

Results of July 2007 TAKS Exit Level Retest Online Comparability Study Analyses

Summary

The Texas Assessment of Knowledge and Skills (TAKS) has been offered online (in addition to paper) for certain grades and subjects 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.

Texas uses a “matched samples comparability analyses” (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 many times. Conversion tables for each replication are retained and aggregated to obtain the final conversion and standard error of linking. The equivalency between online and paper scale scores is then evaluated. A separate conversion table is recommended if the two sets of scores are considered not comparable. Between-mode item level analyses and subgroup analyses for ethnic and gender groups are also performed.

Using this methodology, the comparability analyses were carried out for the July 2007 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 the standard error of linking, and raw score cuts at various performance levels), the comparability between online and paper administrations was assessed. A mode effect between online and paper was found for ELA and mathematics. For ELA, the online version was 1 point easier at the ‘Commended’ level; whereas for mathematics, the online version was 2 points harder at the ‘Met the Standard’ level.

The subgroup analysis results indicated that the online version of the mathematics test appeared more difficult for all subgroups. The item analysis results indicated that for most items where there was a mode effect, students tested in paper scored higher.

The following table summarizes the results for each grade/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
	Math	X	X	X	
	Science				
	Social Studies			X	

Based on the results, Pearson recommended using paper conversion tables for July 2007 exit level online retest in science and social studies and using a separate conversion table for the online retests in ELA and mathematics.

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 (Grade 11) retests of TAKS in English language arts (ELA), mathematics, social studies, and science. 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 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 addition, exit level retest online was offered in July and will be offered again in October. This report describes the results of the comparability studies for the July 2007 TAKS exit level retest online administrations, which took place from July 10 to July 13.

Methodology

In 2005, TEA and Pearson devised a matched samples comparability analysis 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 comparability analyses. For the July 2007 exit level retest comparability studies, a composite of the previous scores in ELA, mathematics, science, and social studies was created. This composite was then used in the matching of samples. A description of this process is included below.

- 1) Using the students who tested in paper, we regressed their July 2007 retest raw score on their ELA, mathematics, science, and social studies scale scores from the spring 2007 primary administration.

For exit level:

$$\hat{Y}_{predicted_rawscore} = \beta_0 + \beta_1 X_{1(prev_ELA)} + \beta_2 X_{2(prev_Math)} + \beta_3 X_{3(prev_SocialStudies)} + \beta_4 X_{4(prev_Science)}.$$

- 2) The resulting regression weights were applied to all students (paper and online) to obtain an estimated raw score (\hat{y}) for each of the students.
- 3) Students were then broken into 20 groups based on the estimated raw score.
- 4) This resulted in a 20 (previous score groups) by 4 (ethnicity groups) by 2 (gender groups) grid that was used in the matched sampling.
- 5) To improve optimal matching, students with missing values on any of the matching variables were dropped from the study.

Table 1 displays the demographic information for the online and paper samples for each subject area tested. The online and paper samples tended to differ in terms of ethnicity proportions. For example, according to table 1, more White and African American students and fewer Hispanic students participated online than on paper. There appears to be no participation differences by gender with regard to the mode tested. In addition, each mode has about the same percentage of special education students. It can also be seen that for most subjects, online students generally had lower raw scores than the paper students, although for social studies and science the difference appeared to be very small. The only exception was for exit level ELA, where the online students scored higher than the paper students on average. Table 2 provides the performance for exit level ELA students by testing mode. Students tested online performed better on the essay than students tested in paper. However, for the multiple-choice items the performance levels appear to be about the same. Studies from previous comparability analyses on ELA have consistently indicated that there exists an item type by mode interaction, namely students tested online have a higher essay score, whereas students tested in paper have a higher scores on multiple-choice items.

The matched samples comparability analysis method is described in the steps below.

1. All students eligible for matching were placed into 20 groups based on the regression of July 2007 retest raw score on the Spring 2007 scale scores as described earlier. Each student testing online with complete data was matched to a student from the available 2007 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 2007 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 paper 2007 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 results, two Pearson psychometricians programmed independent versions of the analysis. The online results based on these two analyses were 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, we paid special attention to differences in the range of scale scores around the “Met the Standard” and “Commended” score levels. Differences at the extremes of the scale are less important, given the purpose and primary uses of the TAKS tests. This standard error procedure is sensitive to sample size such that the standard errors will be greater when the sample sizes are smaller. Therefore, we also considered 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 scaled scores are reported in 10-point units. For a given raw score, if the resulting scales scores from the linking differed by fewer than 5 points, then the scale scores would ideally 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 we considered is 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 result 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 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 use of an alternate score table for the online TAKS administration.

Results

Table 3 provides an overview of the comparisons between online and paper scores. Test-level mode effect was detected for ELA and mathematics, where math was harder online at the ‘Met the Standard’ cut but ELA was easier online at the ‘Commended’ cut. No test-level mode effect was detected for science or social studies.

Tables 4 to 7 detail the comparisons between online and paper for each test. The columns of the tables are as follows:

RS – 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 2007 July TAKS 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 online 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.

For ELA, the differences in raw scores were more than half of a raw score point and the scale score differences were statistically significant at the higher end of the scale. The “Met the Standard” performance level did correspond to a same raw score for the online version compared to the paper version, however, the “Commended” cut score was 1 point higher for online. It should be mentioned that (as has consistently been observed in previous ELA comparability studies) raw score differences appear to be in favor of paper students at the lower end of the scale, and in favor of the online students at the higher end. However, only the differences at the higher end resulted in a significant difference (as manifested by the difference in the “Commended” performance level cut).

For mathematics, the online version of the test was more difficult. The online “Met the Standard” cut score was 2 points lower, and the scale score differences reached statistical significance across most of the scale range.

For science, although the raw score differences were not marked as both practically and statistically significant at any raw score level (refer to table 6, the SIG column), the overall mean raw score difference for the bootstrapped samples is statistically significant (table 8). These are not inconsistent, since the raw score differences/scale score differences at each raw score level in table 6 were indeed statistically significant—the magnitudes, however, were not practically significant.

For social studies the ‘Met the Standard’ cut was one point lower online; however, the raw score differences were less than 0.5 across the score range, and the scale score differences were statistically non-significant across the score range.

For all four subject areas, large differences occur at the lowest and highest scaled score points, because WINSTEPS (the IRT calibration software used in the study) does not estimate abilities for zero and perfect scores and the comparability WINSTEPS runs were centered differently than the paper runs. These differences are not meaningful. Therefore, for each test the online conversion table (if recommended) will have the scale score at the extreme ends set to the paper, as has been consistently done in the past.

Impact Data Analyses

The pass-rate-comparison in table 3 indicates that in general, applying the paper conversion tables to the online students results in a lower percentage of students achieving the ‘Met the Standard’ level; except for science, where there is no difference between the pass-rates regardless of which conversion table is used for online students. Using the paper conversion tables for science and social studies resulted in passing rates more similar to paper for the online 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 item-level analysis compares mean differences of each item between the two testing modes across replications, whereas 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.

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 are listed in table 8. A significance test was performed for each ‘matched’ subgroup using the following equation:

$$Z_{dif} = \frac{\bar{D}_{Diff}}{\sqrt{SE_{Diff}^2}}$$

where \bar{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} and is the bootstrap standard error of the mean differences over the replications, also for each subgroup.

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

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

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 8, there was a consistent significant mode effect for mathematics (online harder). For ELA, large significant mode effects were present for white students (online easier) and for Hispanics students (online harder). For science, significant mode effects were indicated for Hispanics students (online harder), and similarly for male and female (online harder). No significant mode effect for the subgroups was found for social studies.

It is interesting to note the differential mode effect for ELA between White students and Hispanic students. A similar finding was present in 2006 October exit level retest ELA.

Item-Level Analysis

Item-level analysis was performed in a similar way. 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 9 to 12 display the results of item-level comparison across replications for each subject area tested. 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, when there was a significant between-mode mean difference, the paper group tended to score higher for math and science for more items. Specifically, for math, the paper group scored higher on 22 items of the 25 items exhibiting a mode effect; and the online group higher on the other three. For science, the paper group scored higher on 13 of the 17 items exhibiting a mode effect, where the online group scored higher on the other four.

For ELA and social studies, few items exhibited a mode effect. There were six items each in social studies and ELA that showed a mode effect. In social studies, four items were in favor of the paper group and two were in favor of the online group. For ELA, the online group scored lower on the five multiple choice items but scored higher on the essay item.

It should also be noted that for most of the differences, the standard errors were relatively small—which might have contributed to the large number of items showing statistically significant difference between the testing modes for some tests. To help determine the ‘practical’ significance of the differences, the average effect size for each difference was calculated and listed on the last column of these tables. It can be seen that the magnitude of the effect sizes ranged between 0 and 0.22.

Summary and Recommendation

For ELA and mathematics, statistically and practically significant between-mode differences were found. There were also between-mode differences in the performance level cuts. For social studies, although the “Met the Standard” cut scores were different between online and paper, the scale score differences at the cut score levels did not exceed two standard errors of the linking, and the mode effects were not found according to the other criteria. For science, none of the criterion yielded significant differences between online and paper modes.

Based on the comparisons, Pearson recommended using a separate online conversion table for ELA and mathematics, and using the paper pre-equated conversion table for 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). *Standards 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. Student Demographic Information for July 2007 Comparability Analyses*

Grade	Subject	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***	
		CBT	PAP	CBT	PAP	CBT	PAP	CBT	PAP	CBT	PAP	CBT	PAP	CBT	PAP	CBT	PAP	CBT	PAP	CBT	PAP
Exit Level	ELA	123	1023	853	8648	43.67	42.98	43.53	42.98	57	54	28	18	45	61	23	19	5	3	7	7
	Mathematics	157	1215	2643	21064	25.19	26.15	26.43	26.15	39	39	27	23	42	52	29	24	2	1	6	5
	Science	168	1228	3065	25015	25.84	26.03	26.25	26.03	37	38	24	19	45	57	28	22	3	2	4	4
	Social Studies	131	910	591	5446	26.36	26.39	26.54	26.39	32	31	20	14	50	64	28	20	3	2	4	5

* 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 2. Student Performance on July 2007 Exit Level Retest ELA Multiple Choice, Open-Ended Items, and Essay by Test Mode

	ONLINE				PAPER					
Multiple Choice	33.81				33.68					
	0	1	2	3	0	1	2	3		
OE 1	17.47%	64.95%	17.58%	0.00%	20.00%	63.06%	16.94%	0.00%		
OE 2	11.49%	67.64%	20.87%	0.00%	10.88%	64.03%	25.07%	0.02%		
OE 3	27.20%	54.75%	18.05%	0.00%	27.16%	55.94%	16.89%	0.00%		
	0	1	2	3	4	0	1	2	3	4
Essay	3.52%	35.64%	47.71%	12.08%	1.06%	2.61%	42.60%	50.02%	4.60%	0.16%

Table 3. Summary of the July 2007 Exit Level Retest 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****	853	8648	41	41	46.3	50.1	49.8	14 (73)	Alternate Conversion Table (Mode Effect)
			62	61	1.3	2.8	4.0		
Exit Level Math	2643	21064	30	32	22.0	24.6	16.9	46(60)	Alternate Conversion Table (Mode Effect)
			53	53	<0.1	<0.1	<0.1		
Exit Level Science	3065	25015	29	29	33.9	32.0	32.0	0 (55)	Paper Conversion Table (No Mode Effect)
			50	50	<0.1	<0.1	<0.1		
Exit Level Social Studies	591	5446	27	28	42.0	48.6	42.5	0 (55)	Paper Conversion Table (No Mode Effect)
			49	49	0.1	<0.1	<0.1		

*: 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 4. 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	0.33	0.040	1417.39	1454.98	4.481	0.33	37.59		1417
1	1.10	0.118	1530.15	1537.90	10.140	0.10	7.75		1538
2	2.17	0.202	1611.17	1618.93	10.896	0.17	7.76		1619
3	3.23	0.263	1659.58	1667.23	9.801	0.23	7.65		1667
4	4.27	0.307	1694.64	1701.88	8.883	0.27	7.23		1702
5	5.30	0.338	1722.42	1729.18	8.091	0.30	6.76		1729
6	6.32	0.361	1745.61	1751.91	7.407	0.32	6.30		1752
7	7.34	0.377	1765.64	1771.51	6.817	0.34	5.87		1772
8	8.35	0.389	1783.35	1788.83	6.319	0.35	5.48		1789
9	9.36	0.398	1799.32	1804.46	5.901	0.36	5.14		1804
10	10.36	0.405	1813.94	1818.75	5.539	0.36	4.82		1819
11	11.36	0.411	1827.44	1831.98	5.236	0.36	4.54		1832
12	12.36	0.415	1840.06	1844.34	4.976	0.36	4.28		1844
13	13.36	0.419	1851.93	1855.98	4.756	0.36	4.05		1856
14	14.36	0.423	1863.18	1867.02	4.571	0.36	3.84		1867
15	15.36	0.427	1873.90	1877.54	4.408	0.36	3.64		1878
16	16.35	0.430	1884.16	1887.63	4.276	0.35	3.47		1888
17	17.35	0.433	1894.04	1897.35	4.161	0.35	3.30		1897
18	18.34	0.437	1903.59	1906.74	4.066	0.34	3.15		1907
19	19.34	0.440	1912.85	1915.86	3.987	0.34	3.01		1916
20	20.33	0.443	1921.87	1924.74	3.915	0.33	2.87		1925
21	21.32	0.446	1930.66	1933.41	3.856	0.32	2.75		1933
22	22.31	0.448	1939.28	1941.90	3.806	0.31	2.63		1942
23	23.30	0.451	1947.73	1950.24	3.763	0.30	2.51		1950
24	24.29	0.453	1956.05	1958.44	3.728	0.29	2.39		1958
25	25.28	0.455	1964.25	1966.54	3.702	0.28	2.29		1967
26	26.27	0.456	1972.38	1974.56	3.676	0.27	2.18		1975
27	27.26	0.457	1980.42	1982.49	3.655	0.26	2.07		1982
28	28.25	0.457	1988.41	1990.38	3.642	0.25	1.97		1990
29	29.23	0.457	1996.38	1998.23	3.627	0.23	1.85		1998
30	30.22	0.456	2004.31	2006.06	3.619	0.22	1.75		2006
31	31.21	0.454	2012.26	2013.89	3.607	0.21	1.63		2014
32	32.19	0.452	2020.19	2021.71	3.603	0.19	1.52		2022
33	33.17	0.450	2028.17	2029.57	3.602	0.17	1.40		2030
34	34.16	0.446	2036.19	2037.46	3.596	0.16	1.27		2037
35	35.14	0.443	<u>2044.26</u>	<u>2045.39</u>	3.596	0.14	1.14		<u>2045</u>
36	36.12	0.438	<u>2052.42</u>	<u>2053.41</u>	3.598	0.12	0.99		<u>2053</u>
37	37.10	0.433	2060.67	2061.50	3.602	0.10	0.84		2062
38	38.08	0.427	<u>2069.04</u>	<u>2069.71</u>	3.607	0.08	0.67		<u>2072</u>
39	39.05	0.420	2077.55	2078.03	3.609	0.05	0.48		2078
40	40.03	0.413	2086.21	2086.49	3.615	0.03	0.28		2086
41	41.00	0.405	2095.06	2095.12	3.626	0.00	0.05		2100
42	41.97	0.396	2104.13	2103.93	3.633	-0.03	-0.20		2104
43	42.94	0.387	2113.42	2112.94	3.642	-0.06	-0.48		2113
44	43.91	0.377	2123.00	2122.19	3.654	-0.09	-0.81		2122
45	44.88	0.366	2132.88	2131.70	3.664	-0.12	-1.18		2132
46	45.84	0.355	2143.10	2141.50	3.679	-0.16	-1.60		2141
47	46.80	0.344	2153.72	2151.63	3.696	-0.20	-2.09		2152
48	47.76	0.332	2164.79	2162.12	3.715	-0.24	-2.66		2162
49	48.71	0.321	2176.35	2173.02	3.739	-0.29	-3.33		2173
50	49.66	0.309	2188.50	2184.38	3.776	-0.34	-4.12		2184
51	50.60	0.297	2201.31	2196.24	3.817	-0.40	-5.06		2196
52	51.54	0.286	2214.84	2208.66	3.873	-0.46	-6.18		2209
53	52.48	0.275	2229.23	2221.70	3.952	-0.52	-7.53		2222
54	53.40	0.265	2244.58	2235.42	4.055	-0.60	-9.16	*	2235
55	54.32	0.257	2261.01	2249.88	4.193	-0.68	-11.13	*	2250
56	55.23	0.251	2278.67	2265.15	4.372	-0.77	-13.52	*	2265
57	56.14	0.247	2297.70	2281.31	4.599	-0.86	-16.39	*	2281

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG	Final
58	57.03	0.246	2318.26	2298.44	4.894	-0.97	-19.82	*	2298
59	57.92	0.250	2340.51	2316.63	5.297	-1.08	-23.89	*	2317
60	58.79	0.260	2364.64	2335.99	5.880	-1.21	-28.64	*	2336
61	59.67	0.277	2390.82	2356.68	6.754	-1.33	-34.13	*	2357
62	60.54	0.307	2419.29	2378.91	8.065	-1.46	-40.39	*	2400
63	61.42	0.353	2450.34	2402.98	10.039	-1.58	-47.36	*	2403
64	62.32	0.422	2484.34	2429.43	13.049	-1.68	-54.91	*	2429
65	63.24	0.522	2521.79	2458.91	17.665	-1.76	-62.88	*	2459
66	64.19	0.654	2563.31	2492.26	24.450	-1.81	-71.05	*	2492
67	65.18	0.808	2609.80	2530.57	33.756	-1.82	-79.24	*	2531
68	66.22	0.972	2662.74	2576.11	45.696	-1.78	-86.64		2576
69	67.41	1.137	2724.52	2635.92	61.098	-1.59	-88.60		2636
70	68.95	1.257	2798.71	2731.25	82.567	-1.05	-67.45		2731
71	70.58	1.120	2890.56	2865.55	101.07	-0.42	-25.01		2866
72	71.87	0.691	3016.83	3005.36	82.859	-0.13	-11.47		3005
73	72.66	0.246	3157.50	3110.45	34.151	-0.34	-47.05		3158

Table 5. 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	0.32	0.004	1335.51	1381.21	0.567	0.32	45.70		1336
1	1.07	0.013	1477.72	1484.92	1.296	0.07	7.20		1485
2	2.14	0.024	1578.81	1587.39	1.479	0.14	8.58		1587
3	3.21	0.035	1639.30	1648.59	1.544	0.21	9.29		1649
4	4.28	0.045	1683.27	1693.02	1.572	0.28	9.75		1693
5	5.35	0.054	1718.21	1728.31	1.584	0.35	10.10		1728
6	6.41	0.063	1747.45	1757.82	1.588	0.41	10.37		1758
7	7.47	0.071	1772.77	1783.36	1.590	0.47	10.59		1783
8	8.53	0.078	1795.24	1806.01	1.589	0.53	10.77	*	1806
9	9.59	0.085	1815.53	1826.46	1.587	0.59	10.93	*	1826
10	10.64	0.092	1834.13	1845.20	1.585	0.64	11.07	*	1845
11	11.69	0.098	1851.36	1862.55	1.582	0.69	11.19	*	1863
12	12.74	0.104	1867.48	1878.78	1.578	0.74	11.30	*	1879
13	13.79	0.109	1882.67	1894.06	1.575	0.79	11.39	*	1894
14	14.84	0.114	1897.08	1908.56	1.569	0.84	11.48	*	1909
15	15.88	0.119	1910.83	1922.38	1.562	0.88	11.55	*	1922
16	16.92	0.123	1924.01	1935.63	1.555	0.92	11.62	*	1936
17	17.95	0.127	1936.71	1948.39	1.547	0.95	11.68	*	1948
18	18.99	0.131	1948.99	1960.73	1.541	0.99	11.74	*	1961
19	20.02	0.134	1960.91	1972.70	1.536	1.02	11.79	*	1973
20	21.05	0.137	1972.51	1984.36	1.533	1.05	11.85	*	1984
21	22.07	0.139	1983.85	1995.74	1.529	1.07	11.89	*	1996
22	23.10	0.142	1994.95	<u>2006.88</u>	1.527	1.10	11.93	*	<u>2015</u>
23	24.12	0.144	<u>2005.85</u>	2017.83	1.527	1.12	11.98	*	2018
24	25.14	0.145	2016.58	2028.60	1.525	1.14	12.02	*	2029
25	26.15	0.147	2027.17	2039.23	1.526	1.15	12.06	*	2039
26	27.17	0.148	2037.64	<u>2049.75</u>	1.525	1.17	12.11	*	<u>2058</u>
27	28.18	0.148	2048.03	<u>2060.17</u>	1.525	1.18	12.14	*	<u>2060</u>
28	29.19	0.149	<u>2058.34</u>	2070.52	1.525	1.19	12.18	*	2071
29	30.19	0.149	2068.61	2080.83	1.526	1.19	12.22	*	2081
30	31.20	0.149	2078.85	2091.10	1.526	1.20	12.25	*	2100
31	32.20	0.148	2089.09	2101.37	1.527	1.20	12.28	*	2101
32	33.20	0.148	2099.34	2111.66	1.528	1.20	12.32	*	2112
33	34.19	0.146	2109.63	2122.00	1.528	1.19	12.37	*	2122
34	35.19	0.145	2119.99	2132.39	1.531	1.19	12.40	*	2132
35	36.18	0.144	2130.43	2142.87	1.534	1.18	12.44	*	2143
36	37.16	0.142	2140.98	2153.46	1.534	1.16	12.48	*	2153
37	38.15	0.140	2151.67	2164.18	1.538	1.15	12.51	*	2164
38	39.13	0.137	2162.51	2175.07	1.540	1.13	12.56	*	2175
39	40.11	0.134	2173.56	2186.15	1.543	1.11	12.59	*	2186
40	41.09	0.131	2184.83	2197.45	1.546	1.09	12.62	*	2197
41	42.07	0.128	2196.36	2209.03	1.548	1.07	12.67	*	2209
42	43.04	0.124	2208.21	2220.91	1.550	1.04	12.70	*	2221
43	44.01	0.121	2220.41	2233.16	1.551	1.01	12.75	*	2233
44	44.97	0.116	2233.03	2245.82	1.550	0.97	12.79	*	2246
45	45.94	0.112	2246.13	2258.97	1.548	0.94	12.84	*	2259
46	46.90	0.107	2259.80	2272.69	1.544	0.90	12.89	*	2273
47	47.86	0.102	2274.13	2287.07	1.542	0.86	12.94	*	2287
48	48.81	0.096	2289.23	2302.17	1.539	0.81	12.94	*	2302
49	49.76	0.091	2305.18	2318.28	1.557	0.76	13.10	*	2318
50	50.71	0.085	2322.35	2335.55	1.569	0.71	13.20	*	2336
51	51.66	0.078	2340.90	2354.20	1.582	0.66	13.30	*	2354
52	52.60	0.071	2361.15	2374.59	1.600	0.60	13.44	*	2375
53	53.54	0.064	2383.59	2397.18	1.621	0.54	13.59	*	2400
54	54.47	0.056	2408.88	2422.68	1.651	0.47	13.80		2423
55	55.40	0.048	2438.11	2452.21	1.691	0.40	14.10		2452
56	56.33	0.040	2473.05	2487.61	1.755	0.33	14.56		2488
57	57.25	0.031	2517.04	2532.46	1.867	0.25	15.42		2532

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG Final
58	58.17	0.021	2577.58	2595.17	2.140	0.17	17.59	2595
59	59.09	0.011	2678.53	2691.11	1.542	0.09	12.58	2691
60	59.73	0.003	2819.22	2780.86	0.474	-0.27	-38.36	2819

Table 6. 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	0.31	0.003	1404.78	1445.28	0.439	0.31	40.50		1405
1	1.03	0.011	1535.70	1538.58	1.021	0.03	2.88		1539
2	2.06	0.021	1628.82	1632.21	1.191	0.06	3.39		1632
3	3.09	0.031	1684.73	1688.36	1.261	0.09	3.63		1688
4	4.12	0.040	1725.43	1729.21	1.298	0.12	3.78		1729
5	5.14	0.049	1757.82	1761.71	1.320	0.14	3.89		1762
6	6.17	0.057	1784.98	1788.96	1.339	0.17	3.98		1789
7	7.19	0.064	1808.62	1812.65	1.342	0.19	4.03		1813
8	8.22	0.071	1829.53	1833.61	1.348	0.22	4.08		1834
9	9.24	0.078	1848.46	1852.59	1.350	0.24	4.13		1853
10	10.26	0.084	1865.84	1870.01	1.353	0.26	4.17		1870
11	11.28	0.089	1881.99	1886.19	1.353	0.28	4.20		1886
12	12.30	0.095	1897.12	1901.36	1.352	0.30	4.24		1901
13	13.31	0.099	1911.42	1915.69	1.353	0.31	4.27		1916
14	14.33	0.104	1925.03	1929.33	1.350	0.33	4.30		1929
15	15.35	0.108	1938.05	1942.37	1.347	0.35	4.32		1942
16	16.36	0.111	1950.56	1954.90	1.346	0.36	4.34		1955
17	17.37	0.115	1962.65	1967.02	1.343	0.37	4.37		1967
18	18.38	0.117	1974.38	1978.77	1.339	0.38	4.39		1979
19	19.40	0.120	1985.79	1990.20	1.336	0.40	4.41		1990
20	20.41	0.122	1996.94	2001.36	1.330	0.41	4.42		2001
21	21.41	0.124	2007.83	2012.28	1.329	0.41	4.45		2012
22	22.42	0.125	2018.56	2023.03	1.327	0.42	4.47		2023
23	23.43	0.127	<u>2029.14</u>	<u>2033.63</u>	1.323	0.43	4.49		<u>2035</u>
24	24.44	0.127	2039.59	2044.10	1.321	0.44	4.51		2044
25	25.44	0.128	2049.95	2054.47	1.316	0.44	4.52		2054
26	26.44	0.128	<u>2060.23</u>	<u>2064.78</u>	1.316	0.44	4.55		<u>2068</u>
27	27.45	0.128	2070.48	2075.05	1.313	0.45	4.57		2075
28	28.45	0.128	2080.71	2085.30	1.310	0.45	4.59		2085
29	29.45	0.127	2090.94	2095.55	1.308	0.45	4.61		2100
30	30.45	0.127	2101.20	2105.84	1.309	0.45	4.64		2106
31	31.45	0.126	2111.53	2116.20	1.307	0.45	4.67		2116
32	32.45	0.124	2121.94	2126.64	1.306	0.45	4.70		2127
33	33.44	0.122	2132.46	2137.18	1.305	0.44	4.72		2137
34	34.44	0.121	2143.12	2147.88	1.307	0.44	4.76		2148
35	35.43	0.118	2153.96	2158.76	1.309	0.43	4.80		2159
36	36.43	0.116	2165.02	2169.85	1.309	0.43	4.83		2170
37	37.42	0.113	2176.33	2181.21	1.312	0.42	4.88		2181
38	38.41	0.110	2187.94	2192.86	1.315	0.41	4.92		2193
39	39.40	0.107	2199.90	2204.88	1.319	0.40	4.98		2205
40	40.39	0.103	2212.28	2217.32	1.325	0.39	5.04		2217
41	41.38	0.099	2225.16	2230.26	1.331	0.38	5.10		2230
42	42.37	0.095	2238.61	2243.78	1.339	0.37	5.17		2244
43	43.35	0.090	2252.75	2258.00	1.348	0.35	5.25		2258
44	44.33	0.085	2267.73	2273.07	1.358	0.33	5.34		2273
45	45.32	0.080	2283.70	2289.15	1.373	0.32	5.45		2289
46	46.30	0.074	2300.91	2306.48	1.389	0.30	5.57		2306
47	47.28	0.068	2319.66	2325.37	1.409	0.28	5.71		2325
48	48.25	0.062	2340.39	2346.27	1.436	0.25	5.88		2346
49	49.23	0.055	2363.73	2369.82	1.468	0.23	6.09		2370
50	50.20	0.047	2390.65	2397.01	1.515	0.20	6.36		2400
51	51.17	0.039	2422.81	2429.54	1.582	0.17	6.73		2430
52	52.13	0.031	2463.27	2470.59	1.698	0.13	7.32		2471
53	53.09	0.021	2518.92	2527.54	1.972	0.09	8.62		2528
54	54.05	0.011	2611.77	2618.14	1.437	0.05	6.37		2618
55	54.72	0.003	2741.15	2704.33	0.447	-0.28	-36.82		2741

#: Note that the paper conversion was used for the Final column.

Table 7. Summary of Comparability Analysis – July Exit Level Social Studies[#]

RAW	CBT_RS	RS_SD	PAP_SS	CBT_SS	SS_SD	RS_DIF	SS_DIF	SIG	Final
0	0.30	0.007	1445.46	1483.80	0.952	0.30	38.34		1445
1	1.00	0.024	1573.27	1572.98	2.604	0.00	-0.29		1573
2	2.00	0.047	1663.90	1663.45	3.311	0.00	-0.45		1663
3	3.01	0.069	1718.19	1718.11	3.136	0.01	-0.08		1718
4	4.01	0.089	1757.64	1757.79	3.087	0.01	0.15		1758
5	5.02	0.109	1789.08	1789.40	3.055	0.02	0.32		1789
6	6.03	0.127	1815.28	1815.75	3.030	0.03	0.47		1816
7	7.03	0.143	1837.98	1838.58	3.016	0.03	0.60		1839
8	8.04	0.159	1858.13	1858.85	3.005	0.04	0.72		1859
9	9.05	0.174	1876.35	1877.18	2.994	0.05	0.83		1877
10	10.06	0.188	1893.06	1894.00	2.985	0.06	0.94		1894
11	11.07	0.201	1908.57	1909.61	2.978	0.07	1.04		1910
12	12.09	0.213	1923.10	1924.24	2.971	0.09	1.14		1924
13	13.10	0.224	1936.82	1938.05	2.963	0.10	1.23		1938
14	14.11	0.234	1949.86	1951.18	2.956	0.11	1.32		1951
15	15.12	0.244	1962.34	1963.75	2.951	0.12	1.41		1964
16	16.13	0.252	1974.34	1975.83	2.943	0.13	1.49		1976
17	17.14	0.260	1985.92	1987.49	2.938	0.14	1.57		1987
18	18.15	0.266	1997.16	1998.81	2.933	0.15	1.65		1999
19	19.16	0.272	2008.11	2009.84	2.926	0.16	1.73		2010
20	20.17	0.278	2018.80	2020.60	2.923	0.17	1.80		2021
21	21.18	0.282	<u>2029.29</u>	<u>2031.16</u>	2.916	0.18	1.87		<u>2033</u>
22	22.19	0.286	2039.60	2041.54	2.909	0.19	1.94		2042
23	23.20	0.288	2049.76	2051.77	2.906	0.20	2.01		2052
24	24.21	0.290	<u>2059.82</u>	<u>2061.89</u>	2.901	0.21	2.07		<u>2067</u>
25	25.22	0.292	2069.79	2071.93	2.896	0.22	2.14		2072
26	26.22	0.293	2079.70	2081.90	2.892	0.22	2.20		2082
27	27.23	0.292	2089.58	2091.84	2.886	0.23	2.26		2100
28	28.23	0.292	2099.45	2101.77	2.883	0.23	2.32		2102
29	29.24	0.290	2109.34	2111.72	2.880	0.24	2.38		2112
30	30.24	0.288	2119.27	2121.70	2.874	0.24	2.43		2122
31	31.25	0.285	2129.26	2131.75	2.870	0.25	2.49		2132
32	32.25	0.281	2139.34	2141.88	2.868	0.25	2.54		2142
33	33.25	0.277	2149.55	2152.14	2.863	0.25	2.59		2152
34	34.25	0.272	2159.90	2162.55	2.862	0.25	2.65		2163
35	35.25	0.266	2170.44	2173.14	2.859	0.25	2.70		2173
36	36.25	0.260	2181.20	2183.95	2.856	0.25	2.75		2184
37	37.25	0.253	2192.21	2195.01	2.856	0.25	2.80		2195
38	38.24	0.245	2203.53	2206.38	2.856	0.24	2.85		2206
39	39.24	0.237	2215.21	2218.11	2.855	0.24	2.90		2218
40	40.23	0.228	2227.30	2230.25	2.859	0.23	2.95		2230
41	41.23	0.218	2239.89	2242.89	2.860	0.23	3.00		2243
42	42.22	0.208	2253.05	2256.10	2.865	0.22	3.05		2256
43	43.21	0.197	2266.90	2270.00	2.871	0.21	3.10		2270
44	44.20	0.185	2281.57	2284.73	2.880	0.20	3.16		2285
45	45.19	0.172	2297.23	2300.44	2.892	0.19	3.21		2300
46	46.18	0.159	2314.11	2317.38	2.909	0.18	3.27		2317
47	47.16	0.145	2332.51	2335.85	2.931	0.16	3.34		2336
48	48.15	0.131	2352.86	2356.27	2.958	0.15	3.41		2356
49	49.13	0.115	2375.76	2379.26	3.003	0.13	3.50		2400
50	50.11	0.099	2402.20	2405.81	3.062	0.11	3.61		2406
51	51.09	0.081	2433.75	2437.51	3.160	0.09	3.76		2438
52	52.07	0.063	2473.43	2477.45	3.337	0.07	4.02		2477
53	53.05	0.043	2527.98	2532.62	3.781	0.05	4.64		2533
54	54.03	0.022	2618.89	2622.24	2.772	0.03	3.35		2622
55	54.71	0.007	2746.62	2709.32	0.877	-0.29	-37.30		2747

[#]: Note that the paper conversion was used for the Final column.

Table 8. Summary of Item-Level Analyses and the 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	6 (52)	0.02 (0.002)	-0.17 (-0.02)	0.29 (0.03)	-0.05 (-0.01)	-1.08 (-0.10)	1.95 (0.22)
Exit Level Math	25 (60)	-1.11 (-0.17)	-0.99 (-0.15)	-1.18 (-0.18)	-1.12 (-0.17)	-1.36 (-0.21)	-0.82 (-0.12)
Exit Level Science	17 (55)	-0.42 (-0.07)	-0.50 (-0.08)	-0.37 (-0.06)	-0.29 (-0.05)	-0.57 (-0.10)	-0.23 (-0.04)
Exit Level Social Studies	6 (55)	-0.20 (-0.03)	0.18 (0.03)	-0.38 (-0.06)	-0.06 (-0.11)	-0.57 (-0.09)	0.68 (0.11)

*: 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 9. Summary of Item-Level Analysis – Exit Level ELA

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
1	0.83	0.85	-0.03	0.02	-1.67		-0.08
2	0.75	0.78	-0.02	0.02	-1.23		-0.06
3	0.74	0.79	-0.05	0.02	-2.52	*	-0.12
4	0.71	0.73	-0.02	0.02	-1.01		-0.05
5	0.75	0.77	-0.02	0.02	-0.78		-0.04
6	0.82	0.81	0.01	0.02	0.55		0.03
7	0.50	0.53	-0.03	0.02	-1.37		-0.06
8	0.57	0.60	-0.03	0.02	-1.38		-0.06
9	0.78	0.79	0.00	0.02	-0.13		-0.01
10	0.61	0.62	-0.01	0.02	-0.45		-0.02
11	0.81	0.85	-0.04	0.02	-2.19	*	-0.11
12	0.40	0.39	0.01	0.02	0.26		0.01
13	0.79	0.83	-0.04	0.02	-2.12	*	-0.10
14	0.84	0.90	-0.06	0.02	-3.62	*	-0.17
15	0.73	0.74	-0.01	0.02	-0.68		-0.03
16	0.68	0.71	-0.03	0.02	-1.52		-0.07
17	0.63	0.66	-0.03	0.02	-1.35		-0.07
18	0.61	0.62	-0.02	0.02	-0.70		-0.03
19	0.59	0.61	-0.02	0.02	-1.01		-0.05
20	0.74	0.75	-0.01	0.02	-0.69		-0.03
21	0.65	0.71	-0.06	0.02	-2.67	*	-0.12
22	0.53	0.54	-0.01	0.02	-0.56		-0.03
23	0.76	0.76	0.00	0.02	0.03		0.00
24	0.60	0.60	0.00	0.02	0.03		0.00
25	0.73	0.75	-0.02	0.02	-1.01		-0.05
26	0.74	0.76	-0.02	0.02	-1.22		-0.06
27	0.63	0.64	-0.01	0.02	-0.45		-0.02
28	0.60	0.60	0.00	0.02	0.13		0.01
29	1.00	0.98	0.02	0.03	0.88		0.04
30	1.10	1.14	-0.04	0.03	-1.47		-0.07
31	0.91	0.91	0.00	0.03	0.01		0.00
32	0.62	0.64	-0.01	0.02	-0.66		-0.03
33	0.85	0.84	0.01	0.02	0.78		0.03
34	0.92	0.92	0.01	0.01	0.75		0.03
35	0.60	0.60	-0.01	0.02	-0.29		-0.01
36	0.67	0.69	-0.02	0.02	-1.09		-0.05
37	0.80	0.76	0.04	0.02	1.90		0.09

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
38	0.73	0.71	0.01	0.02	0.52		0.02
39	0.74	0.75	-0.01	0.02	-0.74		-0.03
40	0.74	0.73	0.02	0.02	0.79		0.04
41	0.74	0.73	0.01	0.02	0.42		0.02
42	0.76	0.74	0.02	0.02	0.85		0.04
43	0.63	0.61	0.01	0.02	0.55		0.02
44	0.60	0.60	0.00	0.02	0.12		0.00
45	0.78	0.77	0.01	0.02	0.41		0.02
46	0.82	0.80	0.02	0.02	1.20		0.06
47	0.81	0.81	0.00	0.02	0.19		0.01
48	0.85	0.85	0.00	0.02	-0.14		-0.01
49	0.69	0.67	0.03	0.02	1.22		0.06
50	0.65	0.68	-0.02	0.02	-1.07		-0.05
51	0.72	0.72	-0.01	0.02	-0.38		-0.02
52	6.86	6.33	0.53	0.12	4.39	*	0.19

Table 10. Summary of Item-Level Analysis – Exit Level Mathematics

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
1	0.81	0.79	0.02	0.01	1.48		0.04
2	0.76	0.76	-0.01	0.01	-0.66		-0.02
3	0.41	0.45	-0.04	0.01	-3.30	*	-0.08
4	0.55	0.63	-0.08	0.01	-6.13	*	-0.17
5	0.46	0.46	0.00	0.01	0.28		0.01
6	0.41	0.43	-0.02	0.01	-1.51		-0.04
7	0.31	0.32	-0.01	0.01	-0.79		-0.02
8	0.55	0.55	-0.01	0.01	-0.40		-0.01
9	0.51	0.53	-0.03	0.01	-2.24	*	-0.05
10	0.45	0.42	0.03	0.01	2.36	*	0.06
11	0.30	0.36	-0.05	0.01	-4.64	*	-0.12
12	0.56	0.55	0.01	0.01	0.72		0.02
13	0.42	0.42	0.00	0.01	-0.16		0.00
14	0.45	0.40	0.05	0.01	3.37	*	0.10
15	0.49	0.52	-0.04	0.01	-2.67	*	-0.08
16	0.22	0.22	0.00	0.01	-0.11		0.00
17	0.25	0.27	-0.01	0.01	-1.31		-0.03
18	0.25	0.29	-0.04	0.01	-3.13	*	-0.09
19	0.25	0.27	-0.02	0.01	-1.97	*	-0.05
20	0.48	0.44	0.04	0.01	3.01	*	0.08
21	0.20	0.25	-0.05	0.01	-4.66	*	-0.11
22	0.25	0.26	-0.01	0.01	-0.60		-0.02
23	0.38	0.46	-0.07	0.01	-5.49	*	-0.15
24	0.31	0.37	-0.06	0.01	-4.24	*	-0.13
25	0.37	0.44	-0.07	0.01	-6.58	*	-0.14
26	0.41	0.41	0.00	0.01	-0.27		-0.01
27	0.13	0.16	-0.03	0.01	-3.16	*	-0.08
28	0.44	0.47	-0.03	0.01	-2.22	*	-0.06
29	0.51	0.50	0.00	0.01	0.18		0.00
30	0.51	0.52	-0.01	0.01	-0.56		-0.01
31	0.31	0.33	-0.02	0.01	-1.47		-0.04
32	0.17	0.17	0.00	0.01	0.38		0.01
33	0.45	0.47	-0.02	0.01	-1.57		-0.04
34	0.50	0.49	0.01	0.01	0.50		0.01
35	0.30	0.33	-0.04	0.01	-2.81	*	-0.08
36	0.40	0.42	-0.01	0.01	-1.03		-0.03
37	0.63	0.60	0.03	0.01	1.89		0.06

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
38	0.34	0.35	-0.01	0.01	-0.79		-0.02
39	0.36	0.37	-0.01	0.01	-0.81		-0.02
40	0.28	0.29	-0.01	0.01	-0.90		-0.02
41	0.38	0.40	-0.02	0.02	-1.56		-0.05
42	0.71	0.71	0.00	0.01	-0.22		-0.01
43	0.31	0.34	-0.03	0.01	-2.26	*	-0.06
44	0.42	0.45	-0.03	0.01	-2.39	*	-0.06
45	0.46	0.44	0.02	0.01	1.34		0.04
46	0.40	0.40	0.00	0.01	-0.16		0.00
47	0.32	0.32	0.00	0.01	0.10		0.00
48	0.42	0.42	0.00	0.01	0.06		0.00
49	0.36	0.39	-0.03	0.01	-2.25	*	-0.06
50	0.42	0.44	-0.01	0.01	-1.13		-0.03
51	0.38	0.39	-0.01	0.01	-0.69		-0.02
52	0.57	0.59	-0.02	0.01	-1.87		-0.05
53	0.42	0.53	-0.11	0.01	-8.09	*	-0.22
54	0.55	0.57	-0.02	0.01	-1.58		-0.04
55	0.45	0.49	-0.04	0.01	-2.96	*	-0.08
56	0.30	0.34	-0.04	0.01	-3.36	*	-0.09
57	0.41	0.42	-0.01	0.01	-0.72		-0.02
58	0.55	0.57	-0.02	0.01	-1.90		-0.05
59	0.57	0.65	-0.08	0.01	-5.77	*	-0.16
60	0.65	0.67	-0.02	0.01	-2.13	*	-0.05

Table 11. Summary of Item-Level Analysis – Exit Level Science

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
1	0.64	0.66	-0.02	0.01	-1.65		-0.04
2	0.40	0.44	-0.04	0.01	-3.09	*	-0.08
3	0.65	0.67	-0.03	0.01	-2.07	*	-0.05
4	0.62	0.62	-0.01	0.01	-0.47		-0.01
5	0.50	0.47	0.03	0.01	2.94	*	0.07
6	0.47	0.50	-0.03	0.01	-2.65	*	-0.06
7	0.54	0.56	-0.02	0.01	-1.82		-0.04
8	0.21	0.28	-0.07	0.01	-5.75	*	-0.16
9	0.12	0.15	-0.02	0.01	-3.16	*	-0.07
10	0.23	0.22	0.01	0.01	1.07		0.03
11	0.61	0.58	0.03	0.01	2.55	*	0.07
12	0.67	0.66	0.02	0.01	1.51		0.04
13	0.67	0.67	0.00	0.01	-0.03		0.00
14	0.60	0.61	-0.01	0.01	-0.65		-0.02
15	0.61	0.65	-0.04	0.01	-3.56	*	-0.09
16	0.54	0.56	-0.02	0.01	-1.97		-0.05
17	0.30	0.28	0.02	0.01	2.03		0.05
18	0.50	0.53	-0.04	0.01	-2.83	*	-0.07
19	0.54	0.53	0.01	0.01	1.03		0.02
20	0.60	0.61	-0.01	0.01	-0.81		-0.02
21	0.55	0.55	0.00	0.01	-0.04		0.00
22	0.30	0.31	0.00	0.01	-0.15		0.00
23	0.41	0.43	-0.02	0.01	-1.48		-0.04
24	0.46	0.42	0.04	0.01	3.07	*	0.08
25	0.60	0.59	0.01	0.01	0.99		0.03
26	0.50	0.51	-0.01	0.02	-0.75		-0.02
27	0.36	0.35	0.01	0.01	0.49		0.01
28	0.38	0.39	-0.01	0.01	-0.67		-0.02
29	0.58	0.59	-0.01	0.01	-1.01		-0.03
30	0.41	0.41	0.00	0.01	0.19		0.01
31	0.41	0.42	-0.01	0.01	-0.59		-0.01
32	0.42	0.45	-0.03	0.01	-2.64	*	-0.07
33	0.47	0.47	0.00	0.01	-0.12		0.00
34	0.27	0.24	0.03	0.01	2.76	*	0.07
35	0.25	0.28	-0.03	0.01	-2.68	*	-0.06
36	0.26	0.30	-0.04	0.01	-3.14	*	-0.08
37	0.43	0.42	0.01	0.01	0.80		0.02

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
38	0.50	0.49	0.01	0.01	1.18		0.03
39	0.66	0.68	-0.02	0.01	-2.35	*	-0.05
40	0.48	0.48	0.00	0.01	-0.40		-0.01
41	0.32	0.33	-0.02	0.01	-1.31		-0.04
42	0.58	0.57	0.01	0.01	0.61		0.01
43	0.43	0.42	0.01	0.01	0.78		0.02
44	0.28	0.28	0.00	0.01	0.07		0.00
45	0.58	0.60	-0.02	0.01	-1.73		-0.04
46	0.33	0.32	0.00	0.01	0.23		0.01
47	0.38	0.41	-0.02	0.01	-1.83		-0.05
48	0.28	0.27	0.00	0.01	0.19		0.00
49	0.68	0.68	0.00	0.01	0.41		0.01
50	0.49	0.56	-0.07	0.01	-5.39	*	-0.14
51	0.71	0.70	0.01	0.01	1.40		0.03
52	0.50	0.50	0.00	0.01	0.04		0.00
53	0.58	0.58	0.00	0.01	0.02		0.00
54	0.37	0.41	-0.04	0.01	-2.96	*	-0.08
55	0.60	0.60	0.00	0.01	-0.16		0.00

Table 12. Summary of Item-Level Analysis – Exit Level Social Studies

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
1	0.59	0.60	-0.01	0.03	-0.33		-0.02
2	0.71	0.65	0.06	0.03	2.40	*	0.13
3	0.51	0.47	0.04	0.03	1.58		0.09
4	0.67	0.72	-0.05	0.02	-1.98		-0.10
5	0.53	0.51	0.01	0.03	0.50		0.03
6	0.73	0.69	0.03	0.03	1.30		0.08
7	0.46	0.54	-0.08	0.03	-2.69	*	-0.16
8	0.59	0.59	0.00	0.03	-0.06		0.00
9	0.65	0.60	0.04	0.03	1.71		0.09
10	0.49	0.45	0.04	0.03	1.60		0.09
11	0.67	0.65	0.01	0.03	0.48		0.03
12	0.59	0.62	-0.03	0.03	-0.97		-0.06
13	0.48	0.48	-0.01	0.03	-0.26		-0.02
14	0.33	0.29	0.04	0.03	1.67		0.09
15	0.44	0.44	0.00	0.03	0.07		0.00
16	0.48	0.47	0.01	0.03	0.28		0.02
17	0.40	0.43	-0.04	0.03	-1.29		-0.08
18	0.57	0.53	0.05	0.03	1.69		0.09
19	0.40	0.40	-0.01	0.03	-0.28		-0.02
20	0.45	0.50	-0.05	0.03	-1.90		-0.11
21	0.46	0.45	0.01	0.03	0.27		0.02
22	0.39	0.36	0.02	0.03	0.72		0.05
23	0.51	0.53	-0.02	0.03	-0.60		-0.03
23	0.58	0.53	0.05	0.03	2.01		0.11
25	0.39	0.40	-0.02	0.03	-0.59		-0.03
26	0.21	0.25	-0.04	0.03	-1.48		-0.09
27	0.57	0.54	0.03	0.03	1.10		0.06
28	0.31	0.34	-0.02	0.03	-0.87		-0.05
29	0.27	0.29	-0.02	0.03	-0.61		-0.04
30	0.47	0.46	0.01	0.03	0.41		0.02
31	0.43	0.45	-0.02	0.03	-0.52		-0.03
32	0.44	0.42	0.02	0.03	0.65		0.04
33	0.44	0.41	0.03	0.03	0.91		0.05
34	0.59	0.54	0.05	0.03	1.87		0.10
35	0.44	0.47	-0.03	0.03	-1.10		-0.06
36	0.40	0.39	0.01	0.02	0.53		0.03
37	0.19	0.18	0.01	0.02	0.51		0.03

ITEM	CBT_PVAL	PAP_PVAL	DIF_PVAL	DIF_STD	Z_DIF	SIG	EFFECT_SIZE
38	0.33	0.33	0.00	0.03	-0.13		-0.01
39	0.47	0.46	0.01	0.03	0.22		0.01
40	0.54	0.54	0.00	0.03	-0.16		-0.01
41	0.28	0.32	-0.04	0.02	-1.62		-0.09
42	0.50	0.47	0.02	0.03	0.83		0.04
43	0.45	0.43	0.03	0.03	0.96		0.05
44	0.50	0.56	-0.06	0.03	-2.55	*	-0.13
45	0.47	0.49	-0.02	0.03	-0.80		-0.05
46	0.54	0.58	-0.04	0.03	-1.76		-0.09
47	0.54	0.59	-0.05	0.03	-1.61		-0.10
48	0.43	0.45	-0.02	0.02	-0.86		-0.04
49	0.49	0.53	-0.04	0.03	-1.56		-0.09
50	0.57	0.55	0.02	0.03	0.56		0.03
51	0.45	0.46	-0.01	0.03	-0.45		-0.03
52	0.41	0.44	-0.03	0.03	-1.17		-0.07
53	0.40	0.48	-0.08	0.03	-2.98	*	-0.16
54	0.70	0.64	0.06	0.03	2.16	*	0.12
55	0.53	0.61	-0.07	0.03	-2.70	*	-0.15