

## *Measures*

*Texas Assessment of Knowledge and Skills* (TAKS; TEA, 2004a; 2004b, 2004c)

The TAKS is a criteria-referenced reading comprehension test developed by Pearson Educational Measurement in conjunction with the Texas Educational Agency (TEA) and is the Texas accountability test. The TAKS is not timed, with different assessments used for each grade that are criterion-referenced and aligned with grade-based standards from the Texas Essential Knowledge and Skills (TEKS). Test development is based on an interactive item development process in which items are field tested and evaluated for reliability, validity, and bias by language experts from TEA and Pearson, and by educators from across the state. Students read passages (both expository and narrative) and answer questions. After reading each passage, which typically has a title and illustrative pictures, students answer several multiple choice questions designed to access the literal meaning of the passage, vocabulary, and different aspects of critical reasoning about the material read. The goals across the passages are to assess students' comprehension of the literal meaning of the passages, understanding of critical vocabulary, and ability to reason using information in the paragraph. The internal consistency (coefficient alpha) of the Grade 7 test is .89 (TEA, 2004c). In preliminary latent-variable analyses of the students in Grades 6-8 who comprised the parent sample of the students reported here, the TAKS measure loaded strongly on a comprehension factor with other measures of reading comprehension including the WJ-III Passage Comprehension subtest and the GRADE (see below). TAKS 2006 Reading scores account for 44% of the variance in TAKS 2007 Reading scores for students in grades 6-8. Additionally, although the TAKS is a complex measure of comprehension, students who

failed this assessment also had a high rate of both decoding and/or fluency difficulties as well. Other studies evaluating criterion-related validity have compared student performance on TAKS with student performance on national assessments, such as The National Assessment of Educational Progress (NAEP) and the norm-referenced Iowa Tests®, and on college readiness measures (TEA, 2006).

Students are initially assessed in February of each academic year, although students who fail are given additional opportunities to pass. Only initial testing results are discussed in this report. Raw scores are converted to both standard scores as well as lexile scores. Standard scores are the dependent measure used in this report; pre-test results are from spring of these students' fifth grade year, and post-test results are from spring of these students' sixth grade year. Given the "high stakes" nature of this measure, and the fact that it is criterion referenced, in addition to standard scores, proportions of students who passed in the treatment groups were also considered across years.

*Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001)*

Students were administered the subtests of Passage Comprehension and Listening Comprehension, at both pre-test and post-test. The GRADE is a group-based, norm-referenced untimed test. For Passage Comprehension, the students read one or more paragraphs and answer multiple choice questions that require questioning, predicting, clarifying, and summarizing text. With 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 examiner read. Reliability and validity information is provided in the GRADE technical manual (Williams, 2001). Students were administered form A at pre-test and post-test. The dependent measures analyzed were the stanine score of

Listening Comprehension and a standard score computed for Passage Comprehension. The GRADE produces a stanine score for the Passage Comprehension subtest, but for purposes of this study the raw score was prorated to derive a standard score for the GRADE Comprehension Composite, which is typically based on the Passage Comprehension and Sentence Comprehension measures (the latter was not administered).

Coefficient alpha for the Passage Comprehension subtest in the grade 6 sample of 327 struggling readers and 249 typical who contributed data throughout the year was .82 at pre-test time point. Coefficient alpha for the Passage Comprehension subtest in the grade 7-8 sample of 436 struggling readers and 440 typical who contributed data through the year was 0.88 at pre-test time point. The criterion related validity of the GRADE Passage Comprehension with TAKS Reading was 0.64 in a sample of 1421 middle school students in grades 6-8.

*Woodcock-Johnson III Tests of Achievement*

(WJ-III; Woodcock, McGrew, & Mather, 2001)

The WJ-III is a nationally standardized individually administered battery of achievement tests, with excellent psychometric properties (McGrew & Woodcock, 2001). The subtests administered in this study were Letter-Word Identification, Word Attack, Passage Comprehension, and Spelling. These subtests were administered at pre-test and post-test, except for Spelling, which was administered at post-test only. Letter-Word Identification assesses the ability to read real words, and Word Attack examines the ability to read phonetically correct nonsense words. Passage Comprehension utilizes a cloze procedure to assess sentence level comprehension by requiring the student to read a sentence or short passage and fill in missing words based on the overall context. The

Letter-Word Identification, Word Attack, and Passage Comprehension subtests were administered individually to students. The Spelling subtest involves orally dictated words written by the examinee to assess encoding skills, which are related to decoding ability; this measure was modified for group administration by administering a set list of items. Standard scores from these subtests were the dependent measures of interest. Coefficient alphas in the grade 6 sample of 327 struggling readers and 249 typical who contributed data throughout the year for Letter Word Identification, Word Attack, and Passage Comprehension subtests at pre-test were .97, .93, and .94, respectively, and at post-test were .92, .99, and .85, respectively; coefficient alpha for Spelling at post-test was .84. Coefficient alphas in the grade 7-8 sample of 436 struggling readers and 440 typical who contributed data throughout the year for Letter Word Identification, Word Attack, and Passage Comprehension subtests at pre-test were .98, .94, and .96, respectively, and at post-test were .97, .99, and .83, respectively; coefficient alpha for Spelling at post-test was .94. The criterion related validity of WJ-III Letter Word Identification with TAKS Reading was 0.52, WJ-III Word Attack with TAKS Reading was 0.34, and WJ-III Passage Comprehension with TAKS was 0.61 in a sample of 1421 middle school students in grade 6-8.

*Test of Sentence Reading Efficiency (TOSRE; Wagner, in press)*

The TOSRE is a 3-minute, group-based assessment of reading fluency and comprehension. Students are presented with a series of short sentences, and are required to assess their veridicality. The raw score is the number of sentences correctly identified as true or false, minus the number of incorrect responses, within the time limit; skipped sentences are ignored, and if the number of incorrect responses exceeds the number

correct, a raw score of 0 is recorded. The TOSRE was standardized on 2,000 students from grades 4-9. The standard score was the dependent measure utilized. The mean intercorrelation of performances across the five time points in the grade 6 sample of 327 struggling readers and 249 typical readers was .79 for standard scores, and .80 for raw scores. The mean intercorrelation of performances across the five time points in the grade 7-8 sample of 436 struggling readers and 440 typical readers was .96 for standard scores, and .96 for raw scores. These correlations likely under-estimate reliability since some students received intervention and may have changed their rank order over time. The criterion related validity of TOSRE with TAKS Reading was 0.56 in a sample of 1421 middle school students in grade 6-8.

*AIMSweb Reading Maze (Shinn & Shinn, 2002)*

The Maze subtest, a 3-minute, group-administered curriculum based assessment of fluency and comprehension, was administered at all 5 time points. Students are presented with a 150-400 word passage, and for every seventh word after the first sentence students are asked to identify a correct target word from among three choices. The raw score is the number of targets correctly identified within the time limit, and was the dependent measure utilized. AIMSweb provides 15 different stories for sixth grade, and the particular story any individual student received was randomly determined, within school and treatment group. These measures are not equated, although stories were chosen based on reading level. Reliability and validity information is reported (Fuchs, & Fuchs, 1992; Jenkins & Jewell, 1993; Shinn, Deno, & Espin, 2000; Shinn & Shinn, 2002). The mean intercorrelation of performances across the five time points in the grade 6 sample of 327 struggling readers and 249 typical readers was .79. The mean

intercorrelation of performances across the five time points in the grade 7-8 sample of 436 struggling readers and 440 typical readers was .95. Mean intercorrelations of performance are likely to be an under-estimate of reliability since some students received intervention and may have changed their rank order over time. The criterion related validity of AIMSweb Maze Reading with TAKs Reading was 0.39 in a sample of 1421 middle school students in grades 6-8.

*Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999)*

Single-word fluency subtests of Sight Word Efficiency and Phonemic Decoding Efficiency were administered at pre-test and post-test. The participant is given a list of 104 words and asked to read them as accurately and as quickly as possible; the number of words read correctly within 45 seconds is recorded. Psychometric properties are good, with most alternate forms and test retest reliability coefficients at or above .90 in this age range; further information on reliability and validity is provided in the TOWRE manual (Torgesen, Wagner, & Rashotte, 1999). Standard scores from these subtests were the dependent measures. The criterion related validity of TOWRE Sight Word Efficiency with TAKS Reading was 0.40 in a sample of 1421 middle school students in grades 6-8. The criterion related validity of TOWRE Phonemic Decoding Efficiency with TAKs Reading was 0.33.

*Passage Fluency (PF) (Francis, D., Barth, A., Cirino, P., Reed, D., & Fletcher, J. 2008)*

The PF was designed by the authors specifically for this study. The PF consists of graded passages administered as short 1 minute probes to assess fluency of text reading. The entire PF consists of 100 passages for use in grades 6-8 with both narrative and

expository text structure. All passages averaged approximately 500 words each, and ranged in difficulty from 350 to 1400 Lexiles (Lexile Framework, 2007). Within each of 10 “Lexile bands” separated by 110 Lexile units, there are 10 passages, 5 of which are expository and 5 narrative. The passages were derived from former TAKS, TAAS (the precursor to TAKS in Texas), SDAA, or Texas Primary Reading Inventory (TPRI) passages, or else were created specifically for this assessment of reading fluency. A comprehension assessment was also created for these measures, which required participants to read the entire passage, and then answer approximately 8 explicit and implicit multiple-choice comprehension questions (which were developed specifically for the PF measure); however, the comprehension portion was administered only at pre-test, and so is not further discussed.

Students were administered the PF at five time points throughout the academic year, including pre- and post-test. At each time point, students read five stories for one minute each, one from each of five Lexile bands (thus having an overall range of 550 Lexiles). The particular stories any individual student received were randomly determined, within school, grade, and treatment group. Both raw scores (words correct per minute) and linearly equated scores raw scores are available. We recognize that multiple types of equating are possible, but chose linear equating due to its relative ease of implementation, its sufficiency when score distributions are normal, as they were in the present case, and given that all stories were administered to students from the same population using a random equating design. For the purposes of carrying out story equating, the larger sample of 1803 middle school students in grades 6-8 was used. Equating was carried out within grade and time-point, such that the equated scores

eliminated differences between stories in mean differences and in within-story variability at a particular time point, but allowed differences over time and across grade in both mean performance and variability in performance. Therefore differences in mean performance across time points and grades are preserved but any resulting differences are not due to, for example, older students reading easier passages, or students reading difficult passages followed by reading easier passages later in the year.

To illustrate the process, actual words read by a student are converted to a standardized score that corrects for the difficulty of a story. If the story a student is reading is a relatively easier one, their standardized score would be adjusted downward, whereas if the story a student is reading is a relatively difficult one, their standardized score would be adjusted upward. Therefore, it is possible that if one student reads 130 words on a relatively easy story, and another student reads 110 words on a relatively difficult story, both of their standardized scores could be equal to 120 words. The mean intercorrelation of the five stories read at pre-test in the grade 6 sample of 327 struggling readers and 249 typical readers was .87, and for the three stories read at post-test was .86. The mean intercorrelation of the five stories read at pre-test in the grade 7-8 sample of 436 struggling readers and 440 typical readers was .98, and for the three stories read at post-test was .96. The criterion related validity of mean passage fluency with TAKS Reading was 0.50 in a sample of 1421 middle school students in grades 6-8.

*Word List Fluency* (Francis, D., Barth, A., Cirino, P., Reed, D., & Fletcher, J. 2008)

The WLF was also designed by our research team to assess word reading fluency in outside of context on lists of varying difficulty. Students are required to read as many words as possible within one minute, for three word lists. The WLF contains two

different word list pools, constructed lists and passage lists. The first pool, constructed lists, are 21 timed lists subdivided into 7 “easy” lists, 7 lists of “moderate” difficulty, and 7 “challenging” word lists. These words were derived from the Zeno (1991) word bank and each list is comprised of approximately 150 words, constructed on the basis of word length and frequency parameters. In this context, “easy” words were less than 5 letters and frequency values were high, with the individual word lists constructed from a random sample of such words. Words for the “moderate” words lists were constructed in a similar manner, with the word length parameter now 6-10 letters, but with the frequency parameter remaining high. The word length parameter for “challenging” word lists remained at 6-10 letters, but the frequency parameter was now adjusted to low. The second pool, passage lists, consisted of 38 word lists derived from the PF measure described above. For each story, words were identified and duplicates removed, leaving only unique words. These words were then randomly ordered within a passage, and arranged into word lists for each passage.

Students were administered the WLF at five time points throughout the academic year, including pre- and post-test. At each time point, students read three word lists of varying difficulty for one minute each. Students were randomly assigned to read one of three types of word list: a) passage word lists derived from the *same* stories they read to assess PF, b) passage word lists derived from stories that *other* students read to assess PF (but they themselves did not read), or c) constructed word lists. Students reading constructed lists read one easy, one moderate, and one challenging list. Students reading passage lists read one list based on a passage rated at the lowest of five Lexile bands for students in that grade, another list based on a passage at the third of five Lexile bands,

and a third list based on a passage at the fifth of five Lexile bands. Again, the particular stories, and thus, the particular lists, that any individual student received was randomly determined, within school, grade, and treatment group. As with PF, the dependent measure utilized for WLF is a linearly equated-score average of the three, one-minute word list reads, derived in a similar manner as described above. The mean intercorrelation of the three word lists read in the grade 6 sample of 327 struggling readers and 249 typical readers was .92 at pre-test and .89 at post-test. The mean intercorrelation of the three word lists read in the grade 7-8 sample of 436 struggling readers and 440 typical readers was .97 at pre-test and .98 at post-test. The criterion related validity of mean word list fluency with TAKS Reading was 0.36 in a sample of 1421 middle school students in grades 6-8.

*Kaufman Brief Intelligence Test- 2 (K-BIT-2; Kaufman & Kaufman, 2004)*

The K-BIT 2 is an individually administered intellectual screening measure, used in this study for descriptive purposes. The K-BIT 2 was standardized on 2,120 examinees who ranged in age from 4-90, and who matched the 2001 population survey in education level, race/ethnicity, and region. Internal consistency values for the subtests and composite range from .87 to .95, and test-retest reliabilities range from .80 to .95, in the age range of the students in this study (Kaufman & Kaufman, 2004). The K-BIT 2 correlates with several other intellectual measures in the .75 to .90 range, for examinees at the age of the current sample (Kaufman & Kaufman, 2004). The Matrices subtest was administered at pre-test, and requires students to choose a diagram from among five or six choices that either “goes with” a series of other diagrams, or completes a series, or completes a 2 x 2 analogy. The most difficult items use abstract stimuli to complete a 2 x

2 or 3 x 3 matrix. The Verbal Knowledge subtest was administered at post-test, and assesses receptive vocabulary and general information (e.g., nature, geography). The participant is required to choose one of six illustrations that best corresponds to an examiner question. The K-BIT 2 also has a Riddles subtest that was not utilized; instead, the Verbal Knowledge score was prorated for the verbal domain, and therefore verbal and nonverbal standard scores were utilized, as well as a composite. The criterion related validity of KBIT Verbal Knowledge with TAKS Reading was 0.54 in a sample of 1421 middle school students in grades 6-8. The criterion related validity of KBIT Matrices with TAKS Reading was 0.38.

## References

- Francis, D., Barth, A., Cirino, P., Reed, D., & Fletcher, J. (2008). *The Texas Middle School Fluency Assessment*. Austin, TX: Texas Educational Agency.
- Kaufman, A. S. & Kaufman, N.L. (2004). *Kaufman Brief Intelligence Test, Second Edition (K-BIT-2)*. Minneapolis, MN: Pearson Assessment.
- Shinn, M. R., & Shinn, M. M. ( 2002). *AIMSweb Training Workbook: Administration and scoring of reading Maze for use in general outcome measurement*. Edformation, Inc.: Eden Prairie, MN.
- Texas Educational Agency. (2004b). TAKS: Texas Assessment of Knowledge and Skills. Information Booklet: Reading, Grade 6 – Revised. Author. Downloaded from <http://www.tea.state.tx.us/student.assessment/taks/booklets/reading/g6e.pdf> on December 19, 2007.
- Torgesen, J., Wagner, R., & Rashotte, C. (1998). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.
- Wagner, R. K. et al. (in press). *Test of Sentence Reading Efficiency (TOSRE)*. PRO-ED, Inc.: Austin, TX.
- Williams, K. T. (2001). *The Group Reading Assessment and Diagnostic Evaluation (GRADE)*. Teacher's scoring and interpretive manual. American Guidance Service, Inc: Circle Pines, MN.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing.