Appendix 16

Report: Alternative Reading Assessments for Older Students with Dyslexia and Word Recognition Difficulties

REPORT

Alternative Reading Assessments for Older Students with Dyslexia and Word Recognition Difficulties

Submitted to the Texas Education Agency

August 31, 2006

Jack M. Fletcher
Department of Psychology
University of Houston

Cathy Caldwell
Sharon Kalinowski
Center for Academic and Reading Skills
University of Texas Health Science Center – Houston

Kimberly O'Malley

Elizabeth Hanna

Pearson Educational Measurement

David J. Francis
Kim Copeland
Texas Institute for Measurement, Evaluation, and Statistics
University of Houston

Alternative Reading Assessments for Older Students with Dyslexia
And Word Recognition Difficulties

Executive Summary

Previous research showed that a package of bundled accommodations (reading aloud proper nouns and comprehension stems to the student, structured extended time) improved the scores of students with dyslexia on the grade 3 Texas Assessment of Knowledge and Skills (TAKS) reading test. In this report, we investigated whether similar accommodations would be effective for secondary level students with dyslexia or other reading disabilities. Students in grades 7 and 11 a) identified with dyslexia or b) served in special education with reading disabilities and c) without reading disabilities were randomly assigned to take the grade appropriate TAKS reading assessment under accommodated and standard administrations. In grade 7, students were assigned to one of three different testing formats: standard administration, a one-day administration involving just the read aloud accommodations, and a two-day administration that provided read aloud accommodations and structured extended time. In grade 11, students were randomly assigned to a standard administration and a one-day administration involving only the read aloud accommodations.

The grade 7 results revealed that students with reading disabilities showed increased scores and a higher pass rate for both accommodated administrations relative to the standard administration; the effect was significantly greater for students who were allowed to take the TAKS over two days. Students without reading disabilities in both accommodated groups performed better than those who received no accommodations. There was no difference in the one-day and two-day administrations, suggesting no benefit of extended time. In examining this finding, unaccommodated students without reading disabilities scored significantly below their scores on the 2006 state-administered TAKS. The difference in scores suggests that the scores of this group on the experimental TAKS were aberrantly low, perhaps reflecting lack of interest in the test. By reading each question aloud on the accommodated version, readers without disabilities were essentially paced on the test, mitigating any impact of potential lack of motivation. If the higher scores of the accommodated readers without disabilities were due to the accommodations, their scores should have been higher than on the state-administered TAKS, which was not the case.

For students in grade 11, there was no improvement in either group of readers by virtue of the accommodations; simply providing accommodations without extended time did not lead to improved scores. However, an interesting finding is that even though most readers with disabilities did not take the state-administered TAKS, virtually all obtained meaningful scores at grades 11 and 7 on our equivalent experimental version, implying that the TAKS is an appropriate test for these students.

The following primary recommendations emerge from the study:

- The bundled accommodations, including structured extended time, should be extended to students in middle school who show clear evidence of decoding problems.
- More students with reading disabilities should be encouraged to take the TAKS, especially in an accommodated administration.
- More research is needed on accommodations for high school students.

Objective

The objective of this study is to evaluate the effects of a bundled accommodation package for older students with disabilities involving word recognition on the Texas Assessment of Knowledge and Skills (TAKS). We propose to administer an alternative, accommodated assessment to students with significant word recognition difficulties in grades 7 and 11 to determine whether a bundle of accommodations effective with third graders is also effective with older students.

Nature of the Accommodations

In 2003, a working group convened by the Texas Education Agency assessed the literature on accommodations and worked with the Dyslexia Statewide Focus Group to develop the accommodations and design the grade 3 study. The group concluded that the critical consideration for accommodating students with dyslexia and word reading difficulties on the TAKS was to minimize the impact of their difficulties with word decoding without changing the capacity of the TAKS to measure reading comprehension. It was determined that three accommodations might potentially reduce the impact of decoding problems on an accommodated administration without invalidating the TAKS as an assessment of reading comprehension: 1) reading aloud the proper nouns; 2) reading aloud the comprehension stems; and 3) structured extended time, i.e., extending the testing session into two blocks.

The rationale behind these accommodations is as follows: 1) Given the decoding difficulties specifically associated with dyslexia, it is important not to require students with dyslexia and related reading disabilities to decode proper nouns that are often idiosyncratic and do not contribute to understanding the meaning of a passage. Therefore, *reading proper nouns* may reduce reliance on decoding skills without invalidating the test. 2) The TAKS reading test is a comprehension assessment that determines how well the student understands the passage that he or she reads. The determination of comprehension is not dependent on the student's ability to read the comprehension questions. Therefore, no "read aloud" accommodation can be provided while the student actually reads the TAKS passages. However, the *comprehension stems can be read* to the student, and the student can provide a written response in the TAKS booklet in the standard manner, thus reducing demands on word decoding skills. 3) When decoding words is difficult, students with reading difficulties commonly experience fatigue and diminished interest during the task. Consequently, *extending the testing session into two blocks* may be beneficial to counteract effects of fatigue and loss of interest due to word decoding difficulties.

Research Design and Questions

To address the impact of accommodations on the TAKS performance of secondary students, we conducted comparisons of students who either had significant word recognition difficulties or were average decoders in grades 7 and 11. The students with reading disabilities were identified with dyslexia and/or served in special education for reading problems, allowing expansion of the original target group to a broader group of students with reading disabilities. Within each group of good and poor readers, pairs

of demographically matched students – a poor reader and an average reader – were randomly assigned to receive an alternative accommodated administration of the TAKS or a standard administration of the TAKS. In grade 7, we separated the effects of the read aloud accommodations (reading proper nouns and stems) from the effects of extended time by provided one-day and two-day accommodations, both incorporating the read aloud accommodations. Thus, we used a two (poor word recognition vs. average word recognition) X three (standard vs. one-day vs. two-day administration) design for grade 7. In grade 11, the structure of the test makes extended time difficult, so we compared a one-day vs. standard administration, or a two (poor word recognition vs. average word recognition) X two (standard vs. one-day administration) design. We also determined whether performance varied with a variety of student characteristics, including the student's overall verbal ability (vocabulary), critical reasoning skills, reading fluency, listening comprehension, and initial level of decoding ability.

Four specific questions can be addressed with this design: 1) Do students with word recognition difficulties perform better on the accommodated administrations of the TAKS than on the standard administration? The comparison of the two groups with word recognition difficulties will provide evidence for answering this question; 2) Is there a differential effect of the administrations of the TAKS for students with and without word recognition difficulties, i.e., is there a group by administration condition interaction? Improved performance by virtue of accommodations only in the group with reading disability would indicate that the accommodations were specific to the disability and did not invalidate the TAKS capacity to assess reading comprehension. 3) Is there a difference in the effects of the reading accommodations versus structured extended time? A comparison of the one-day and two-day accommodations provided in grade 7 will address this question; 4) Does performance on the TAKS, including the alternative assessment for children with word recognition difficulties, vary by virtue of students' vocabulary level, listening comprehension, initial level of decoding, reading fluency skills, and critical reasoning ability? These variables will be tested as covariates in secondary analyses involving the group with reading disabilities.

Methods

Participants

Schools. The students were recruited from five suburban districts in the Houston area (Crosby, Deer Park, Fort Bend, Pasadena, and Goose Creek). In these 5 districts, 27 schools—10 high schools and 17 middle schools- participated, representing 113 classrooms and teachers. We began discussions with district personnel in October 2005, but some time was required to obtain district permissions and identify potential students. Actual screenings began in January 2006. We continued conducting the study through the spring, finishing in May.

Students. The students screened for this study were secondary students in either grade 7 or 11 who were identified and served in public school programs specifically designed for students with dyslexia and/or served in special education programs because of a learning disability in reading. These students were compared with students classified as average readers from the same schools who were similar in demographics.

We screened a total of 1144 students. Of these, 442 were identified as poor readers across both grade levels, and 702 were nominated as average readers. A group of 600 students (grade 7 = 359; grade 11 = 241) represented the final sample, having established eligibility and completed the experimental TAKS. In grade 7, the 168 students with reading disabilities included 46 identified with dyslexia, 99 served in special education, and 23 identified with dyslexia and served in special education; there were 191students without disabilities. In grade 11, the 112 students with reading disabilities included 29 with dyslexia, 77 served in special education and 6 identified with dyslexia and served in special education; there were 129 students without disabilities.

To establish eligibility, the students were initially administered the Letter-Word and Word Attack subtests of the Woodcock-Johnson III Tests of Achievement (WJ; Woodcock, McGrew, & Mather, 2001). The administration of these WJ subtests, which measure reading of real words and pseudowords in isolation, was undertaken to ensure that the student was impaired in reading decoding skills, a necessary criterion because of the accommodations were specifically designed for students with word reading difficulties, identification procedures for special programs vary across schools, and because some students show significant improvement in reading after eligibility is established. Students were accepted into the group with reading disabilities only if they read below the 26th percentile on the Basic Reading Skills cluster scale of the WJ. During the initial screening, we also administered the Kaufman Brief Intelligence Test (K-BIT) Verbal Knowledge and Matrices tests, as well as the Test of Word Reading Efficiency (TOWRE) to all students screened as potential poor reader participants in the study.

To avoid excluding poor readers who were accurate but slow, we cross-checked students identified as poor readers who scored below the 35th percentile in the Woodcock Johnson, with the TOWRE to test their fluency. If their TOWRE score was less than the 26th percentile, the students were included in the poor reader group. This involved 9 students in grade 7. Whereas 2 students in grade 11 qualified in this manner, one of those students refused to take the test, so only one was actually involved in the TAKS testing. We excluded these 10 students from the analysis as there were so few. Some good readers obtained TOWRE fluency scores below the 26th percentile; a few poor readers obtained TOWRE scores above the 25th percentile. Excluding these students from the analysis (N = 64) did not alter the pattern of results.

Students in the comparison group were screened by the same WJ instruments. They were first nominated by their teachers or counselors as being average readers; they were also selected to match the demographic characteristics of the poor readers from the same school. In total, we screened 424 seventh grade students and 277 eleventh grade nominated students. The “qualified” students had WJ scores that were above the 34th percentile and below the 80th percentile, leaving 454 students. Based on closeness in demographic characteristics, pairs – one poor reader and one average reader- of students were scheduled to take randomly assigned TAKS tests. However, when the actual TAKS administration day arrived, if one of the selected average readers was unavailable, another qualified average reader was chosen based merely on convenience and availability, as it was not feasible to reschedule TAKS testing so late in the school year.

Parents were notified of the study and given the opportunity to refuse participation. In addition, students sometimes did not want to participate. In general, the reasons for refusal were simply that the student was overwhelmed at the amount of work/study he or she already had to do, or a sentiment that there was" nothing in it for me." Thirty-one students did not participate due to parental restrictions or student refusals, and 23 students were unable to complete the actual test due to illness, suspension or other factors. In general, we did not return to schools for the purpose of testing absentees, but there were times when a student was reassigned to the next day's administration.

Testing Procedures

Test Descriptions

Texas Assessment of Knowledge and Skills *(Experimental Version).* The study was conducted with forms of the grade 7 and 11 TAKS developed in 2005. Each version was designed so that the format was the same and the test was equivalent in difficulty with the 2006 TAKS. No modifications were made to the TAKS booklets; only instructions for test examiners were modified. For the grade 7 two-day accommodated administrations, the instructions were modified by the Center for Academic and Reading Skills (CARS) and the Texas Education Agency (TEA) so that the TAKS could be administered in two sessions. In the first session, the student completed the first three stories in the test; in the second session, the last two stories were completed. In grade 11, only a one-day accommodated administration was possible because it was difficult to divide the test into two days as some questions required reading of multiple stories. We administered, but did not score, the open-ended writing items as these were not accommodated. The booklets and scoring sheets were provided and scored by Pearson Educational Measurement. A data file was provided with appropriate student identifiers. These data were merged with demographic data and other performance data collected by CARS and Texas Institute of Measurement, Evaluation, and Statistics (TIMES).

Woodcock-Johnson PsychoEducational Test Battery-III Basic Reading *(WJIII; Woodcock et al., 2001).* The WJIII Basic Reading composite is composed of the Letter Word Identification subtest and Word Attack subtest. Letter Word Identification assesses the ability to read real words; Word Attack examines the ability of children who read phonetically correct nonsense words. We used the composite because it is more reliable than individual subtests. It has reliability (split-half) of .91-.97 in the age range of interest.

Test of Word Reading Efficiency *(TOWRE; Torgesen et al., 2001).* The TOWRE is an individually administered test of speeded reading of single words that is commonly used as an outcome in reading intervention studies. The participant is given a list of 104 words and asked to read them as fast as possible. The number of words read correctly within 45 seconds is recorded. Separate subtests are available for word reading efficiency and phonemic decoding efficiency, reflecting speeded reading real words and pseudowords. We gave one form of each subtest. Test-retest reliabilities range from .83-.92 for each subtest in the age range of interest.

Kaufman Brief Intelligence Test-2 *(K-BIT; Kaufman, 1999).* The K-BIT consists of 3 subtests designed to provide a short form assessment of intelligence. We used two subtests to assess vocabulary

(Verbal Reasoning) and reasoning (Matrices). Verbal Reasoning is a receptive vocabulary task, requiring the child to select pictures based on a word, and also asks for responses that reflect general world knowledge. Matrices is a nonverbal problem solving test in which the student selects one response from several possibilities that completes a relation among stimuli. Vocabulary and reasoning skills are widely accepted as determinants of reading comprehension. The test-retest reliability of these two subtests exceed .90 in the age range of interest, with internal consistency above .87.

Group Reading Assessment and Diagnostic Evaluation *(GRADE; Williams, 2002).* The GRADE is a group-based, norm-referenced untimed test that assesses a variety of dimensions of reading ability. We used the Listening Comprehension subtest from the GRADE. For Listening Comprehension, the examiner reads aloud one or two sentences and the student marks one of four pictures that conveys the meaning of what the teacher said. Although it does not require reading, it does require understanding of vocabulary and non-literal language, along with the ability to make inferences. Recent reviews of the GRADE indicate that the technical qualities of the GRADE are strong (Buros, 2003), with excellent reliability and validity and a representative test development and normative process.

Examiners

We utilized trained examiners to ensure consistency and reliability of all administrations. They were trained extensively and observed. They worked in teams to screen students and then to administer the TAKS tests. All examiners signed oaths of test security and confidentiality as mandated by the TEA. There were generally 2-3 examiners in most administrations, although in some cases there were more. The number of examiners was dictated by the number of students in each examination. To promote consistency in test administrations, we did not test groups of more than 20 total students.

Test Procedures

Students were told that they were involved in a study to evaluate different ways to make the TAKS a better test instrument, but that their results would not affect them personally. We asked for their cooperation in giving their best efforts, emphasizing that this would eventually help other students. In addition to a keychain, all students who completed the TAKS were given a personalized certificate of appreciation thanking them for their community service. Whereas these certificates were sent later, the students knew they were being awarded and the certificates were valuable to our getting students' active participation.

To help ensure that testing standards were consistent, all students within a school were tested in small groups in a room provided by the school. These groups ranged in size from 4 students to 20 students per administration. They consistently contained students from the special education, dyslexia and average reader groups. These numbers determined the sizes of each test administration. The accommodated versions were more challenging to administer to the larger groups. We attempted to neutralize any difficulties by utilizing additional examiners. The highly scripted instructions reduced any potential examiner bias.

The administration instructions were highly scripted, read straight from the manual and given with minimal deviation. For the standard administration, the examiners followed the manual, reading directions and monitoring the students, but not otherwise interacting with the group. The other examiners served as other “eyes” or helpers with any exceptions or needs that arose. For the accommodated administrations, the examiners followed the scripted guidelines developed for oral administration of the TAKS. Students also had proper nouns booklets to read and follow along with the examiner. Students in both groups responded in the standard manner by working in their TAKS booklets, and bubbling in responses in their TAKS answer document. In accommodated administrations, the extra examiners were useful in determining pace and readiness of students to proceed to the next questions. Given the group make-up and dynamics, this proved extremely helpful. In addition, at the end of each TAKS administration, when there were additional assessments given, the “extra” examiners were often able to leave the TAKS room, to do the individual assessments. This enabled us to expedite the entire process.

In the grade 7 study, where schools had smaller groups of students, all students who were assigned either of the accommodated versions of the TAKS test began the day together. First, they were given the Listening Comprehension portion of the GRADE test. This was followed by the TAKS administration. After the first three stories and associated questions were finished, those students who were assigned to the two-day accommodated version were dismissed, and those assigned to the one-day accommodated version took a short snack break, and then finished the final two stories and questions. The two-day accommodated group returned the next day to finish the exam. However, in four schools where a test administration would have exceeded 20 students, we split the accommodated versions into two completely separate administrations. Average readers were given the K-BIT and TOWRE assessments upon completion of the TAKS.

In the grade 11 study, all students were first given the GRADE Listening Comprehension test, followed by the TAKS test. Average readers were also given the K-BIT and TOWRE assessments after the TAKS administration.

We allowed students to return to their classrooms when they were finished taking the tests. This was a departure from the routine TAKS administration but was necessitated by time constraints of the schools. We did notice that some students did try to rush through their tests. This appeared to happen more with eleventh graders than with seventh graders.

Statistical Analysis

From a statistical viewpoint, the *first question* involves the magnitude of the effect of accommodations for students with word recognition problems. Framed in this way, this question presumes that the effect of accommodations is different for dyslexic and non-dyslexic students, i.e., that group and assessment type will interact. Thus, the *second question* involves the group by accommodation interaction. The *third question* entails further evaluation of differences among administration conditions in grade 7, which was tested within groups. *The fourth question* is essentially the effect of the covariates (word reading ability, fluency, vocabulary, listening comprehension, critical

reasoning) on group and accommodation status. These covariates were tested in terms of an interaction with group.

TIMES used SAS PROC GLM to test for interactions of group and condition through a series of ANOVAs at grades 7 and 11. Then an analysis of covariance model was fit within the group of readers with disabilities that looked at the effect of accommodation status, each covariate, and their interaction on TAKS performance.

The GLM analysis examined performance on the TAKS as a continuous scaled score. However, the TAKS also has established scale score cut-points that define state-mandated proficiency levels. Thus, on the basis of each student's TAKS continuous score, it was possible to determine whether the student had met the state's passing standard. After converting each student's continuous score on the TAKS to Pass/Fail on the basis of the established cut-point, we conducted a chi square analysis within each reader group to evaluate the effects of accommodations.

Results

Grade 7

Demographic characteristics. The groups did not differ in sociodemographic factors. Table 1 shows that the grade 7 students with disability (M = 162.2; SD = 6.5) were about 5 months older than the students without disability (M = 157.0; SD = 5.3), $F(1,353) = 70.6$, $p < .0001$. This likely reflects the effects of retention in the older group. More importantly, there was no interaction of group and condition, $F(2,353) < 1$, so comparisons within groups are not affected by the age difference. A series of Group (2) by Accommodation Condition chi-square comparisons within each group were not significant for gender, ethnicity, or free lunch status. Thus, the randomization procedures worked well and the within group comparisons are not biased by differences in initial level of reading ability and sociodemographic factors.

***Comparison of groups and conditions*.** Table 2 provides means, standard deviations, and ranges by group and administration condition for the grade 7 TAKS reading test that are not adjusted for the covariates, along with the covariates and the state-administered TAKS. As expected, the groups with reading problems had word reading scores that were at the 10-11th percentile and well below the comparison groups, who were slightly above average. Although students were not selected on the basis of their development of fluency, vocabulary, and critical reasoning, the scores are comparable within groups with and without reading disabilities. Performance on measures of reading fluency, vocabulary, critical reasoning, and listening comprehension was significantly lower for poor readers than for average readers ($p < .0001$). Average readers scored consistently in the average range on these covariate measures.

The GLM analysis showed that the Group (Poor Reader, Good Reader) X Accommodation Condition (Standard, One- Day, Two- Day) interaction was not statistically significant, $F(2, 353) = 1.51$, $p = .22$. The main effects of Group, $F(1, 353) = 294.7$, $p = .0001$, and Accommodation Condition, $F(2, 353) = 13.33$, $p = .0001$, were significant. As expected, collapsing across Conditions, performance was higher

in good readers than poor readers; collapsing across group, the three accommodated conditions were significantly different from one another. Table 2 shows that performance on the two-day accommodation was higher than performance on the standard administration, with the one-day performance in between.

Since we had hypothesized that the Group X Accommodation Condition interaction would be significant, we examined the effects of the different accommodation conditions within each group in a series of planned two-group contrasts, adjusting the critical value of alpha for the number of comparisons (.05/3 = .017). Different patterns emerged and are shown in Figure 1. Within the group with reading disabilities, there was a stepwise effect showing that students in the two-day administration obtained higher scores than students in the one-day (p = .013), and standard administrations (p = .0001). Performance was also higher in the one-day than in the standard administration, but this effect was weaker (p = .08). In contrast, within the group without reading disabilities, performance did not differ between the one-day and two-day accommodations (p = .85), but both the one- day, (p = .008), and two-day administrations, (p = .005) were higher than the standard administration.

These results show the expected effects of the accommodations within the group with reading disabilities. They also show that structured extended time adds to the effects of the two read aloud accommodations in the group with reading disabilities. In contrast, the effects for the students without disabilities do not show differential impact of extended time and the reading accommodations. In trying to understand this finding, we compared the state-administered TAKS scores with the scores from the experimental version. Table 3 shows that the unaccommodated students without disabilities obtained scores on the experimental test that were markedly below their scores on the state-administered test. In contrast, the scores of the two accommodated groups on the experimental test are very similar to their scores on the state-administered TAKS. Similarly, Figure 2, which shows the distribution of performance on the experimental and state TAKS, demonstrates that the unaccommodated good readers show better performance on the state-administered test and that performance is much more variable, with some students having extremely low performance. If the accommodations resulted in better performance on the experimental TAKS, we would expect higher scores than on the state TAKS, which is not the case. Apparently other factors, such as lack of interest in the test and the absence of pacing introduced by the accommodations, led some unaccommodated good readers to exert inadequate effort.

Effects of the covariates. Table 3 also provides means and standard deviations for the 5 covariates. Each covariate was analyzed within the group with reading disabilities. For this analysis, the major question is whether a covariate interacted with the accommodations. Such an interaction would indicate that the effect of the accommodations would vary depending on the level of ability on the covariate. For example, children with reading disabilities and higher decoding scores may be more likely to benefit from accommodations than those with lower decoding scores. In fact, for the WJIII Basic Reading score, the measure of decoding, the interaction was not significant, $F(2,162) < 1$. This pattern was apparent for each of the covariates: TOWRE, $F(2,162) < 1$; K-BIT Verbal Knowledge, $F(2,162) < 1$; K-BIT Matrices, $F(2,162) < 1$; and GRADE Listening Comprehension, $F(2,162) < 1$. Thus, effects of the accommodations did not vary by the level of decoding ability, reading fluency, vocabulary, critical reasoning, or listening comprehension. This does not mean that the covariates did not predict TAKS

performance. In each analysis, the effect of the covariate was significant, indicating that students with higher levels of performance obtained higher scores on the TAKS and are therefore more likely to pass the TAKS. Finally, adding the covariates to the models did not eliminate the overall pattern of results seen in the unadjusted models.

Pass rate analysis. The GLM analysis demonstrates that performance on the grade 7 TAKS was significantly higher for children with reading disabilities who took the TAKS under this "bundled" accommodations condition, an average improvement of about 100 scale score points in the two-day accommodation condition versus the standard administration (Table 2). However, this analysis did not address whether students with reading disabilities were more likely to pass the TAKS if the test was taken under accommodated conditions. It is possible that average TAKS scale scores could be affected with no effects on the pass rates. Thus, to determine if accommodations differentially affected pass rates for children with and without reading disabilities, we conducted a chi square analysis for differences in pass rates by condition. For the students with reading disabilities, 3 X 2 chi-square was significant, x^2 (2, N = 168) = 8.03. p = .018. The pass rates were 7% for standard administration, 17% for the one-day administration, and 27% for the two-day administration. For students without reading disabilities, the chi square was also significant, X (2, N = 191) = 7.07, p = .029. The pass rate was only 68% for the standard administration, 84% for the one-day accommodation, and 85% for the two-day accommodation. Again, the difference across conditions for the good readers may reflect the unusually low performance on the experimental version of the TAKS, especially given that over 80% of these unaccommodated readers without disabilities passed the state-administered TAKS.

Grade 11

Demographic characteristics. The groups did not differ in age and demographic factors. As with the grade 7 study, Table 4 shows that the grade 11 students with disabilities (M = 204.2; SD = 4.7) were about 6 months older than the students without disabilities (M = 210.2; SD = 8.6), F (1,237) = 45.2, p < .0001. This likely reflects the effects of retention in the older group. As in grade 7, there was no interactions of group and condition, F (1,237) < 1, so comparisons within groups are not affected by the age difference. A series of Group (2) by Accommodation Condition (2) chi-square comparisons within each group were not significant for gender, ethnicity, or free lunch status (p < .10). Thus, the randomization procedures worked well and the within group comparisons are not biased by differences in initial level of reading ability and sociodemographic factors.

Comparison of groups and conditions. Table 5 provides means, standard deviations, and ranges by group and administration condition for the grade 11 experimental TAKS reading test that are not adjusted for the covariates, along with the covariates. Only raw scores are available as the writing items were not scored since performance on these unaccommodated items could obscure the effects of the reading accommodations. As expected, the groups with reading problems had word reading scores that were at the 6-7th percentile, somewhat lower than in grade 7 and well below the grade 11 average

reader groups, who were slightly above average. Although students were not selected on the basis of their development of fluency, vocabulary, and critical reasoning, the scores are comparable within groups with and without reading disabilities. The group with reading disabilities shows significant difficulties with reading fluency and lower performance on measures of vocabulary, critical reasoning, and listening comprehension compared to the average readers ($p < .0001$), who scored consistently in the average range on the covariates.

The GLM analysis showed that the Group (Poor Reader, Good Reader) X Accommodation Condition (Standard, One-Day, Two-Day) interaction was not statistically significant, $F(1,237) < 1$. The main effect of Group, $F(1,237) = 175.2$, $p = .0001$ was significant, but not Accommodation Condition, $F(1,237) < 1$. As expected, collapsing across Conditions, performance was higher in good readers than poor readers; collapsing across group, the two accommodated conditions were not significantly different from one another. Table 5 shows mean scores on the TAKS that are virtually identical across accommodations within each group. The read aloud accommodations had little impact on the TAKS performance of grade 11 students. Given the absence of condition effects, we did not further pursue these findings.

Effects of the covariates. Table 5 also provides means and standard deviations for the 5 covariates. Each covariate was analyzed within the group with reading disabilities. No evidence emerged across the 5 covariates for an interaction with the accommodations: WJIII Basic Reading score, $F(1,111) < 1$; TOWRE, $F(1,111) < 1$; K-BIT Verbal Knowledge, $F(1,111) < 1$; K-BIT Matrices, $F(1,111) < 1$; and GRADE Listening Comprehension, $F(1,111) < 1$. As in grade 7, effects of the accommodations did not vary by the level of decoding ability, reading fluency, vocabulary, critical reasoning, or listening comprehension. In each analysis, the affect of the covariate was significant; indicating that students with higher levels of performance obtained higher scores on the TAKS. Also, adding the covariates to the models did not change the overall pattern of results seen in the unadjusted models, i.e., no condition effects.

Pass rate analysis. We did not attempt a pass rate analysis. A great deal of effort would be required to determine a method for creating scale scores just from the reading items of the experimental TAKS and to score the constructed written responses. Since there is no basis for differences in pass rates given the lack of evidence for interactions or main effects involving condition, this question was not pursued.

Discussion

These results indicate no benefit of the one-day read aloud accommodations of the grade 11 students. In grade 7, the interaction indicating that only students with reading disabilities benefited from the accommodations was not obtained. However, it was apparent that the accommodations were effective for grade 7 students with reading disabilities. It also appeared that the interaction failed to emerge because of aberrantly low performance by the students who were average readers who took the TAKS under the standard administration condition. We believe that the interaction is result of this extremely low performance, which reflects a pass rate of only 68% and a much more variable pattern of performance (Figure 2). Many factors could have created this unusually low performance, including lack of interest in taking the TAKS again and other factors that reflected less than adequate effort. It would have been harder for students without disabilities to exert poor effort in the accommodated conditions because of the natural pacing that was built into the experience due to the read aloud accommodations. There was a clear improvement in the pass rate; 27% for students with disabilities who received both the read aloud accommodations and the structured extended time achieved a passing score. In grade 7, the extended time accommodation provided a clear value-added impact relevant to the read aloud accommodations, which were less robust compared to the standard administration. The extended time component is likely effective not just because it gives more time to complete the TAKS, which is possible in the current administration requirements. It may be the structuring of the experience through the pacing and the reduction of fatigue that serves as the key to the benefit of this accommodation. Many children with reading disabilities have organizational and attentional difficulties. Reading is a fatiguing experience for them. By structuring the experience, construct irrelevant effects of their disabilities on reading performance are better controlled.

The absence of benefits for the grade 11 students was clearly apparent. The research should be completed on other accommodations of the high school TAKS. In particular, methods for providing structured extended time should be explored. The high school exit exam is a demanding test and the fatigue experienced by both good and poor readers was readily apparent (see Appendix 1). It would also be helpful to explore ways of modifying the tests and separating out the requirements for written instruction since these items may confound the assessment of reading comprehension. Although the rationale for including constructed written responses is certainly understandable, it may be helpful to separately score these items in an effort to account for variability in the performance of high risk subgroups on the TAKS.

This study not only extended the evaluation of the accommodations to older students, it also included students in special education. An interesting finding from the experimental version of the TAKS that was used by the study is the lack of evidence for a floor effect indicating that the TAKS was too difficult for them. All of the students who took the experimental version of the TAKS, which was equated to the 2006 TAKS, obtained meaningful scores that were clearly different from zero. Concerns that the standard TAKS is not an appropriate test for students with reading disabilities were not supported by this pattern. At the same time, those students who took the state-administered TAKS often obtained

extremely low scores. The basis for this discrepancy is not apparent since the tests were equated in level of difficulty. Many students with disabilities did not have the opportunity take the state- administered TAKS (see Table 5), so more encouragement to districts to permit students with disabilities to participate in either the standard or accommodated administration of the TAKS reading test is encouraged.

Recommendations

- The bundled accommodations, including structured extended time, should be extended to middle school students with reading disabilities.
- More students with reading disabilities should be encouraged to take the TAKS, especially in an accommodated administration.
- More research is needed on accommodations for high school students.

Appendix 1: Examiner Observations

During this study, there were often suggestions made as to the test administration itself. At the end of the field testing, we asked the examiners to share any thoughts or suggestions from these observations. These suggestions are often anecdotal and meant merely to make the administration better in the event that any type of accommodation becomes implemented.

The timing of the administration, while unavoidable, was not ideal because of all the other testing that occurs in the spring, including benchmark tests and standardized tests such as TAKS, SDAA, AP, and SAT. In addition, there were also final exams and routine classroom tests. For the middle schoolers, we ran into conflicts with extra-curricular events, such as band competitions or performances; for the high school students, there were also end-of-year activities, such as prom. Within the schools, there was concern about the amount of testing being done during the spring and whether this would have any effect on either student participation or study results. We tried to make the screening and TAKS administration as painless as possible for all concerned. Many school personnel felt that the TAKS was inappropriate for students who had only taken the SDAA prior to this TAKS administration but responded when we emphasized that this administration would not count and that it was important to learn more about how students with disabilities would perform on the TAKS. Despite these logistic difficulties, the final sample sizes were adequate, reflecting significant effort on the part of the participating schools.

In terms of cooperation, we did find some students to be “burned out” on testing in general. There were instances of students sleeping through the administrations or being otherwise uncooperative. Some of this may be attributed to the diversity of students taking these administrations. Some of students with reading disabilities had never taken a TAKS test before, while others had just completed one. For the most part, however, students did appear to try to give their best efforts. We did note that some students became impatient waiting for the next stem to be read and tended to work ahead. Whenever possible, examiners discouraged this practice. In grade 11, fatigue was an obvious factor for some in both the accommodated version and the standard administration. Later in the day, they often seemed to run out of steam. Structured extended time might have benefited these students.

The oral instructions were irritating to many of the students. The language was sometimes repetitive and abstract so that some students seemed to lose attention while listening. In particular, the specific instructions about the accommodations to the test were embedded into such long paragraphs that the students could miss the instruction entirely. In revising the instructions, we would urge the writer to read them aloud to identify when the language felt stiff and awkward and to actually pilot the instructions with the appropriate age group. For example, referring to the web site as a “viewing and representing piece” was confusing to some eleventh graders. In addition, it would be much clearer to refer to the written pieces as stories or essays. Instructions should be clearer and more to the point. Also, the eleventh grade instructions did not mention the permitted use of dictionaries.

In this administration the eleventh grade proper noun lists could be pruned of simple words such as Tim or Cal. We had students laugh or roll their eyes as short words like these were read. On the other hand, the seventh grade students needed the more difficult words read to them.

Timing and pacing presented distinct challenges. Pacing of the accommodated version to a group of students with mixed abilities was difficult. This should be less of an issue if the accommodations are provided only to students with reading disabilities. Despite instructions to the contrary, students would often work ahead, making it difficult to know when everyone was finished with the question at hand. Also, sometimes one or two slower students could take up to an hour reading a passage or 20 minutes completing a question. This could cause frustration with other students and led to a loss of focus, including sleeping, doodling, staring into space, versus activity related to the reading passage. At the same time, it also seemed that in the accommodated version, students sometimes answered the question with less reflection than when they were on their own. Our gut feeling was that some felt “rushed” by working at a group pace. This pace was not set by the examiner; rather it was set by the students who were very aware of fellow test-takers.

References

Buros Institute (2003). *Group Reading Assessment and Diagnostic Evaluation*. Omaha, NE: Author.

Fletcher, J.M., Francis, D.J., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., & Vaughn, S. (2006). Effects of accommodations on high stakes testing for students with reading disabilities. *Exceptional Children.*

Torgeson, J.K., Wagner, R., & Raschote, C. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Williams, K.T. (2002). *Group Reading Assessment and Diagnostic Evaluation*. Circle Pines, MN: American Guidance Services.

Table 1. Age, gender, ethnicity, and free lunch status by group and condition for grade 7

	GROUP/CONDITION					
	Poor Reader			Average Reader		
	Standard	One-Day	Two-Day	Standard	One-Day	Two-Day
Number of Students	56	53	59	66	61	64
Age (months) M	161.2	163.1	162.5	156.5	156.8	157.7
SD	6.6	6.2	6.7	5.5	4.9	5.4
Gender Female (%)	32	34	25	35	34	25
Ethnicity (%)						
White	50	42	32	42	33	27
African-American	12	9	7	8	8	5
Hispanic	38	49	61	50	59	69
Asian/Other	0	0	0	0	0	0
Free Lunch Status (%)	57	64	69	61	61	59

Table 2. Means and standard deviations for the WJIII Basic Reading, TOWRE, K-BIT Verbal Knowledge and Matrices, GRADE Listening Comprehension, and grade 7 Experimental TAKS Reading Scale score.

	GROUP/CONDITION					
	Poor Reader			Average Reader		
	Standard	One-Day	Two-Day	Standard	One-Day	Two-Day
N	56	53	59	66	61	64
Basic Reading Skills Cluster[a]						
M	9.5	11.2	10.8	54.7	55.9	56.8
SD	7.8	7.1	7.3	13.5	13.6	13.9
Range	1-25	1-25	1-25	36-78	35-79	35-79
TOWRE[b]						
M	78.6	76.2	77.5	101.0	103.4	105.6
SD	10.2	8.9	11.1	9.7	10.4	11.9
Range	51-99	57-98	52-111	81-120	84-126	85-130
K-BIT[b] Verbal Knowledge						
M	89.5	86.7	86.4	100.2	102.3	97.9
SD	10.7	10.1	11.0	11.7	10.3	9.9
Range	65-105	65-110	65-115	70-130	75-130	80-135
K-BIT Matrices[b]						
M	91.0	88.1	91.5	98.2	99.5	97.0
SD	14.3	15.5	14.9	12.7	11.1	12.9
Range	53-119	53-116	55-119	69-128	72-128	56-124
GRADE Listening Comprehension[c]						
M	10.1	10.0	10.0	11.8	12.0	11.6
SD	2.4	2.2	2.2	2.0	2.0	2.0
Range	6-15	4-15	4-15	8-15	7-16	6-16
TAKS: Experimental Scale Score						
M	1895.0	1936.5	1994.7	2137.2	2197.0	2196.0
SD	118.3	162.3	146.5	166.4	11.7	92.0
Range	1684-2192	1626-2332	1750-2400	1730-2541	1934-2439	2009-2439

Notes. WJ-III = Woodcock-Johnson Achievement Tests-III; TOWRE = Test of Word Reading Efficiency; K-BIT = Kaufman Brief Intelligence Test-2; GRADE = Group Reading Achievement Diagnostic Evaluation; TAKS = Texas Assessment of Knowledge and Skills reading test. [a]Percentile; [b]Mean (SD) = 100 (15); [c]Raw Score.

Table 3. Means and standard deviations (raw scores) for the grade 7 State-Administered TAKS and Experimental TAKS by Group and Condition.

	GROUP/CONDITION					
	Poor Reader			Average Reader		
	Standard	One-Day	Two-Day	Standard	One-Day	Two-Day
N	18	18	31	63	60	60
TAKS: State Administered						
M	24.4	25.9	23.7	39.6	40.6	40.4
SD	16.7	17.4	15.9	5.5	4.7	4.8
Range	0-43	0-44	0-43	18-47	25-48	24-48
N	18	18	31	63	60	60
TAKS: Experimental						
M	24.6	32.1	30.2	34.4	38.1	38.5
SD	9.4	7.3	8.9	9.3	5.3	4.7
Range	11-39	19-44	14-45	9-47	22-46	27-46

Notes. TAKS = Texas Assessment of Knowledge and Skills Reading Test.

Table 4. Age, gender, ethnicity, and free lunch status by group and condition for grade 11.

	GROUP/CONDITION			
	Poor Reader		Average Reader	
	Standard	One-Day	Standard	One-Day
N	59	53	60	69
Age (months) M	210.7	209.6	204.5	204.0
SD	7.8	9.4	5.2	4.2
Gender Female(%)	36	40	37	32
Ethnicity				
White	36	49	37	44
African-American	19	13	12	9
Hispanic	42	38	47	45
Asian/Other	3	0	5	3
Free Lunch Status (%)	42	40	37	41

Table 5. Means and standard deviations for the WJIII Basic Reading, TOWRE, K-BIT Verbal Knowledge and Matrices, GRADE Listening Comprehension, Grade 11 TAKS Reading scale score, and grade 11 State-Administered TAKS by Group and Condition.

	GROUP/CONDITION			
	Poor Reader		Average Reader	
	Standard	One-Day	Standard	One-Day
N	59	53	60	69
Basic Reading Skills Cluster[a]				
M	6.4	7.6	51.9	52.3
SD	7.0	8.0	11.8	12.7
Range	1-25	1-25	35-79	35-79
TOWRE[b]				
M	72.1	69.0	96.4	99.1
SD	12.6	10.5	10.1	9.2
Range	54-96	45-99	73-116	74-118
K-BIT[b] Verbal Knowledge				
M	84.3	86.5	97.2	98.5
SD	10.7	13.1	9.9	10.3
Range	55-105	55-120	80-130	75-120
K-BIT Matrices[b]				
M	90.1	89.8	96.2	97.3
SD	14.3	15.6	9.8	10.0
Range	55-132	55-125	70-120	72-120
GRADE Listening Comprehension[a]				
M	11.6	10.9	13.6	13.3
SD	2.9	3.4	2.0	2.2
Range	3-16	1-16	9-17	7-17
TAKS: Experimental[c]				
M	17.6	17.5	27.1	27.8
SD	8.3	6.6	4.1	3.5
Range	4-32	6-30	9-32	15-33

Notes. WJ-III = Woodcock-Johnson Achievement Tests-III; TOWRE = Test of Word Reading Efficiency; K-BIT = Kaufman Brief Intelligence Test-2; GRADE = Group Reading Assessment and Diagnostic Evaluation; TAKS = Texas Assessment of Knowledge and Skills Reading Test. [a]Percentile; [b]Mean (SD) = 100 (15); [c]Raw Score.

Figure 1. Relation of performance on the Woodcock-Johnson III Basic Reading Skills cluster and Texas Assessment of Knowledge and Skills (TAKS) scale score for groups with poor decoding (PR) and average decoding (AD) by condition (accommodated (ACC) vs. standard administrations

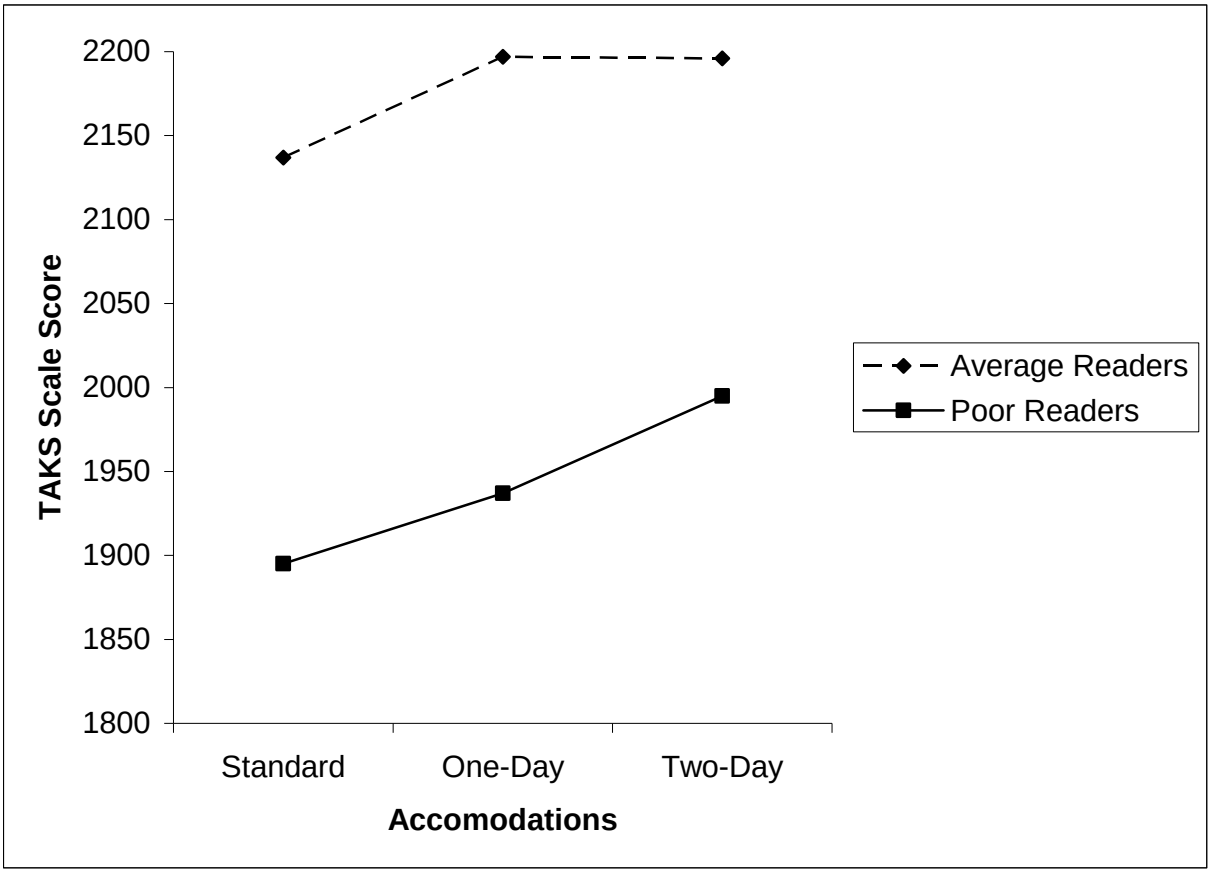
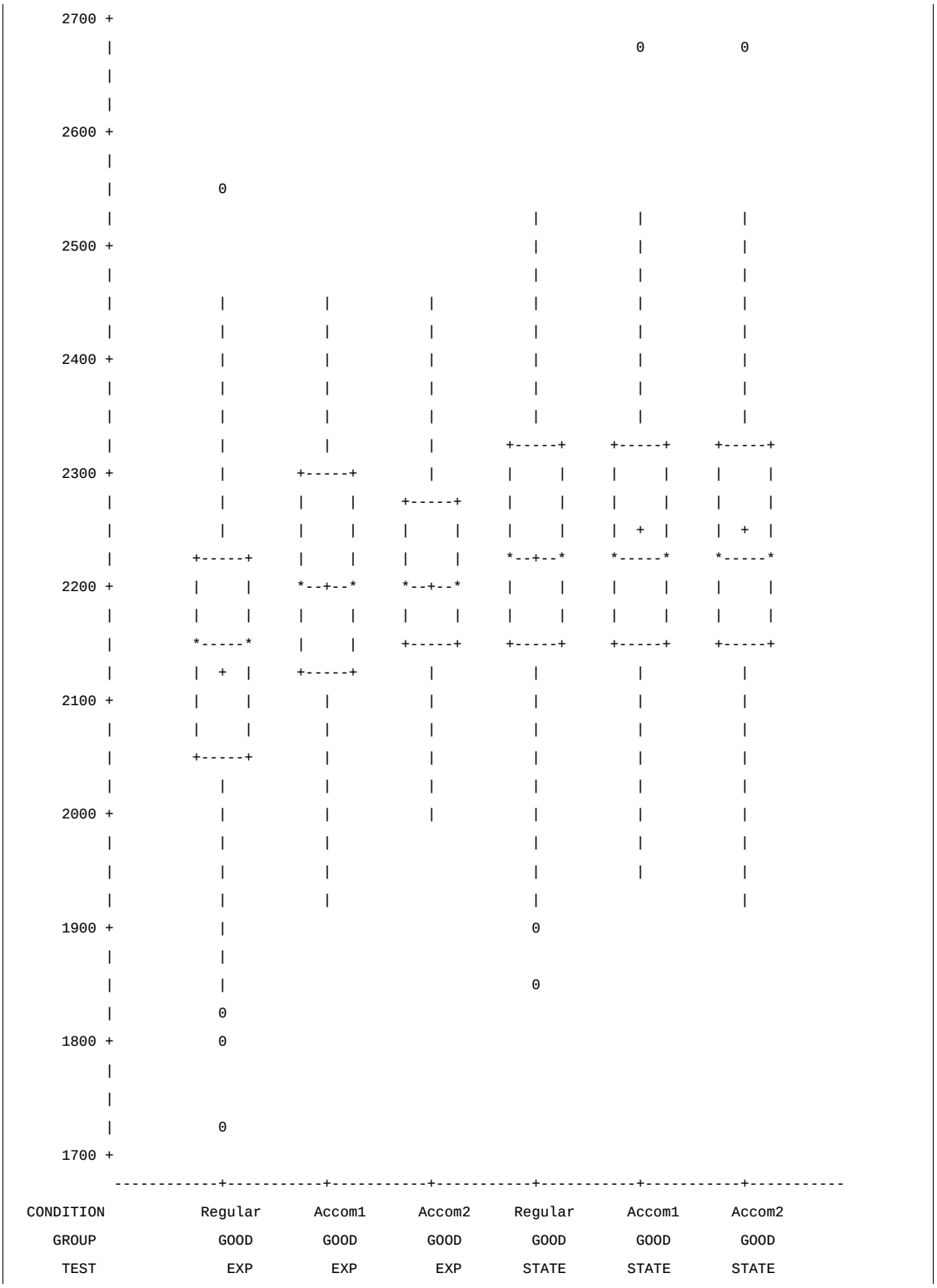
Figure 2. Frequency distributions of experimental TAKS vs. state TAKS in grade 7 average readers by accommodation condition.