APPENDIX 5

ONLINE PILOT STUDY REPORT

Report on the Grade 8 Online Pilot Study

Summary

This is a summary of a report regarding the Grade 8 Online Pilot conducted in the spring 2004 at the request of the Texas Education Agency, Division of Student Assessment, by Pearson Educational Measurement.

September 1, 2004

Report on the Grade 8 Online Pilot Study
Summary

Introduction

Since the fall of 2002, districts have had the option of administering the Algebra I end-of-course examination online. The success of this program sparked interest in expanding Texas's application of online (computer-administered) testing. The feasibility of expanding the applications of online testing required study. Consequently, in December 2003 the Student Assessment Division of TEA authorized a pilot study of online Texas Assessment of Knowledge and Skills (TAKS) testing for Grade 8.

From March 22, 2004, to April 2, 2004, online pilot versions of the Grade 8 reading, mathematics, and social studies TAKS tests were administered to 22,583 students in 216 campuses representing 172 districts across the state. Each campus administered only one subject-area test online. The online tests were administered in a secure environment and were constructed to be approximately equal in difficulty to the operational TAKS tests administered in the spring of 2004. A post-administration survey was provided to each participating district in order to collect information regarding the online pilot.

The purpose of the pilot was threefold. First, the pilot would provide information regarding the general feasibility of implementing computerized testing on a large scale. Second, the pilot would help determine whether local infrastructure (equipment and personnel) is sufficient to support computerized testing on a large scale. Finally, comparability in performance between paper-based and electronic administrations could be evaluated.

This report provides a brief summary of the results of the Grade 8 Online Pilot.

Post-Test Survey

A post-test survey was distributed to all participating districts after both the online pilot and the spring 2004 operational testing windows had closed. The survey consisted of eight topic sections, each with two to six specific questions to be answered on a 5-point Likert scale and an area for narrative comments or suggestions. The survey also contained a final section of four open-ended questions. Post-test survey data were received from 86 districts.

Quantitative results from the post-test survey are summarized in the following table. In general, participants responded that the pilot was successful. However, only 65 percent of the respondents indicated they had had enough computers for the pilot testing. Only 39 percent of the respondents indicated that the computer rooms were configured for security. Ninety percent of the respondents indicated that having less material to handle was an improvement. Fifty-one percent of the districts responded that computer availability was an issue. Almost 60 percent of the respondents recommended pursuing online testing statewide.

Post-Test Survey Results						
	Percent Responding					
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Not Applicable
Training						
The user manuals were easy to understand.	13	62	4	19	2	0
The telephone (WebEx) training session was useful.	15	27	19	12	5	22
Training was adequate and met your needs.	11	46	13	18	7	6
Customer Support						
The Help Desk was available when needed.	40	31	7	2	1	19
The Help Desk was able to answer your questions.	35	34	9	1	1	20
Any problems were quickly resolved.	33	34	7	6	4	16
The quality of customer service met your expectations.	35	34	9	1	3	17
Ease of Setup						
Downloading the software was easy.	21	56	6	9	1	6
The student registration option you used was adequate.	21	57	3	6	2	10
Preloading student information into the system was easy.	23	45	12	6	3	10
Assigning students to test sessions was easily accomplished.	26	47	5	13	6	5
The mechanism for adding students was user-friendly.	24	51	7	11	1	7
Moving students to different test sessions was easy.	15	43	10	14	1	16
Test sessions were easily scheduled using the eMeasurement site.	21	64	6	5	0	5
Test Administration						
Arranging the testing room was easily accomplished.	29	51	6	12	2	0
Sufficient computers were available for testing.	25	40	7	19	9	0
The computer response time was satisfactory.	32	53	1	8	5	1
There were no technological disruptions in test administration.	24	42	7	12	15	0
Test Security						
Test security was easy to maintain.	30	45	8	9	7	0
Seating arrangements did not allow students to see others' answers.	13	36	12	23	16	0

Post-Test Survey Results (continued)						
	Percent Responding					
	Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Not Applicable**
Test Results						
Reports were easily generated.	44	37	6	8	1	3
Results were easy to understand.	47	46	2	2	0	2
Student Reactions						
Students found online testing engaging.	28	49	15	6	2	0
Students adapted quickly to testing online.	34	58	3	2	2	0
Generally speaking, student reaction was positive.	36	43	13	6	2	0
Test Coordinator / Administrator Reactions						
Administering a test online is easier than administering a paper-and-pencil test.	28	31	17	17	5	1
Preparation for online testing was more manageable than paper testing.	22	34	21	15	7	1
Having less test material inventory to manage was an improvement.	45	45	3	2	2	1
Generally speaking, teacher reaction to online testing was positive.	27	42	15	12	3	1
Computer availability is not an issue on your campus.	17	21	9	30	21	1
You would recommend pursuing online testing for statewide assessment.	32	27	16	15	8	1

Comparability Analyses

Although the idea of online testing has some appeal, the feasibility of such a new test administration modality must be studied. Ongoing research is necessary to examine the comparability of the computer-administered tests with the existing paper-and-pencil formats. According to professional standards of practice, research studies should be conducted to demonstrate that students are not disadvantaged in any way by taking the tests in an electronic form instead of the typical paper-and-pencil format. For example, the American Psychological Association's *Guidelines for Computer-Based Tests and Interpretations* (1986) states "when interpreting scores from the computerized versions of conventional tests, the equivalence of scores from computerized versions should be established and documented before using norms or cut scores obtained from conventional tests" (p. 18). Furthermore, the joint *Standards for Educational and Psychological Testing* (AERA, 1999) also recommends empirical validation of computerized versions of tests in Standard 4.11: "A clear rationale and supporting evidence should be provided for any claim that scores earned on different forms of a test may be used interchangeably" (p. 57).

Comparability studies explore the possibility of effects that result from the use of computers to administer tests. Studies such as the current one are needed in order to demonstrate that students are not disadvantaged in any way by taking the tests in an electronic form rather than the typical paper-and-pencil format. Comparability studies are desirable to ensure that test score interpretations remain valid for their intended purpose across computer and paper-and-pencil formats. Comparability is particularly critical when the testing is high stakes in nature, as in Texas, where student performance is used for the Student Success Initiative, high school graduation, and other accountability purposes.

Online Pilot Sample

Data from a total of 22,584 students representing 172 districts were collected during the online pilot testing window. The following table reports some of the characteristics of study participants.

Grade 8 Mathematics				
	Pilot		Statewide	
	Number	Percent	Number	Percent
Total Number*	7,401	3**	285,712	100
Male	3,692	50	141,092	49
Female	3,708	50	144,363	51
American Indian	27	0	968	0
Asian	223	3	8,558	3
African American	394	5	39,603	14
Hispanic	3,612	49	113,351	40
White	3,144	42	122,751	43

*Some cases lost after matching.
** Total pilot represents 3 percent of entire statewide population.

Grade 8 Reading				
	Pilot		Statewide	
	Number	Percent	Number	Percent
Total Number*	5,482	2**	285,996	100
Male	2,808	51	140,871	49
Female	2,674	49	144,862	51
American Indian	20	0	978	0
Asian	122	2	8,512	3
African American	461	8	39,660	14
Hispanic	2,324	42	112,987	40
White	2,555	47	123,373	43

*Some cases lost after matching.
** Total pilot represents 2 percent of entire statewide population.

Grade 8 Social Studies				
	Pilot		Statewide	
	Number	Percent	Number	Percent
Total Number*	7,821	3**	287,741	100
Male	3,969	51	142,473	50
Female	3,851	49	144,995	50
American Indian	35	0	988	0
Asian	186	2	8,527	3
African American	780	10	40,089	14
Hispanic	3,276	42	113,697	40
White	3,542	45	123,944	43

*Some cases lost after matching.
** Total pilot represents 3 percent of entire statewide population.

Descriptive statistics for the raw score (number of items correct) on both the paper and online tests are provided by subject area in the following table. The mean raw score on the online test was lower (by about four to five points) than the mean raw score on the paper test for all subject areas. Also presented in this table is the internal consistency reliability estimate, coefficient Alpha. Coefficient Alpha was in general high (above 0.85) for all tests and relatively similar between online and paper modes.

Comparisons of Raw Scores for Paper-Based and Online Administrations					
	Mean	**SD**	**Min**	**Max**	**Alpha**
Mathematics					
Paper Raw Score	33	10.2	0	50	0.92
Online Raw Score	28	10.0	3	50	0.90
Reading					
Paper Raw Score	40	6.2	7	48	0.85
Online Raw Score	36	8.4	4	48	0.90
Social Studies					
Paper Raw Score	33	8.9	0	48	0.90
Online Raw Score	29	8.6	3	48	0.87

The following table shows the differences between the online and paper-based administrations in terms of raw scores required to meet both the "Met the Standard" and "Commended" passing standards. Because the online test questions were administered previously in a paper-based format, two different equating methodologies were used to link the online tests to the paper-based scale for comparability purposes. This first method is labeled "2004 Online FT RIDs Fixed" and gives the raw score corresponding to the two standards for the online pilot test form that results from using the common-item field-test data as the equating anchor. The second method, labeled "2004 Online LIVE RIDs Fixed," gives the raw score corresponding to the two standards for the online pilot test form that results from using the 2004 operational paper-based item responses as equating anchors.

This table reiterates that the online assessments are between three and five raw score points more difficult than the live paper-based administration. However, it is assumed that student motivation is higher during live testing and, as such, the results of the current investigation are affected by student motivation.

Comparison of Raw Score Standards Across Conditions and Equating Methods			
	2004 Live Paper	**2004 Online FT RIDs Fixed**	**2004 Online LIVE RIDs Fixed**
Mathematics			
Met the Standard	28	25	22
Commended	45	44	42
Reading			
Met the Standard	31	27	25
Commended	45	43	42
Social Studies			
Met the Standard	22	19	17
Commended	42	41	39