Skills (TAKS) has been offered online (in addition to paper) for certain grades and subject areas since 2005. The joint *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999, Standard 4.10) and the *Guidelines for Computer-Based Tests and Interpretations* (APA, 1986) both speak to the need for the evaluation of the score equivalence across computer-based and paper-based tests.

Traditionally, Texas uses a "matched samples comparability analyses", or MSCA (Way, Davis, & Fitzpatrick, 2006), design to study equivalence of the resulting raw-score-to-scale-score conversions between online and paper testing modes. In this design, a bootstrap sampling approach is used to select online and paper samples where each selected online student is matched to a paper student with the same gender, ethnicity, and level of proficiency on previous test scores. Once both the paper sample and online sample are selected, raw-score-to-scale-score conversions are calculated applying the Rasch measurement model. This sampling is repeated for a predetermined number of times. Conversion tables for each replication are retained and aggregated to obtain the final conversion and the standard errors of linking. The equivalency between online and paper scale scores is then evaluated. A separate conversion table is recommended if the two sets of scale scores are considered not comparable.

Beginning with the October 2007 exit level comparability studies, a modified MSCA was implemented due to the large number of first-time testers participating in the retest administrations. The first-time testers differed from the re-testers in a few systematic ways. First of all, the first-time testers had higher mean raw scores than the re-testers, and secondly there was a stronger association between the raw scores and previous TAKS scores for the first-time testers. For the same reasons the modified MSCA approach was adopted in the March/April and July 2008 exit level comparability studies.

The modified MSCA implemented in the July studies utilized student grade level in April 2008 as an additional matching variable to include the first-time testers in constructing the online scale score conversions. In other words, the online first-time testers were only matched to paper first-time testers, where the proficiency was defined by the grade 10 TAKS scores in April 2008. The online re-testers were only matched to paper re-testers, where the proficiency level was defined by the exit level TAKS scores in April 2007. Between-mode item-level analyses and subgroup analyses for ethnic and gender subgroups were performed in addition to the overall test-level analysis.

Using this revised methodology, the comparability analyses were carried out for the July 2008 TAKS exit level administrations in the following subject areas:

- English language arts (ELA);
- mathematics;
- science; and
- social studies.

With the criteria used to evaluate the equivalence (raw score differences that matter, scale score differences compared to two standard errors of linking, and raw score cuts at various performance levels), the comparability between online and paper administrations was assessed. For the July administrations, a mode effect between online and paper was found for ELA with the online version being 1 point lower at the 'Met the Standard' and 1 point higher at the 'Commended' performance levels. The mode of testing was found to have no impact on mathematics, science and social studies. The cut scores were the same between online and paper at both the 'Met the Standard' and 'Commended' levels; no differences were found at any of the raw score points between online and paper for the three subjects.

The subgroup analysis results indicated that a significant mode effect between online and paper was found for the Hispanic subgroup in ELA (with online being harder) and for the male subgroup in science (with online being easier). Moreover, the online version appeared slightly more difficult than the paper version for all subgroups in ELA. Additionally, the item level analysis results indicated that across subject areas, items exhibiting a significant mode effect tended to be slightly easier in the paper version.

The following table summarizes the comparability analysis results for each subject area tested.

Grade	Subject	Scale score conversions differ by more than two standard errors of linking throughout most of the raw score range	Raw scores differences greater than or equal to half of a raw score point throughout most of the raw score range	Cut Score Difference	
				Met the Standard	Commended
Exit Level Retest	ELA	X	X	X	X
	Math				
	Science				
	Social Studies				

Based on the comparability analysis results, Pearson recommended using a separate conversion table for the July 2008 TAKS exit level ELA online retest and the paper pre-equated conversion tables for the exit level online retests in mathematics, science and social studies.

Introduction

Online comparability studies were first implemented in Texas in 2005 covering the Texas Assessment of Knowledge and Skills (TAKS) in grade 8 reading, mathematics and social studies as well as exit level retests of TAKS in English language arts (ELA), mathematics, science and social studies. The 2006 comparability studies included TAKS grade 8 reading, mathematics, science, and social studies, grade 9 reading and mathematics, and all exit level July and October exit level retests. Further expansion for online testing in 2007 added TAKS grade 7 reading and mathematics and grade 10 ELA, mathematics, science, and social studies. In 2008 the TAKS exit level retests were offered both online and on paper in March, April, and July. The exit level retests are planned to be offered both online and on paper again in October. This report summarizes the comparability study results from the July administration.

Methodology

In 2005, TEA and Pearson devised a matched samples comparability analysis (MSCA) plan using a bootstrap approach in which students in the online group would be matched to students from the paper group on their previous TAKS test scores. Based on recommendations from the Texas Technical Advisory Committee (TTAC), additional demographic variables were considered as matching variables beginning in 2006. The decision was made to include previous TAKS scores, ethnicity, and gender as matching variables in the matched samples comparability analyses.

When the October 2007 exit level retest comparability studies were conducted, it was noted that a great number of students would have been excluded from the comparability analyses because of missing April 2007 TAKS exit level scores. Further investigation indicated that these students differed from the re-testers in the following ways: their mean raw score was higher and their raw scores were more highly correlated with their previous TAKS scores (as indicated by higher r-square values) than those of the re-testers in that October administration. Furthermore, these students were in grade 10 in April 2007. Excluding these students may have had an undesirable impact because they systematically differed from the re-testers.

A similar situation occurred in both the March and April 2008 online retest administrations in that first-time testers as well as re-testers participated in the administration (Note that the phrase "first-time testers" used in this document refers to students who participate in the TAKS exit level retests for the first time). Although it is desirable to include the first-time testers in the study, when the first-time testing sample is too small, a problem could occur where not enough students would be assigned to certain matched groups. For this reason, Pearson developed a decision rule for whether or not to include first-time testers in the comparability study. Specifically, first-time testers will be included if *a), the online first-time testers have at least 500 students, or b), the online first-time testers account for at least 20% of all online testers*. If the number of online first-time testers is less than 500, which accounts for less than 20% of all online testers, than these first-time testers will be excluded (from both modes) and only re-testers will be included in the comparability analysis.

In order to appropriately match first-time testers in the comparability analyses when they met the sample size requirement, modifications were made to the comparability method in October 2007 and a stratified MSCA approach has been adopted since then. The stratified MSCA uses a separate y-hat variable for the first-time testers because these students' previous TAKS scores were based on grade 10 (and not exit level) performance. In the bootstrapping process, these first-time testers were matched with first-time testers *only* and not matched to the re-testers. Similarly, the re-testers were matched with re-testers *only*, not with the first-time testers. The detailed procedures are provided in the following section.

In July 2008 the total numbers of online first-time testers were 130 (3.9% of the total online students), 129 (3.9% of the total online students), and 116 (9.5% of the total online students) for mathematics, science, and ELA, respectively. These counts and percentages did not reach the level that would be sufficient for a comparability study, therefore the first-time testers were excluded from the comparability analyses for these three subjects. A stratified MSCA analysis was only conducted for social studies, for which approximately 23% of the online students were the first-time testers.

<u>The Stratified MSCA Method for July 2008 Exit Level Retests</u>

The following is a list of steps in the 'modified' comparability analyses.

1) For the retest students who tested in the paper mode, their July 2008 retest raw scores were regressed on their ELA, mathematics, science, and social studies scale scores from the April 2008 primary administration.
For students with TAKS exit level (XL) scores in April of 2008 (retesters):

$$\hat{Y}_{predicted_rawscore} = \beta_0 + \beta_1 X_{1(prev_XL_ELA)} + \beta_2 X_{2(prev_XL_Math)} + \beta_3 X_{3(prev_XL_Science)} + \beta_4 X_{4(prev_XL_SocialStudies)}$$

For students with TAKS grade 10 scores in April of 2008 (first-time testers):

$$\hat{Y}_{predicted_rawscore} = \beta_0 + \beta_1 X_{1(prev_G10_ELA)} + \beta_2 X_{2(prev_G10_Math)} + \beta_3 X_{3(prev_G10_Science)} + \beta_4 X_{4(prev_G10_SocialStudies)}$$

In other words, for the July 2008 exit level retest comparability analyses, the regression analysis was repeated separately for students with grade 10 TAKS scores in April 2008 and for students with exit level TAKS scores in April 2008.

2) The resulting regression weights were applied to all students (paper and online) to obtain an estimated raw score (y-hat) for each of the students. For the July 2008 exit level retest comparability analyses, online students' y-hat scores were generated using appropriate regression weights depending on whether they took grade 10 TAKS in April 2008 or they took exit level TAKS in April 2008. The regression weights for each group are shown in Table 1.

3) Students in each group (first-time testers and re-testers) were then broken into either 10 or 20 groups based on the estimated raw score. Due to the nature of the matched-sample bootstrap approach, we need to have enough students participating who are first-time testers to use the modified approach. The criteria for including first-time testers and breaking them into groups are shown in the following grid:

Proportion \ Sample Size	>=500	<500
>=20%	20 groups	10 groups
<20%	20 groups	Exclude

Therefore, the first-time testers would be excluded from the analyses if their count was less than 500 **AND** was less than 20% of the total online students (first-time testers and re-testers combined). In the July administration, both first-time testers and re-testers were included in the analyses for social studies only. There were only 123 first-time testers who took the social studies online but accounted for more than 20% of all online testers. One hundred and nineteen first-time testers were eligible for inclusion in the comparability analysis.

In order to modify the procedure such that re-testers would only be matched to re-testers and first-time testers would only be matched to first-time testers, the following steps were implemented:

I. Grouped the retest students (online plus paper) with previous exit level information into 10 (or 20) score groups based on their y-hat values. This produced a ranking variable ranging from 0 to 9 (or 19), for all re-testers (students with exit level information).

II. Grouped the first-time students (online plus paper) with previous grade 10 information into 10 or 20 score groups based on their y-hat values. This also produced a ranking variable ranging from 0 to 9 (or 19), for all first-time testers (students with grade 10 information).

III. Added a constant to the exit level ranking from step I, such that the ranking wasn't duplicated between the two groups of students. For example, if 20 groups were used for both first-time testers and re-testers, then the rank variable would range from 0 to 39, where 0-19 was for the first-time testers, and 20-39 was for the re-testers.

IV. Separated the online group of students from the paper group of students to form the base of online and paper samples, and then proceeded with the bootstrap process.

Please note that if the students all had exit level TAKS score in April 2008, then there would have only had 20 groups for each subject area, which would have become the regular MSCA approach.

4) This grouping from step 3 could result in a 40 (previous score groups—20 for first-time testers plus 20 for re-testers) by 4 (ethnicity groups) by 2 (gender groups) grid that was used in the matched sampling. Note that the total number of groups could be different from 40, depending on the number of groups for the first-time testers as well as the re-testers.

5) To improve optimal matching, students with missing values on any of the matching variables were dropped from the study.

Implementation of the Analysis Procedures

The regular/standard matched samples comparability analyses were conducted for ELA, mathematics, and science with first-time testers being excluded. The modified matched samples comparability analysis method was used for social studies. The key steps are described below. Please note that with the modified approach, the number of score groups has become 30 for social studies instead of 20 as in regular/standard comparability analyses for ELA, mathematics, and science.

1. All students eligible for matching were placed into score groups based on the regression of July 2008 retest raw score on the April 2008 scale scores as described earlier. Each student testing online with complete data was matched to a student from the available 2008 July paper TAKS data with an identical profile on the matching variables.
2. Online versus paper comparability analyses were performed using matched groups of students. The following steps were repeated 100 times (500 for ELA):
    a. A bootstrap sample of students was drawn from the online participants.
    b. A matched sample was drawn at random from the available July 2008 paper TAKS data.
    c. A raw score-to-raw score equating was carried out with the bootstrap samples.
    d. The raw score equivalents were transformed to scale scores using the appropriate paper pre-equated score conversion tables and linear interpolation.
3. Online scale score conversions for each raw score were based on the average of the conversions calculated over each of the 100 replications (500 for ELA). These average scale score values comprised the alternate online conversion table.
4. The standard deviation of online scale score conversions at each raw score represented the conditional bootstrap standard errors of the linking.
5. To evaluate comparability, raw score points for which the difference between the online and paper scale score conversions exceeded two standard errors of the linking were noted. Also, the differences between the paper-raw-score equivalent and the corresponding paper raw score were calculated, and differences greater than 0.5 were noted.

To verify the comparability analysis results, two Pearson psychometricians programmed independent versions of the analysis. The online results were then compared to the paper results. The recommendations regarding the use of a separate online score conversion table were made based on a set of rules that TEA adopted in 2006. Pearson recommended the consideration of three pieces of information: the standard error of the linking, the magnitude of the raw score differences, and the rounding differences for cut scores.

- The standard error of the linking criterion was suggested by Dorans and Lawrence (1990): "To assess equivalence, it is convenient to compute the difference between the equating function and the identity transformation, and to divide this difference by the standard error of equating. If the resultant ratio falls within a bandwidth of plus or minus two, then the equating function is deemed to be within sampling error of the identity function" (p. 247). In using this procedure, special attention was paid to differences in the range of scale scores around the "Met the Standard" and "Commended" performance levels. Differences at the two extremes of the scale are less important, given the purpose

and primary uses of the TAKS tests. This standard error procedure is sensitive to the sample size such that the standard errors will be greater when the sample sizes are smaller. Therefore, we also considered the following additional criteria.

- The magnitude of the raw score differences was evaluated using the criterion of differences that matter (DTM; Dorans & Feigenbaum, 1994). This was originally developed in the context of the SAT where scale scores are reported in 10-point units. For a given raw score, if the resulting scale scores from the linking differed by fewer than 5 points, then the scale scores would likely be rounded to the same value and would be considered equivalent. This process was adapted to other tests and the DTM was considered to be a half of a score unit for unrounded scores (Dorans, Holland, Thayer, & Tateneni, 2003). For the TAKS, the DTM was considered to be half of a raw score point. For a given proficiency level, if the corresponding raw scores from the linking differed by less than half of a raw score point, then the two could be considered equivalent.

- The third piece of information considered was the rounding differences for the cut scores. The raw-score-to -scale-score conversions for the paper and online tests were compared to see if they resulted in different raw score cut points across the two modes of test administration. Cuts were evaluated for both the "Met the Standard" and the "Commended Performance" levels. Pearson recommended that this information be used in conjunction with the magnitude of the raw score differences, and the statistical significance of the differences (based on the Dorans and Lawrence [1990] two standard errors of the linking).

In addition to those three pieces of information, subgroup and item-level analyses were conducted, and the impact of using the alternate score table was considered. Results from these analyses were used to inform decision-making in borderline cases. Using these pieces of information, overall psychometric judgment determined the recommendation for whether to use an alternate score table for the online TAKS administration.

Results

The summary of samples and regression results are presented in Table 1. As can be seen, the mean scores were higher and the relationship between the raw scores and previous TAKS scores was much stronger for students who took the social studies retest for the first time than those who were re-testers. Table 2 displays the demographic information and descriptive statistics for the online and paper samples for each subject area tested. Across all subject areas, the Hispanic students accounted for the majority of all the students tested, and the online and paper samples tended to be similar in terms of gender and ethnicity proportions. As observed consistently in the previous TAKS exit level administrations, a higher proportion of Hispanic students took the test on paper whereas a higher proportion of white students took the tests online. Unlikely the March and April administrations, however, a higher proportion of African American students took the tests online in July.

It can also be seen from Table 2 that for ELA, mathematics and science, online students generally had higher raw scores than the paper students; whereas for social studies, online students generally had lower raw scores than the paper students---although for all subject areas the difference appeared to be very small.

Studies from previous comparability analyses on ELA retest administrations have consistently indicated an item type by mode interaction, namely students tested online tend to have higher essay scores, whereas students tested in paper tend to have higher scores on multiple-choice items. Table 3 provides the performance for July exit level ELA students by testing mode, which was consistent with the previous findings.

Table 4 provides an overview of the results of the July online and paper comparability studies. A test-level mode effect was detected only for ELA. The 'Met the Standard' cut was lower for online (meaning that the online version was harder) while the 'Commended' cut was higher for online (meaning that the online version was easier).

Tables 5 to 8 detail the comparisons between online and paper for each July retest. The columns of the tables are as follows:

RAW – Paper test raw score
CBT_RS – Equivalent raw scores on the online test based on the comparability linking. Note that a higher equivalent raw score indicates that the online version of the test was more difficult.

RS_SD – Standard deviation of the equivalent raw scores over the 100/500 replications.
PAP_SS – Paper test scale score conversions, based on the 2008 July pre-equated scales.
CBT_SS – Equivalent scale scores on the online test based on the comparability linking. Again, higher equivalent scale scores indicate that the online version of the test was more difficult.
SS_SD – Standard deviation of the equivalent scale scores over the 100/500 replications.
RS_DIF – Difference between paper-raw-score equivalent and paper raw score.
SS_DIF – Difference between online scale score equivalent and paper scale score.
SIG – Raw score points where scale score differences exceed two standard errors of the linking and where the difference in raw scores is greater than half a point are noted by "*".
FINAL– Final recommended online scale score conversion.

Note that the "*" in the SIG column indicates both statistical and practical significance, based on the recommendations from the TTAC. In addition, the scale scores that corresponded to the '1-SEM and '2-SEM performance cuts are underlined and italicized; whereas the scale scores that corresponded to the 'Met Standard' and 'Commended' performance cuts are highlighted and bolded.

For ELA, as shown in Table 5, student performance was impacted by the test mode. The differences in raw scores between paper and online exceeded half of a raw score point throughout much of the raw score range and the scale score differences were statistically significant throughout the scale. The online "Met Standard" cut score was 1 point lower than the paper version, while the online "Commended" cut was 1 point higher.

For mathematics, science, and social studies (see Tables 6-8), there was no score difference shown as both practically and statistically significant. Both the "Met the Standard" and "Commended" cut scores were the same for the paper and online version of the tests. Based on these criteria, it was concluded that the testing mode did not significantly impact student performance for these subject areas.

It has been consistently observed in the previous comparability studies that large differences occur at the lowest and highest scaled score points for all subject areas. A similar pattern has shown in the July administration, though less obvious for mathematics, science, and social studies than for ELA. This is because WINSTEPS (the IRT calibration software used in the study) does not estimate abilities for zero and perfect scores. These differences are not meaningful. Therefore, for each test the online conversion table has the scale score at the two extreme ends set to the paper as has been consistently done in the past.

Impact Data Analyses

The pass-rate comparison shown in Table 4 indicates that applying the paper conversion tables to the online students taking the ELA exit level retest resulted in a lower percentage of students achieving the 'Met the Standard' level and a higher percentage of students achieving the 'Commended' level. Using the alternate (online) conversion tables resulted in passing rates for the online students that were more similar to passing rates for paper students.

Additional Analysis

As consistent with previous online comparability studies, two sets of additional analyses were conducted: the subgroup analysis and the item-level analysis. The subgroup analysis compares mean differences of the total raw scores between the two testing modes across replications for male, female, White, Hispanic and African American student groups separately, whereas the item-level analysis compares mean differences of each item between the two testing modes across replications.

Subgroup Analysis

The mean raw score differences (and the mean effect size, see Cohen [1992]) between the online and paper testing modes for each subgroup for each subject area tested in July are listed in Table 9. A significance test was performed for each 'matched' subgroup using the following equation:

$$Zdif = \frac{\overline{D}_{Diff}}{\sqrt{SE^2_{Diff}}}$$

where $\overline{D}_{Diff}$ is the grand mean of the differences between mean online and mean paper essay scores over the replications for each subgroup; and SE_{diff} is the bootstrap standard error of the mean differences over the replications, also for each subgroup.

The effect size between two group means <u>at each replication</u> was calculated by the following equation:

$$EffectSize = \frac{\overline{X}_{Group1} - \overline{X}_{Group2}}{\sqrt{\frac{(SD^2_{Group1} + SD^2_{Group2})}{2}}}$$

The effect sizes for the raw scores were based on the averages of the effect sizes over the replications.

As can be seen in Table 9, online was significantly harder for Hispanic students taking the ELA retest and significantly easier for male students taking the science retest. The effect sizes, however, were small for both subjects. There was no significant mode effect presented for any subgroups for mathematics and social studies.

Item-Level Analysis

Item-level analysis was performed in a similar way to the subgroup analysis. The mean item raw score differences across replications between the online and paper testing modes for each subject area were computed, and the effect sizes were also calculated. Tables 10 to 13 display the results of item-level comparison across replications for each subject area tested in July. The columns of the tables are as follows:

CBT_PVAL: Mean item score for the online students across 100/500 replications.
PAP_PVAL: Mean item score for the paper students across 100/500 replications.
DIF_PVAL: Mean item score differences between online and paper students across 100/500 replications
DIF_STD: Standard deviation of the mean differences across 100/500 replications
Z_DIF: Z statistic for the mean item score differences
SIG: Items where the Z_DIF statistic was greater than 2 are noted by "*".
EFFECT_SIZE: Mean effect size over 100/500 replications.

As can be seen from the tables, only 5 items exhibit a mode effect in ELA, with two of them in favor of paper. For social studies, there were 7 items exhibiting a mode effect, with 5 of them in favor of paper. For science, the online group scored higher on 7 of the 11 items. A relatively larger number of items showed a mode effect in mathematics. The paper group scored higher on 12 items of the 23 items exhibiting a mode effect.

The average effect size across iterations for each item difficulty difference was calculated and listed on the last column of these tables. It can be seen that the magnitude of the effect sizes ranged from 0.09 to 0.15 for ELA, from 0.06 to 0.15 for mathematics, from 0.05 to 0.11 for science, and from 0.11 to 0.15 for social studies.

Summary and Recommendation

For the four subjects analyzed in the July 2008 administration, a test-level mode effect was only found in ELA. Based on the analysis results, Pearson recommended using a separate online conversion table for students taking the online ELA exit level retest and the paper conversion tables for students taking the online exit level retests in mathematics, science, and social studies.

References

American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessments (APA). (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.

American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). (1999). S*tandards for educational and psychological testing*. Washington, DC: AERA.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1)*,* 155-159.

Dorans, N. J., & Feigenbaum, M. D. (1994). *Equating issues engendered by changes to the SAT and PSAT/NMSQT*. (ETS Research Memorandum 94-10). Princeton, NJ: Educational Testing Service.

Dorans, N. J., Holland, P. W., Thayer, D. T., & Tateneni, K. (2003). Invariance of score linking across gender groups for three Advanced Placement Program Examinations. In N. J. Dorans, (Ed.), *Population invariance of score linking: Theory and applications to Advanced Placement Program Examinations.* (ETS RR-03-27, pp. 79-118). Princeton, NJ: Educational Testing Service.

Dorans, N. J., & Lawrence, I. M. (1990). Checking the statistical equivalence of nearly identical test forms. *Applied Measurement in Education*, *3*, 245-254.

Way, W. D., Davis, L. L., & Fitzpatrick, S. (2006). *Score comparability of online and paper administrations of the Texas Assessment of Knowledge and Skills*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Table 1. Summary of Sample Statistics and Regression Results, by Student Population and Subject Area Tested
July 2008

Subject Tested	Student Population	Intercept	Prev_ELA	Prev_Math	Prev_Scie	Prev_Sost	r-square	PaperN	PaperMean	OnlineN	OnlineMean
ELA	First timers	--	--	--	--	--	--	--	--	--	--
	Re-testers	-190.453	0.071	0.007	0.012	0.022	0.454	7544	42.70	1097	42.42
Math	First timers	--	--	--	--	--	--	--	--	--	--
	Re-testers	-120.677	0.011	0.047	0.010	0.004	0.299	21562	27.98	3193	28.34
Science	First timers	--	--	--	--	--	--	--	--	--	--
	Re-testers	-102.715	0.012	0.010	0.024	0.015	0.339	20467	26.51	3216	27.02
Social Studies	First timers	-77.043	0.012	-0.008	0.021	0.026	0.590	1027	36.22	119	36.94
	Re-testers	-82.184	0.019	0.006	0.011	0.016	0.238	3194	23.16	418	23.28

Table 2. Student Demographic and Descriptive Information for July 2008 Comparability Analyses*

	Number of Campuses		Number of Students		Mean Raw Score		Estimated Raw Score, or y-hat		Demographic Information**											
									Male		White		Hispanic		African American		Other Ethnicity		Special Education***	
Subject	CBT	*PAP*	CBT	*PAP*	CBT	*PAP*	CBT	*PAP*	CBT	*PAP*	CBT	*PAP*	CBT	*PAP*	CBT	*PAP*	CBT	*PAP*	CBT	*PAP*
ELA	118	*1047*	1097	*7544*	42.70	*42.42*	43.38	*42.42*	55	*53*	18	*15*	57	*64*	20	*17*	5	*3*	6	*5*
Mathematics	154	*1302*	3193	*21562*	28.34	*27.98*	28.26	*27.98*	41	*42*	24	*22*	45	*54*	29	*23*	2	*1*	5	*4*
Science	151	*1306*	3216	*20467*	27.02	*26.51*	26.99	*26.51*	31	*32*	19	*17*	51	*59*	27	*23*	3	*2*	3	*3*
Social Studies	117	*864*	537	*4221*	26.30	*26.34*	26.11	*26.34*	37	*37*	16	*14*	57	*65*	23	*17*	4	*4*	4	*4*

*: CBT-Online administration; PAP-Paper administration.
**: Cell entries represent percentages rounded to the nearest integer. Due to rounding, the percentages may not add up to 100%.
***: Special education status is included in the demographics table, but was not used as a matching variable.

Table 3. Student Performance* on July 2008 Exit Level ELA Multiple Choice, Open-Ended Items, and Essay by Test Mode

	ONLINE					PAPER				
Multiple Choice	32.75					32.35				
	0	1	2	3		0	1	2	3	
OE 1	11.67%	56.43%	31.81%	0.09%		13.55%	57.67%	28.71%	0.07%	
OE 2	21.15%	54.15%	24.70%	0.00%		23.36%	56.16%	20.43%	0.05%	
OE 3	30.90%	51.60%	17.50%	0.00%		37.46%	48.37%	14.12%	0.05%	
	0	1	2	3	4	0	1	2	3	4
Essay	13.22%	23.43%	43.76%	18.14%	1.46%	5.77%	24.36%	55.33%	13.71%	0.84%

*: based on the eligible students included in the analyses.

Table 4. Summary of the July 2008 Exit Level Online TAKS Comparability Analyses

Grade /Subject	Sample Size		Raw Score Cuts*		Pass Rates (%) Comparison**			Number of Raw Score Points with Meaningful Difference***	Decision (Conclusion)
	Online	Paper	Online	Paper	Paper w/Paper	Online w/Online	Online w/Paper		
Exit Level ELA****	1097	*7544*	**40**	41	55.8	54.9	53.3	42 (73)	Alternate Conversion Table (Mode Effect)
			64	63	1.6	1.3	2.2		
Exit Level Math	3193	*21562*	33	**33**	29.8	28.1	29.8	0 (60)	Paper Conversion Table (No Mode Effect)
			53	**53**	0.0	0.0	0.0		
Exit Level Science	3216	*20467*	30	**30**	35.1	31.8	35.1	0 (55)	Paper Conversion Table (No Mode Effect)
			50	**50**	0.0	0.0	0.0		
Exit Level Social Studies	537	*4221*	28	**28**	38.9	36.8	38.9	0 (55)	Paper Conversion Table (No Mode Effect)
			49	**49**	1.3	2.1	1.3		

*: Raw score points corresponding to 'Met Standards' (top) and 'Commended' (bottom) levels. Final RS cuts in bold.
**: Pass rates based on different conversion tables.
***: Meaningful differences require both scale score statistical significance and raw score practical significance. Total maximum RS points shown in parentheses.
****: Note that for ELA, an essay score of 2 or above is required in addition to the scale score requirements to be considered passing.

Table 5. Summary of Comparability Analysis – July Exit Level ELA

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG	Final
0	0.36	0.019	1427.83	1469.04	2.103	0.36	41.21		1428
1	1.21	0.060	1540.80	1558.04	4.864	0.21	17.24		1558
2	2.41	0.113	1622.13	1642.23	5.525	0.41	20.10		1642
3	3.60	0.162	1670.81	1692.10	5.696	0.60	21.29	*	1692
4	4.78	0.205	1706.14	1727.96	5.584	0.78	21.82	*	1728
5	5.95	0.244	1734.16	1756.11	5.354	0.95	21.96	*	1756
6	7.11	0.278	1757.54	1779.50	5.182	1.11	21.96	*	1779
7	8.25	0.309	1777.75	1799.61	5.070	1.25	21.86	*	1800
8	9.39	0.337	1795.61	1817.37	4.989	1.39	21.76	*	1817
9	10.51	0.361	1811.70	1833.32	4.895	1.51	21.62	*	1833
10	11.62	0.383	1826.41	1847.84	4.807	1.62	21.43	*	1848
11	12.73	0.402	1839.97	1861.21	4.714	1.73	21.24	*	1861
12	13.82	0.419	1852.62	1873.63	4.625	1.82	21.00	*	1874
13	14.90	0.434	1864.49	1885.24	4.536	1.90	20.75	*	1885
14	15.97	0.447	1875.71	1896.17	4.454	1.97	20.46	*	1896
15	17.03	0.458	1886.37	1906.52	4.374	2.03	20.15	*	1907
16	18.08	0.467	1896.54	1916.37	4.296	2.08	19.83	*	1916
17	19.12	0.475	1906.30	1925.78	4.223	2.12	19.48	*	1926
18	20.15	0.482	1915.68	1934.80	4.155	2.15	19.12	*	1935
19	21.17	0.487	1924.75	1943.50	4.086	2.17	18.75	*	1943
20	22.19	0.491	1933.53	1951.89	4.020	2.19	18.36	*	1952
21	23.19	0.494	1942.08	1960.02	3.960	2.19	17.94	*	1960
22	24.19	0.496	1950.40	1967.93	3.901	2.19	17.53	*	1968
23	25.17	0.497	1958.52	1975.63	3.844	2.17	17.11	*	1976
24	26.15	0.497	1966.49	1983.16	3.790	2.15	16.67	*	1983
25	27.13	0.496	1974.31	1990.54	3.735	2.13	16.23	*	1991
26	28.09	0.495	1982.00	1997.77	3.687	2.09	15.77	*	1998
27	29.05	0.492	1989.59	2004.91	3.643	2.05	15.31	*	2005
28	30.01	0.489	1997.08	2011.95	3.599	2.01	14.87	*	2012
29	30.96	0.485	2004.52	2018.92	3.555	1.96	14.40	*	2019
30	31.90	0.481	2011.89	2025.82	3.514	1.90	13.93	*	2026
31	32.84	0.476	2019.24	2032.69	3.479	1.84	13.45	*	2033
32	33.78	0.470	2026.54	*2039.53*	3.444	1.78	12.99	*	*2045*
33	34.71	0.464	2033.85	2046.36	3.410	1.71	12.51	*	2046
34	35.63	0.457	*2041.17*	2053.19	3.376	1.63	12.02	*	2053
35	36.56	0.450	2048.51	2060.04	3.348	1.56	11.53	*	2060
36	37.48	0.442	2055.89	*2066.93*	3.324	1.48	11.04	*	*2072*
37	38.40	0.434	2063.32	2073.87	3.299	1.40	10.55	*	2074
38	39.32	0.426	*2070.84*	2080.87	3.272	1.32	10.04	*	2081
39	40.23	0.417	2078.44	2087.95	3.255	1.23	9.51	*	2088
40	41.14	0.408	2086.13	**2095.14**	3.236	1.14	9.01	*	**2100**
41	42.06	0.398	**2093.97**	2102.44	3.217	1.06	8.46	*	2102
42	42.97	0.389	2101.95	2109.87	3.203	0.97	7.92	*	2110
43	43.88	0.379	2110.10	2117.46	3.190	0.88	7.36	*	2117
44	44.79	0.369	2118.44	2125.21	3.179	0.79	6.77	*	2125
45	45.70	0.358	2127.00	2133.16	3.170	0.70	6.17		2133
46	46.61	0.348	2135.79	2141.33	3.168	0.61	5.54		2141
47	47.52	0.338	2144.86	2149.75	3.173	0.52	4.89		2150
48	48.43	0.327	2154.24	2158.45	3.180	0.43	4.20		2158
49	49.34	0.317	2163.97	2167.45	3.194	0.34	3.48		2167
50	50.26	0.307	2174.08	2176.79	3.215	0.26	2.71		2177
51	51.17	0.298	2184.62	2186.52	3.242	0.17	1.89		2187
52	52.08	0.288	2195.65	2196.68	3.279	0.08	1.02		2197
53	53.00	0.279	2207.24	2207.33	3.324	0.00	0.09		2207
54	53.92	0.271	2219.44	2218.54	3.384	-0.08	-0.90		2219
55	54.84	0.264	2232.36	2230.40	3.465	-0.16	-1.97		2230
56	55.77	0.257	2246.10	2242.99	3.570	-0.23	-3.11		2243
57	56.70	0.251	2260.77	2256.43	3.716	-0.30	-4.34		2256

Table 5. Summary of Comparability Analysis – July Exit Level ELA (Continued)

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG	Final
58	57.64	0.247	2276.53	2270.89	3.912	-0.36	-5.64		2271
59	58.59	0.245	2293.55	2286.53	4.182	-0.41	-7.03		2287
60	59.54	0.246	2312.10	2303.60	4.573	-0.46	-8.49		2304
61	60.51	0.253	2332.42	2322.43	5.140	-0.49	-10.00		2322
62	61.49	0.268	2354.88	2343.38	6.012	-0.51	-11.50		2343
63	62.48	0.295	**2379.93**	2367.00	7.399	-0.52	-12.93		2367
64	63.49	0.340	2408.14	**2393.95**	9.620	-0.51	-14.18		**2400**
65	64.52	0.401	2440.23	2425.10	13.058	-0.48	-15.13		2425
66	65.56	0.472	2477.24	2461.57	17.977	-0.44	-15.67		2462
67	66.62	0.529	2520.72	2505.00	24.154	-0.38	-15.72		2505
68	67.69	0.542	2573.41	2558.73	30.671	-0.31	-14.69		2559
69	68.82	0.494	2640.21	2630.72	36.700	-0.18	-9.49		2631
70	70.03	0.458	2728.19	2736.26	47.455	0.03	8.07		2736
71	71.17	0.475	2842.09	2872.47	66.901	0.17	30.38		2872
72	72.12	0.341	2995.81	3015.58	53.520	0.12	19.76		3016
73	72.74	0.129	3154.79	3113.11	20.463	-0.26	-41.68		3155

Table 6. Summary of Comparability Analysis – July Exit Level Mathematics

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG	Final
0	0.30	0.005	1311.55	1353.99	0.674	0.30	42.44		1312
1	1.00	0.015	1453.23	1452.75	1.826	-0.00	-0.48		1453
2	2.00	0.028	1554.75	1553.88	2.313	-0.00	-0.87		1555
3	2.99	0.040	1615.85	1615.12	2.135	-0.01	-0.73		1616
4	3.99	0.051	1660.41	1659.71	2.051	-0.01	-0.70		1660
5	4.98	0.060	1695.92	1695.21	1.995	-0.02	-0.71		1696
6	5.98	0.069	1725.73	1724.99	1.952	-0.02	-0.74		1726
7	6.97	0.078	1751.60	1750.84	1.915	-0.03	-0.76		1752
8	7.97	0.085	1774.60	1773.81	1.887	-0.03	-0.79		1775
9	8.96	0.092	1795.42	1794.60	1.862	-0.04	-0.82		1795
10	9.96	0.099	1814.53	1813.68	1.840	-0.04	-0.85		1815
11	10.95	0.105	1832.27	1831.39	1.823	-0.05	-0.88		1832
12	11.95	0.111	1848.88	1847.98	1.807	-0.05	-0.90		1849
13	12.94	0.116	1864.62	1863.70	1.794	-0.06	-0.92		1865
14	13.94	0.121	1879.48	1878.55	1.774	-0.06	-0.93		1879
15	14.93	0.126	1893.68	1892.74	1.762	-0.07	-0.94		1894
16	15.93	0.130	1907.32	1906.37	1.753	-0.07	-0.95		1907
17	16.93	0.134	1920.48	1919.52	1.743	-0.07	-0.96		1920
18	17.93	0.137	1933.21	1932.25	1.733	-0.07	-0.96		1933
19	18.92	0.140	1945.58	1944.61	1.724	-0.08	-0.97		1946
20	19.92	0.143	1957.63	1956.66	1.716	-0.08	-0.97		1958
21	20.92	0.146	1969.41	1968.45	1.707	-0.08	-0.96		1969
22	21.92	0.148	1980.94	1979.98	1.697	-0.08	-0.96		1981
23	22.92	0.150	1992.23	1991.29	1.685	-0.08	-0.94		1992
24	23.92	0.151	2003.37	2002.43	1.681	-0.08	-0.94		2003
25	24.92	0.153	*2014.37*	*2013.45*	1.674	-0.08	-0.92		*2015*
26	25.92	0.154	2025.25	2024.34	1.666	-0.08	-0.91		2025
27	26.92	0.154	2036.02	2035.13	1.659	-0.08	-0.89		2036
28	27.92	0.155	2046.73	2045.86	1.654	-0.08	-0.87		2047
29	28.92	0.155	*2057.38*	*2056.53*	1.646	-0.08	-0.85		*2058*
30	29.92	0.154	2068.00	2067.18	1.639	-0.08	-0.82		2068
31	30.92	0.154	2078.62	2077.82	1.632	-0.08	-0.80		2079
32	31.93	0.153	2089.24	2088.47	1.626	-0.07	-0.77		2089
33	32.93	0.152	**2099.90**	**2099.16**	1.619	-0.07	-0.74		**2100**
34	33.93	0.150	2110.61	2109.90	1.612	-0.07	-0.71		2111
35	34.94	0.148	2121.40	2120.73	1.607	-0.06	-0.67		2121
36	35.94	0.146	2132.30	2131.66	1.602	-0.06	-0.64		2132
37	36.94	0.144	2143.33	2142.73	1.595	-0.06	-0.60		2143
38	37.95	0.141	2154.51	2153.95	1.589	-0.05	-0.56		2155
39	38.95	0.138	2165.88	2165.35	1.583	-0.05	-0.53		2166
40	39.96	0.135	2177.47	2176.99	1.578	-0.04	-0.48		2177
41	40.96	0.131	2189.32	2188.88	1.572	-0.04	-0.44		2189
42	41.97	0.128	2201.47	2201.07	1.567	-0.03	-0.40		2201
43	42.97	0.124	2213.97	2213.62	1.563	-0.03	-0.35		2214
44	43.97	0.119	2226.88	2226.57	1.558	-0.03	-0.31		2227
45	44.98	0.114	2240.26	2240.00	1.554	-0.02	-0.26		2240
46	45.98	0.109	2254.20	2253.99	1.550	-0.02	-0.21		2254
47	46.99	0.104	2268.79	2268.63	1.545	-0.01	-0.16		2269
48	47.99	0.098	2284.14	2284.03	1.542	-0.01	-0.11		2284
49	48.99	0.092	2300.41	2300.36	1.540	-0.01	-0.05		2300
50	50.00	0.085	2317.79	2317.79	1.538	-0.00	0.00		2318
51	51.00	0.079	2336.53	2336.59	1.538	0.00	0.06		2337
52	52.00	0.072	2356.95	2357.07	1.537	0.00	0.12		2357
53	53.00	0.064	**2379.54**	**2379.72**	1.539	0.00	0.18		**2400**
54	54.01	0.056	2404.97	2405.22	1.544	0.01	0.25		2405
55	55.01	0.048	2434.32	2434.64	1.554	0.01	0.32		2434
56	56.01	0.039	2469.35	2469.76	1.573	0.01	0.41		2469
57	57.01	0.030	2513.42	2513.95	1.615	0.01	0.53		2513

Table 6. Summary of Comparability Analysis – July Exit Level Mathematics (Continued)

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG	Final
58	58.01	0.021	2573.99	2574.74	1.746	0.01	0.75		2574
59	59.00	0.011	2675.16	2675.70	1.331	0.00	0.54		2675
60	59.70	0.003	2817.41	2774.88	0.459	-0.30	-42.53		2817

Table 7. Summary of Comparability Analysis – July Exit Level Science

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG	Final
0	0.30	0.004	1379.42	1418.54	0.499	0.30	39.12		1379
1	1.00	0.012	1509.99	1509.74	1.353	-0.00	-0.25		1510
2	2.00	0.023	1603.81	1603.52	1.683	0.00	-0.29		1604
3	3.00	0.033	1660.43	1660.40	1.566	0.00	-0.03		1660
4	4.01	0.042	1701.84	1701.96	1.516	0.01	0.12		1702
5	5.01	0.050	1734.93	1735.15	1.483	0.01	0.22		1735
6	6.01	0.057	1762.79	1763.09	1.458	0.01	0.30		1763
7	7.02	0.064	1787.05	1787.41	1.438	0.02	0.36		1787
8	8.02	0.070	1808.68	1809.08	1.421	0.02	0.40		1809
9	9.03	0.076	1828.33	1828.76	1.405	0.03	0.43		1828
10	10.03	0.081	1846.42	1846.86	1.391	0.03	0.44		1846
11	11.03	0.085	1863.27	1863.73	1.383	0.03	0.46		1863
12	12.03	0.090	1879.16	1879.61	1.368	0.03	0.45		1879
13	13.03	0.094	1894.11	1894.56	1.355	0.03	0.45		1894
14	14.03	0.097	1908.37	1908.81	1.347	0.03	0.44		1908
15	15.03	0.100	1922.05	1922.47	1.338	0.03	0.42		1922
16	16.03	0.103	1935.23	1935.64	1.330	0.03	0.41		1935
17	17.03	0.106	1947.98	1948.36	1.322	0.03	0.38		1948
18	18.03	0.108	1960.36	1960.72	1.314	0.03	0.36		1960
19	19.03	0.110	1972.43	1972.76	1.308	0.03	0.33		1972
20	20.03	0.112	1984.23	1984.53	1.297	0.03	0.30		1984
21	21.02	0.113	1995.74	1996.01	1.293	0.02	0.27		1996
22	22.02	0.114	2007.11	2007.35	1.288	0.02	0.24		2007
23	23.02	0.115	2018.31	2018.52	1.283	0.02	0.21		2018
24	24.02	0.116	*2029.38*	*2029.55*	1.278	0.02	0.17		*2035*
25	25.01	0.117	2040.35	2040.49	1.273	0.01	0.14		2040
26	26.01	0.117	2051.24	2051.35	1.268	0.01	0.11		2051
27	27.01	0.117	*2062.08*	*2062.15*	1.265	0.01	0.07		*2068*
28	28.00	0.117	2072.90	2072.94	1.261	0.00	0.04		2073
29	29.00	0.116	2083.72	2083.73	1.257	0.00	0.01		2084
30	30.00	0.115	**2094.57**	**2094.55**	1.255	-0.00	-0.02		**2100**
31	30.99	0.114	2105.47	2105.42	1.252	-0.01	-0.05		2105
32	31.99	0.113	2116.44	2116.36	1.248	-0.01	-0.08		2116
33	32.99	0.112	2127.53	2127.42	1.246	-0.01	-0.11		2128
34	33.99	0.110	2138.76	2138.62	1.244	-0.01	-0.14		2139
35	34.98	0.108	2150.15	2149.99	1.243	-0.02	-0.16		2150
36	35.98	0.106	2161.76	2161.57	1.241	-0.02	-0.19		2162
37	36.98	0.103	2173.61	2173.40	1.239	-0.02	-0.21		2174
38	37.98	0.101	2185.76	2185.52	1.239	-0.02	-0.24		2186
39	38.98	0.098	2198.26	2198.00	1.238	-0.02	-0.26		2198
40	39.98	0.094	2211.17	2210.89	1.239	-0.02	-0.28		2211
41	40.98	0.091	2224.58	2224.28	1.240	-0.02	-0.30		2225
42	41.98	0.087	2238.55	2238.24	1.241	-0.02	-0.31		2239
43	42.98	0.083	2253.22	2252.89	1.242	-0.02	-0.33		2253
44	43.98	0.078	2268.70	2268.35	1.244	-0.02	-0.35		2269
45	44.98	0.073	2285.19	2284.83	1.245	-0.02	-0.36		2285
46	45.98	0.068	2302.90	2302.53	1.249	-0.02	-0.37		2303
47	46.98	0.062	2322.16	2321.78	1.253	-0.02	-0.38		2322
48	47.98	0.056	2343.38	2343.00	1.258	-0.02	-0.38		2343
49	48.98	0.050	2367.21	2366.83	1.264	-0.02	-0.38		2367
50	49.98	0.043	**2394.62**	**2394.24**	1.272	-0.02	-0.38		**2400**
51	50.99	0.036	2427.26	2426.90	1.286	-0.01	-0.36		2427
52	51.99	0.028	2468.33	2468.00	1.307	-0.01	-0.33		2468
53	52.99	0.019	2524.49	2524.25	1.368	-0.01	-0.24		2524
54	54.00	0.010	2617.85	2617.58	1.078	-0.00	-0.27		2618
55	54.70	0.003	2748.99	2709.49	0.401	-0.30	-39.50		2749

Table 8. Summary of Comparability Analysis – March Exit Level Social Studies

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG	Final
0	0.30	0.011	1415.25	1453.53	1.402	0.30	38.28		1415
1	0.99	0.036	1543.59	1542.48	3.973	-0.01	-1.11		1544
2	1.99	0.069	1635.95	1634.20	5.202	-0.01	-1.75		1636
3	2.99	0.099	1691.83	1690.57	4.869	-0.01	-1.26		1692
4	3.98	0.128	1732.75	1731.70	4.754	-0.02	-1.05		1733
5	4.98	0.155	1765.46	1764.54	4.693	-0.02	-0.92		1765
6	5.98	0.180	1792.98	1792.15	4.658	-0.02	-0.83		1793
7	6.98	0.203	1816.99	1816.22	4.625	-0.02	-0.77		1817
8	7.97	0.225	1838.31	1837.60	4.592	-0.03	-0.71		1838
9	8.97	0.245	1857.67	1857.00	4.571	-0.03	-0.67		1858
10	9.97	0.264	1875.49	1874.85	4.553	-0.03	-0.64		1875
11	10.97	0.281	1892.09	1891.48	4.541	-0.03	-0.61		1892
12	11.97	0.297	1907.70	1907.11	4.526	-0.03	-0.59		1908
13	12.97	0.312	1922.48	1921.90	4.500	-0.03	-0.58		1922
14	13.97	0.326	1936.49	1935.95	4.481	-0.03	-0.54		1936
15	14.96	0.338	1949.98	1949.44	4.477	-0.04	-0.54		1950
16	15.96	0.349	1962.97	1962.44	4.461	-0.04	-0.53		1963
17	16.96	0.359	1975.54	1975.03	4.449	-0.04	-0.51		1976
18	17.96	0.367	1987.76	1987.26	4.436	-0.04	-0.50		1988
19	18.96	0.375	1999.68	1999.19	4.425	-0.04	-0.49		2000
20	19.96	0.382	2011.35	2010.88	4.414	-0.04	-0.47		2011
21	20.96	0.387	*2022.81*	*2022.35*	4.402	-0.04	-0.46		*2033*
22	21.96	0.391	2034.09	2033.64	4.391	-0.04	-0.45		2034
23	22.96	0.395	2045.24	2044.80	4.381	-0.04	-0.44		2045
24	23.96	0.397	2056.27	2055.85	4.368	-0.04	-0.42		2056
25	24.96	0.399	*2067.22*	*2066.82*	4.356	-0.04	-0.40		*2067*
26	25.96	0.399	2078.12	2077.73	4.346	-0.04	-0.39		2078
27	26.97	0.399	2088.99	2088.62	4.334	-0.03	-0.37		2089
28	27.97	0.397	**2099.86**	**2099.51**	4.322	-0.03	-0.35		**2100**
29	28.97	0.395	2110.75	2110.42	4.310	-0.03	-0.33		2111
30	29.97	0.392	2121.69	2121.38	4.297	-0.03	-0.31		2122
31	30.97	0.387	2132.70	2132.42	4.284	-0.03	-0.28		2133
32	31.98	0.382	2143.81	2143.55	4.272	-0.02	-0.26		2144
33	32.98	0.376	2155.05	2154.82	4.259	-0.02	-0.23		2155
34	33.98	0.369	2166.45	2166.25	4.245	-0.02	-0.20		2166
35	34.98	0.362	2178.03	2177.86	4.233	-0.02	-0.17		2178
36	35.99	0.353	2189.85	2189.71	4.220	-0.01	-0.14		2190
37	36.99	0.344	2201.93	2201.83	4.205	-0.01	-0.10		2202
38	37.99	0.334	2214.32	2214.26	4.193	-0.01	-0.06		2214
39	38.99	0.322	2227.08	2227.06	4.179	-0.01	-0.02		2227
40	40.00	0.310	2240.26	2240.29	4.169	-0.00	0.03		2240
41	41.00	0.297	2253.95	2254.02	4.156	0.00	0.07		2254
42	42.00	0.284	2268.21	2268.33	4.144	0.00	0.12		2268
43	43.01	0.269	2283.17	2283.35	4.136	0.01	0.18		2283
44	44.01	0.253	2298.96	2299.20	4.130	0.01	0.24		2299
45	45.01	0.237	2315.75	2316.05	4.135	0.01	0.30		2316
46	46.01	0.219	2333.84	2334.20	4.133	0.01	0.36		2334
47	47.01	0.201	2353.40	2353.84	4.131	0.01	0.44		2353
48	48.02	0.181	2374.92	2375.45	4.140	0.02	0.53		2375
49	49.02	0.160	**2399.00**	**2399.62**	4.159	0.02	0.62		**2400**
50	50.02	0.138	2426.62	2427.36	4.192	0.02	0.74		2427
51	51.01	0.114	2459.37	2460.27	4.250	0.01	0.90		2459
52	52.01	0.089	2500.26	2501.39	4.369	0.01	1.13		2500
53	53.01	0.062	2556.02	2557.62	4.711	0.01	1.60		2556
54	54.01	0.032	2648.14	2649.21	3.613	0.01	1.07		2648
55	54.70	0.010	2776.26	2738.06	1.274	-0.30	-38.20		2776

Table 9. Summary of July Exit Level Item-Level Analyses and Subgroup Analyses

Grade/ Subject	Number of Items with Significant Mode Differences*	Mean RS Difference (Effect Size) between Online and Paper over Replications**	Subgroup Analysis--Mean RS Difference (Effect Size) between Online and Paper over Replications**				
			Gender		Ethnicity		
			Male	Female	African American	Hispanic	White
Exit Level ELA	5 (52)	**-0.84 (-0.07)**	-0.93 (-0.08)	-0.74 (-0.07)	-0.80 (-0.08)	**-0.95 (-0.08)**	-0.05 (<-0.01)
Exit Level Math	23 (60)	0.07 (0.01)	0.24 (0.03)	-0.05 (-0.01)	0.05 (0.01)	0.11 (0.02)	-0.04 (-0.01)
Exit Level Science	11 (55)	<-0.01 (<-0.01)	**0.50 (0.08)**	-0.24 (-0.04)	0.10 (0.02)	-0.17 (-0.03)	0.29 (0.05)
Exit Level Social Studies	7 (55)	0.06 (0.01)	1.03 (0.11)	-0.50 (-0.06)	-0.16 (-0.02)	<0.01 (<-0.01)	0.55 (0.06)

*: Items with significant mean differences (p<.05) over replications. Total number of items in parentheses.
**: RS Significant Differences (p<.05) are in bold-face. Negative values indicate a lower mean RS for the online group.

Table 10. Summary of Item-Level Analysis – July Exit Level ELA

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
1	0.73	0.77	-0.04	0.02	-2.23	*	-0.09
2	0.79	0.81	-0.03	0.02	-1.74		-0.07
3	0.89	0.90	-0.01	0.01	-1.19		-0.05
4	0.74	0.74	0.00	0.02	-0.22		-0.01
5	0.70	0.74	-0.04	0.02	-2.00		-0.09
6	0.83	0.83	0.00	0.02	-0.25		-0.01
7	0.70	0.75	-0.05	0.02	-2.79	*	-0.11
8	0.54	0.57	-0.03	0.02	-1.61		-0.07
9	0.80	0.80	0.01	0.02	0.52		0.02
10	0.70	0.72	-0.01	0.02	-0.70		-0.03
11	0.69	0.69	0.01	0.02	0.47		0.02
12	0.62	0.64	-0.02	0.02	-0.95		-0.04
13	0.57	0.58	-0.01	0.02	-0.51		-0.02
14	0.72	0.75	-0.02	0.02	-1.35		-0.05
15	0.59	0.60	-0.02	0.02	-0.89		-0.04
16	0.57	0.58	-0.01	0.02	-0.62		-0.03
17	0.76	0.78	-0.02	0.02	-0.96		-0.04
18	0.66	0.70	-0.04	0.02	-1.95		-0.08
19	0.61	0.63	-0.01	0.02	-0.57		-0.02
20	0.81	0.84	-0.03	0.01	-1.86		-0.07
21	0.73	0.74	-0.01	0.02	-0.50		-0.02
22	0.46	0.47	-0.02	0.02	-0.81		-0.03
23	0.71	0.73	-0.02	0.02	-1.43		-0.06
24	0.57	0.58	0.00	0.02	-0.04		0.00
25	0.60	0.62	-0.03	0.02	-1.28		-0.05
26	0.85	0.88	-0.02	0.01	-1.78		-0.07
27	0.87	0.87	0.00	0.01	0.06		0.00
28	0.67	0.69	-0.02	0.02	-1.21		-0.05
29	1.20	1.18	0.02	0.02	0.97		0.04
30	1.04	1.00	0.04	0.02	1.46		0.05
31	0.87	0.80	0.07	0.03	2.68	*	0.10
32	0.78	0.75	0.03	0.02	1.68		0.07
33	0.83	0.83	-0.01	0.02	-0.60		-0.02
34	0.63	0.65	-0.02	0.02	-0.79		-0.03
35	0.74	0.77	-0.03	0.02	-1.58		-0.06
36	0.85	0.84	0.01	0.02	0.44		0.02
37	0.57	0.53	0.04	0.02	1.93		0.08

Table 10. Summary of Item-Level Analysis – July Exit Level ELA (Continued)

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
38	0.62	0.60	0.02	0.02	1.18		0.05
39	0.75	0.75	0.00	0.02	0.04		0.00
40	0.70	0.66	0.03	0.02	1.68		0.07
41	0.64	0.65	-0.01	0.02	-0.63		-0.02
42	0.52	0.53	0.00	0.02	-0.10		0.00
43	0.80	0.78	0.02	0.02	1.17		0.05
44	0.53	0.54	-0.01	0.02	-0.68		-0.03
45	0.44	0.46	-0.02	0.02	-0.92		-0.04
46	0.68	0.67	0.01	0.02	0.64		0.03
47	0.75	0.77	-0.01	0.02	-0.81		-0.03
48	0.43	0.44	-0.01	0.02	-0.42		-0.02
49	0.73	0.68	0.05	0.02	2.51	*	0.10
50	0.56	0.56	0.00	0.02	-0.17		-0.01
51	0.72	0.74	-0.01	0.02	-0.87		-0.03
52	6.84	7.38	-0.54	0.13	-4.09	*	-0.15

Table 11. Summary of Item-Level Analysis – July Exit Level Mathematics

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
1	0.77	0.80	-0.03	0.01	-2.84	*	-0.07
2	0.68	0.72	-0.03	0.01	-2.95	*	-0.07
3	0.78	0.78	0.00	0.01	0.41		0.01
4	0.86	0.86	0.00	0.01	0.12		0.00
5	0.49	0.47	0.03	0.01	2.04	*	0.06
6	0.49	0.48	0.01	0.01	0.74		0.02
7	0.43	0.42	0.02	0.01	1.34		0.03
8	0.55	0.62	-0.07	0.01	-6.21	*	-0.15
9	0.38	0.43	-0.05	0.01	-3.85	*	-0.09
10	0.47	0.46	0.01	0.01	0.93		0.02
11	0.22	0.24	-0.02	0.01	-1.80		-0.04
12	0.46	0.44	0.02	0.01	1.88		0.04
13	0.37	0.34	0.03	0.01	2.17	*	0.06
14	0.32	0.36	-0.04	0.01	-2.98	*	-0.08
15	0.46	0.42	0.04	0.01	3.67	*	0.08
16	0.30	0.32	-0.02	0.01	-1.58		-0.03
17	0.42	0.40	0.01	0.01	1.26		0.03
18	0.24	0.23	0.01	0.01	1.03		0.03
19	0.53	0.59	-0.06	0.01	-4.53	*	-0.13
20	0.42	0.42	0.00	0.01	-0.22		-0.01
21	0.33	0.37	-0.04	0.01	-3.91	*	-0.09
22	0.53	0.48	0.05	0.01	4.08	*	0.10
23	0.50	0.50	0.01	0.01	0.54		0.01
24	0.45	0.46	-0.01	0.01	-0.54		-0.01
25	0.29	0.29	0.00	0.01	-0.07		0.00
26	0.28	0.31	-0.03	0.01	-2.41	*	-0.06
27	0.41	0.40	0.01	0.01	0.65		0.02
28	0.59	0.59	0.00	0.01	0.22		0.01
29	0.40	0.39	0.00	0.01	0.33		0.01
30	0.48	0.48	0.00	0.01	-0.10		0.00
31	0.32	0.33	-0.01	0.01	-0.93		-0.02
32	0.69	0.66	0.03	0.01	2.56	*	0.06
33	0.43	0.47	-0.03	0.01	-2.57	*	-0.06
34	0.50	0.49	0.02	0.01	1.44		0.03
35	0.38	0.39	-0.01	0.01	-0.92		-0.02
36	0.46	0.42	0.03	0.01	2.53	*	0.06
37	0.29	0.27	0.03	0.01	2.44	*	0.06

Table 11. Summary of Item-Level Analysis – July Exit Level Mathematics (Continued)

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
38	0.32	0.37	-0.05	0.01	-3.93	*	-0.10
39	0.58	0.50	0.07	0.01	7.00	*	0.15
40	0.45	0.44	0.01	0.01	0.80		0.02
41	0.34	0.35	-0.01	0.01	-0.90		-0.02
42	0.46	0.45	0.02	0.01	1.59		0.04
43	0.63	0.61	0.02	0.01	1.30		0.03
44	0.29	0.28	0.01	0.01	0.88		0.02
45	0.37	0.34	0.03	0.01	2.46	*	0.07
46	0.51	0.50	0.02	0.01	1.40		0.04
47	0.56	0.57	-0.01	0.01	-0.77		-0.02
48	0.59	0.56	0.03	0.01	2.69	*	0.06
49	0.49	0.52	-0.03	0.01	-2.31	*	-0.06
50	0.22	0.23	-0.01	0.01	-0.95		-0.02
51	0.69	0.67	0.02	0.01	1.50		0.04
52	0.38	0.40	-0.03	0.01	-2.20	*	-0.06
53	0.41	0.38	0.03	0.01	2.42	*	0.06
54	0.44	0.42	0.02	0.01	1.61		0.04
55	0.53	0.53	-0.01	0.01	-0.44		-0.01
56	0.59	0.59	0.00	0.01	-0.16		0.00
57	0.73	0.72	0.02	0.01	1.43		0.03
58	0.56	0.57	-0.01	0.01	-0.76		-0.02
59	0.60	0.60	0.00	0.01	-0.34		-0.01
60	0.60	0.58	0.02	0.01	1.58		0.04

Table 12. Summary of Item-Level Analysis – July Exit Level Science

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
1	0.69	0.72	-0.03	0.01	-2.94	*	-0.07
2	0.61	0.66	-0.05	0.01	-3.87	*	-0.11
3	0.81	0.79	0.01	0.01	1.21		0.03
4	0.66	0.67	-0.01	0.01	-0.48		-0.01
5	0.55	0.56	-0.01	0.01	-0.81		-0.02
6	0.43	0.45	-0.02	0.01	-1.31		-0.03
7	0.72	0.70	0.02	0.01	1.74		0.04
8	0.53	0.51	0.02	0.01	1.82		0.04
9	0.41	0.41	0.00	0.01	-0.14		0.00
10	0.52	0.50	0.02	0.01	1.31		0.03
11	0.48	0.46	0.03	0.01	2.37	*	0.06
12	0.42	0.41	0.01	0.01	0.64		0.02
13	0.51	0.47	0.04	0.01	3.56	*	0.09
14	0.25	0.25	0.00	0.01	-0.09		0.00
15	0.49	0.51	-0.02	0.01	-1.37		-0.03
16	0.38	0.38	0.00	0.01	0.00		0.00
17	0.58	0.56	0.02	0.01	1.46		0.03
18	0.56	0.54	0.02	0.01	1.83		0.04
19	0.65	0.66	-0.01	0.01	-0.73		-0.02
20	0.32	0.31	0.01	0.01	1.28		0.03
21	0.45	0.43	0.03	0.01	2.16	*	0.05
22	0.45	0.45	0.00	0.01	0.15		0.00
23	0.37	0.40	-0.03	0.01	-2.36	*	-0.06
24	0.31	0.31	0.01	0.01	0.50		0.01
25	0.26	0.27	-0.01	0.01	-0.54		-0.01
26	0.23	0.23	0.01	0.01	0.57		0.01
27	0.53	0.54	-0.01	0.01	-0.81		-0.02
28	0.56	0.58	-0.02	0.01	-1.67		-0.04
29	0.48	0.51	-0.03	0.01	-2.13	*	-0.06
30	0.25	0.23	0.01	0.01	0.91		0.02
31	0.26	0.27	-0.01	0.01	-1.62		-0.03
32	0.44	0.41	0.03	0.01	2.32	*	0.06
33	0.32	0.34	-0.01	0.01	-0.92		-0.02
34	0.45	0.41	0.04	0.01	3.48	*	0.08
35	0.41	0.43	-0.02	0.01	-1.43		-0.03
36	0.40	0.41	-0.01	0.01	-0.70		-0.02
37	0.25	0.26	-0.01	0.01	-0.48		-0.01

Table 12. Summary of Item-Level Analysis – July Exit Level Science (Continued)

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
38	0.37	0.38	-0.01	0.01	-0.99		-0.03
39	0.44	0.43	0.01	0.01	0.84		0.02
40	0.40	0.38	0.02	0.01	1.51		0.04
41	0.42	0.43	-0.01	0.01	-0.66		-0.02
42	0.37	0.34	0.02	0.01	2.03	*	0.05
43	0.48	0.50	-0.02	0.01	-1.30		-0.03
44	0.41	0.42	-0.01	0.01	-0.96		-0.02
45	0.43	0.43	-0.01	0.01	-0.46		-0.01
46	0.67	0.69	-0.02	0.01	-1.53		-0.04
47	0.59	0.58	0.01	0.01	0.59		0.01
48	0.71	0.73	-0.02	0.01	-1.71		-0.04
49	0.69	0.69	-0.01	0.01	-0.52		-0.01
50	0.71	0.69	0.02	0.01	1.89		0.04
51	0.51	0.51	0.01	0.01	0.55		0.01
52	0.59	0.61	-0.02	0.01	-1.94		-0.05
53	0.71	0.71	0.00	0.01	-0.37		-0.01
54	0.69	0.70	-0.02	0.01	-1.38		-0.03
55	0.85	0.83	0.02	0.01	2.04	*	0.05

Table 13. Summary of Item-Level Analysis – July Exit Level Social Studies

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
1	0.78	0.79	-0.01	0.02	-0.27		-0.01
2	0.44	0.40	0.05	0.03	1.47		0.09
3	0.69	0.70	-0.01	0.03	-0.29		-0.02
4	0.52	0.50	0.02	0.03	0.74		0.04
5	0.68	0.64	0.04	0.03	1.47		0.09
6	0.62	0.58	0.03	0.03	1.16		0.07
7	0.40	0.39	0.01	0.03	0.24		0.01
8	0.64	0.69	-0.05	0.02	-2.16	*	-0.11
9	0.59	0.63	-0.03	0.03	-1.14		-0.07
10	0.78	0.74	0.04	0.02	1.58		0.09
11	0.52	0.55	-0.03	0.03	-1.02		-0.06
12	0.56	0.56	0.01	0.03	0.21		0.01
13	0.60	0.58	0.02	0.03	0.73		0.04
14	0.57	0.54	0.03	0.02	1.42		0.07
15	0.44	0.47	-0.03	0.03	-0.96		-0.06
16	0.54	0.47	0.07	0.03	2.56	*	0.13
17	0.49	0.45	0.04	0.03	1.44		0.08
18	0.42	0.41	0.01	0.03	0.25		0.01
19	0.48	0.52	-0.04	0.03	-1.23		-0.07
20	0.47	0.45	0.02	0.03	0.87		0.05
21	0.40	0.44	-0.04	0.03	-1.49		-0.08
22	0.44	0.41	0.03	0.03	1.14		0.06
23	0.51	0.49	0.02	0.03	0.70		0.04
24	0.26	0.31	-0.05	0.02	-2.29	*	-0.12
25	0.37	0.38	-0.02	0.03	-0.58		-0.03
26	0.34	0.30	0.03	0.03	1.18		0.07
27	0.25	0.31	-0.06	0.02	-2.32	*	-0.12
28	0.30	0.31	-0.01	0.03	-0.22		-0.01
29	0.31	0.29	0.02	0.03	0.88		0.05
30	0.24	0.27	-0.02	0.03	-0.84		-0.05
31	0.23	0.20	0.03	0.02	1.31		0.07
32	0.35	0.36	-0.01	0.03	-0.46		-0.03
33	0.69	0.65	0.04	0.03	1.51		0.09
34	0.43	0.42	0.01	0.03	0.23		0.01
35	0.49	0.49	0.00	0.03	-0.14		-0.01
36	0.37	0.36	0.00	0.03	0.09		0.01
37	0.43	0.43	-0.01	0.03	-0.33		-0.02

Table 13. Summary of Item-Level Analysis – July Exit Level Social Studies (Continued)

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
38	0.39	0.36	0.04	0.03	1.31		0.08
39	0.26	0.26	0.00	0.03	0.03		0.00
40	0.48	0.48	0.01	0.03	0.23		0.01
41	0.46	0.46	0.00	0.03	0.00		0.00
42	0.54	0.54	0.00	0.03	0.08		0.00
43	0.56	0.55	0.01	0.03	0.30		0.02
44	0.34	0.28	0.06	0.02	2.31	*	0.12
45	0.58	0.54	0.04	0.03	1.37		0.08
46	0.62	0.64	-0.02	0.03	-0.81		-0.04
47	0.43	0.45	-0.01	0.03	-0.52		-0.03
48	0.66	0.72	-0.06	0.03	-2.45	*	-0.14
49	0.58	0.62	-0.03	0.03	-1.30		-0.07
50	0.59	0.62	-0.03	0.03	-1.16		-0.06
51	0.69	0.69	0.01	0.03	0.23		0.01
52	0.42	0.42	0.00	0.03	0.13		0.01
53	0.42	0.43	0.00	0.03	-0.09		-0.01
54	0.41	0.48	-0.07	0.03	-2.80	*	-0.15
55	0.17	0.18	-0.01	0.02	-0.50		-0.03