Differences in the Relationship of Oral Reading Fluency and High-Stakes Measures of Reading Comprehension

Jeanne Wanzek,
Florida State University, Tallahassee

Greg Roberts,
University of Texas at Austin

Sylvia Linan-Thompson,
University of Texas at Austin

Sharon Vaughn,
University of Texas at Austin

Althea L. Woodruff, and
University of Texas at Austin

Christy S. Murray
University of Texas at Austin

Abstract

The current study examined the predictive validity of oral reading fluency measures across first, second, and third grades for two reading achievement measures at the end of third grade. Oral reading fluency measures were administered to students from first grade to third. The *Texas Assessment of Knowledge and Skills* and the *Stanford Achievement Test* were also administered in the third grade. Oral reading fluency was a reliable predictor of student success on both measures. Data suggest that greater student growth in oral reading fluency is needed through the grade levels to ensure high probabilities of success on the nationally normed measure, as compared to what is needed for the state-normed measure. Implications for practice and future research are discussed.

Keywords

reading curriculum-based measurement; predictive validity; high-stakes testing

Recent federal education policy has increased the focus on accountability standards for student academic success and has mandated annual assessment of student outcomes in reading and language arts beginning in third grade (No Child Left Behind Act, 2002). All 50 states and the District of Columbia now require schools to administer a state-defined assessment measuring student reading and language arts outcomes. Measures have been developed in most states and normed at the state level, but some states use commercially prepared standardized achievement measures normed at the national level, such as the *Iowa Test of Basic Skills* (Hoover, Dunbar, & Frisbie, 2005) and the *Stanford Achievement Test* (SAT; Harcourt Educational Measurement, 2003). Whether states administer state-

developed measures (state normed or criterion referenced) or nationally normed measures, student achievement at preselected grade levels is measured and reported by all schools. The results of these measures are used to monitor school and district performance, identify schools in need of improvement, and, in some cases, determine student retention (Florida Statute, 2002; No Child Left Behind Act, 2002). A given student's school experience, as well as a school's status in providing education to students, is often shaped by performance on these measures, prompting many to refer to the measures as "high-stakes" assessments (e.g., Zimmerman & Dibenedetto, 2008).

Beginning in the 2005–2006 school year, all schools were required to administer annual measures of accountability in reading by third grade. This grade is often considered a pivotal year in students' reading because instruction on how to read often fades in most curricula by the end of third grade, when the focus shifts to reading to learn (Snow, Burns, & Griffin, 1998). As a result, the expectation is that if effective instruction is in place, students will demonstrate grade-level reading abilities by third grade.

Measurement of Oral Reading Fluency

To effectively support students in becoming successful readers by third grade, school psychologists, teachers, and administrators need ongoing information to monitor their progress throughout the grade levels before high-stakes accountability assessment. Many schools and districts use screening and progress-monitoring assessments throughout the early grade levels to identify students in need of early reading intervention, to specify instructional requirements, and to monitor student progress (Vaughn et al., 2008). Curriculum-based measurement is a means for screening and monitoring students' progress throughout the elementary grades (Deno, 1985, 2003), which in the area of reading, most often uses oral reading fluency, as measured by 1-minute samples of a student's oral reading rate and accuracy (Deno, 2003).

The validity of oral reading fluency, as an indicator of student progress in reading, has been examined in a number of studies, with results demonstrating high correlations between oral reading fluency and measures of word recognition and reading comprehension, when the measures are administered concurrently or in close proximity (Deno, 1985; Fuchs, Fuchs, & Maxwell, 1988; Hosp & Fuchs, 2005; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Shinn, Good, Knutson, Tilly, & Collins, 1992). Accurately determining how students are progressing toward meeting third-grade accountability standards is of high interest to educators and parents. For this reason, the measures of student progress (e.g., oral reading fluency) that are used to make instructional decisions throughout the grade levels must reliably inform expected student performance on the third-grade outcome measures (whether state developed or nationally normed).

Oral Reading Fluency's Concurrent and Predictive Validity on High-Stakes Measures

The criterion-related validity data between oral reading fluency and state-developed criterion assessments within grade levels are considerable. For example, McGlinchey and Hixson (2004) reported correlations of .63 to .81 between oral reading fluency (administered 2 weeks before the state test) and the state-developed Michigan test, for eight successive cohorts of fourth-grade students. Seventy-four percent of students with a fluency score of 100 words correct per minute or more were correctly classified as passing the state test. Good, Simmons, and Kame'enui (2001) reported that third-grade spring-administration fluency scores equaling 110 words per minute or more predicted a .96 probability of passing the Oregon state test. Similarly, Stage and Jacobsen (2001) examined fourth-grade oral

reading fluency scores as well as slope in predicting student success on the state-developed Washington reading assessment given at the end of fourth grade. Oral reading fluency scores increased the predictive power of performance on the state reading assessment by 30% over base rates. The authors also reported that September oral reading fluency scores provided better predictive power than did student slope across fourth grade. Correlations were .43 or .44 between each fourth-grade oral reading fluency measurement point and the standard scores of the fourth-grade Washington state reading test.

Wood (2006) extended the research beyond one grade level by examining the concurrent relationship of oral reading fluency with state assessment performance within several grade levels. Students in third through fifth grades were administered oral reading fluency tests 2 months before the administration of the state-developed Colorado reading test (administered in February for third graders and March for fourth and fifth graders). The relationship between oral reading fluency scores and subsequent scores on the state test was equivalent within each of the three grade levels (range = .70–.75). In addition, oral reading fluency scores uniquely predicted student scores on the state test over and above the previous year's state test score. Wood also examined cut scores on oral reading fluency that demonstrated high passing rates on the state test within the same grade level. For example, third graders scoring 96 words correct per minute or more had a 98% chance of passing the third-grade test. Cut scores for fourth and fifth grade were 117 words and 135 words for passing the respective fourth- and fifth-grade state tests.

Silberglitt, Burns, Madyun, and Lail (2006) also examined the relationship between oral reading fluency and state-normed tests within several grade levels. As in Wood's study (2006), oral reading fluency measures were administered within 2 months of the administration of the state reading test (Minnesota's state test). Correlations of .50 or higher were found for each grade level (third, fifth, seventh, eighth). The magnitude of the relationship between oral reading fluency and performance on the state test declined as the grade level increased.

As seen in these studies, most previous research has examined the within-grade-level relationship of oral reading fluency with state-developed measures; however, some studies have examined the within-grade-level relationship of oral reading fluency with nationally normed standardized tests of reading comprehension, which are sometimes used by states in high-stakes testing. Schilling, Carlisle, Scott, and Zeng (2007) examined the relationship of oral reading fluency and the *Iowa Test of Basic Skills*, which was administered at the end of Grades 1, 2, and 3. Oral reading fluency was administered in the winter and spring of first grade and in the fall, winter, and spring of second and third grades. Within-grade-level correlations between oral reading fluency and the *Iowa Test of Basic Skills* comprehension were .69 to .74 for first grade, .68 to .75 for second grade, and .63 to .65 for third grade. Compared to Schilling et al., Klein and Jimerson (2005) reported slightly higher correlations for the concurrent relationship between oral reading fluency and the ninth edition of the SAT total reading scores (SAT-9; vocabulary and comprehension scores). In the two cohorts of students, ranges were reported as follows: .80 to .84 for first grade, .74 to .77 for second grade, and .77 to .81 for third grade.

Roehrig, Petscher, Nettles, Hudson, and Torgesen (2008) recently investigated the within-grade-level relationship between oral reading fluency and both the state-developed Florida reading test and the nationally normed SAT, 10th edition (SAT-10). Oral reading fluency scores from September, December, and February/March of third grade were used to predict performance on the reading portion of the state test administered in February/March. The full sample was split in two, to allow for a calibration sample and a cross-validation sample. The authors reported high correlations between oral reading fluency and both the state-

developed test (*r* = .66–.71) and the SAT-10 (*r* = .68–.71), with the concurrent administration of oral reading fluency (February/March) demonstrating the strongest correlation for both criterion measures.

These studies support the validity of oral reading fluency in relation to student reading achievement on state-developed and nationally normed high-stakes measures. Moreover, oral reading fluency measures a construct similar to what is being assessed by the high-stakes measures—presumably, reading competence or reading ability. However, the value of oral reading fluency in predicting later performance, across grade levels, has received less attention. In particular, whether there are differences in rates of growth necessary to achieve proficiency on these measures has not been studied.

Oral Reading Fluency's Predictive Validity Across Grade Levels

Spear-Swerling (2006) investigated students' oral reading fluency in third grade (winter administration) and its relationship to fourth-grade performance on the reading portion of the Connecticut state reading test. Third-grade oral reading fluency contributed unique prediction to student scores on the fourth-grade state test (*r* = .65), although students' third-grade Woodcock–Johnson comprehension scores accounted for the most variance in the fourth-grade state test scores.

Crawford, Tindal, and Stieber (2001) also examined the predictive power of oral reading fluency across one grade level, in relation to the Oregon state reading test. The correlation between students' second-grade oral reading fluency scores (administered in January) and performance on the third-grade Oregon state reading test (administered in March; *r* = .66) was higher than the correlation between students' third-grade oral reading fluency scores (administered in January) and their performance on the third-grade state reading test (*r* = .60). All students reading at 72 words correct per minute or higher in January of second grade passed the third-grade state test.

A third study of predictive validity across one grade level (Riedel, 2007) investigated the relationship of first-grade oral reading fluency and end-of-second-grade comprehension scores on the *TerraNova* (CTB/McGraw-Hill, 2003), a measure used by some states to make high-stakes decisions. Correlations were .49 for winter first-grade oral reading fluency and .54 for spring first-grade oral reading fluency.

Klein and Jimerson (2005) examined the longitudinal relationship between spring first-grade oral reading fluency and third-grade SAT-9, which resulted in a moderate relationship (*r* = .68). Hintze and Silberglitt (2005) also found moderate relationships between oral reading fluency and the Minnesota state test for Grades 1, 2, and 3. They compared the relative benefits of discriminant analysis, logistic regression, and receiver-operating characteristic curve analysis. Correlations between oral reading fluency and third-grade state test performance ranged from .49 to .58 for first-grade oral reading fluency, .61 to .68 for second-grade oral reading fluency, and .66 to .69 for third-grade oral reading fluency. Oral reading fluency was a consistently reliable and accurate predictor of performance on the state test, with logistic regression demonstrating high diagnostic accuracy and similar cut scores, whether the state test was used as the sole criterion or the oral reading fluency assessments were used as successive criteria across grades, ultimately predicting success on the state test. However, both logistic regression and the receiver-operating characteristic analysis using oral reading fluency assessments successively yielded the greatest diagnostic accuracy, and combining logistic regression and receiver-operating characteristic analysis provided the most reliable estimates of target scores for later success on a high-stakes measure (Silberglitt & Hintze, 2005).

More recently, Baker et al. (2008) examined oral reading fluency intercept and slope across two grade levels in predicting performance on the Oregon state reading test and SAT-10. Correlations from first- and second-grade oral reading fluency in relation to end-of-second-grade SAT-10 scores ranged from .63 to .80 and demonstrated a gradual increase in relationship strength with each oral reading fluency administration. A slightly weaker relationship was reported between oral reading fluency administrations in Grades 2 and 3 and the end-of-third-grade Oregon state reading test (range .58 to .68). Including oral reading fluency slope with the intercept in the prediction models increased the amount of variance accounted for in both the SAT-10 and the Oregon state test.

Before Baker et al. (2008), the research examining oral reading fluency across grade levels had relied on discrete oral reading fluency scores to predict performance on state-developed reading tests in later grades. Even studies that used multivariate analyses posited a series of individual oral reading fluency values as independent variables, either in a stepwise fashion or as separate indicators in a multiple regression model. Baker et al. demonstrated the usefulness of students' slope across grade levels in predicting later reading achievement on high-stakes measures. However, because the two criterion measures were administered in different years, the authors were unable to compare the student growth that would predict proficient levels on each criterion measure. No previous study has provided this direct comparison. Thus, we sought to conduct a direct assessment of the relationship of trends in oral reading fluency over time (across grade levels) and later performance on two types of high-stakes assessments currently in use: one nationally normed measure and one state-developed measure.

Purpose

Our purpose is to extend the research on the predictive validity of oral reading fluency across grade levels in relation to two types of measures used by states for accountability purposes: first, a nationally normed measure, the SAT-10 (Harcourt Educational Measurement, 2003); second, a state-developed test, the *Texas Assessment of Knowledge and Skills* (TAKS; Texas Education Agency, 2004). Of particular interest is the interrelationship of students' status in the winter of first grade, their rate of reading progress over the subsequent 2.5 years, and their third-grade performance on the two measures. The use of oral reading fluency to monitor student progress assumes that trends in reading achievement are related to later performance. As such, two questions were addressed: What parameters describe the best-fitting curves across first, second, and third grades on a nationally normed measure and a state-normed measure?

Given these parameters, what are the probabilities of later success on the nationally normed measure and the state-normed measure given oral reading fluency performance at earlier points in time?

Method

Participants

Students from one school district in Texas (six elementary schools) were followed from first grade to third. The schools are high-poverty Title I schools with a high percentage of minority students. Third-grade students in the district's English classrooms were included in the study if they (a) had obtained scores on the third-grade Texas state assessment (TAKS) and the SAT-10 comprehension subtest given in the spring of 2006 and (b) had at least one oral reading fluency score between winter of first grade and spring of third grade. Students in the district's bilingual classrooms were not included in the study. A total of 461 students composed the sample (90% of the third-grade class). Approximately 66% of the total sample

was Hispanic, with approximately 75% of the sample participating in the free or reduced-price lunch program. Table 1 summarizes the demographic data.

Measures

Oral reading fluency—Oral reading fluency was assessed with the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS; Good & Kaminski, 2002), by administering the oral reading fluency benchmark passages to individual students in the winter and spring of first grade and in the fall, winter, and spring of second and third grades. During each administration, students were asked to read three grade-level passages aloud for 1 minute. Words omitted, words substituted, and hesitations of more than 3 seconds are scored as errors. Words self-corrected within 3 seconds are scored as accurate answers. The median score of the three passages was recorded and used for analyses. Alternate-form reliabilities range from .89 to .96 (Good, Kaminski, et al., 2001). Concurrent validity with the *Test of Oral Reading Fluency* (Children's Educational Resources, 1987) ranges from .91 to .96 (Good, Kaminski, et al., 2001).

SAT-10—The SAT-10 is a group-administered, norm-referenced test. In this study, the reading comprehension subtest (Form A) was administered in third grade at the beginning of May. The reading comprehension subtest requires the student to read passages and answer multiple-choice questions about the passage. The internal consistency reliability coefficient is .93 for the reading comprehension subtest (Form A) of Grade 3. The SAT-10 reading comprehension subtest highly correlates ($r = .80–.85$) with the SAT-9 reading comprehension subtest at the primary level. Scaled scores were used for analyses in this study. According to the publisher's guidelines, proficiency on this subtest represents mastery of the grade-level skills necessary for success at the next grade level, defined as a scaled score of 634 or higher.

TAKS—TAKS is a group-administered test measuring the students' mastery of the Texas state curriculum in Grades 3 through 9. The third-grade reading portion of the test was administered at the end of February of the students' third-grade year. In the state of Texas, students must pass the TAKS to be eligible for promotion to fourth grade. The TAKS reading test consists of 36 multiple-choice questions related to various passages, read independently by the student. Passages include narrative, expository, and mixed text (both narrative and expository) and are 500 to 700 words long. In this study, proficiency information for each student is based on the 2006 state-determined standard. Scaled scores were used for analyses in this study. A scaled score of 2100 on the TAKS is considered passing, representing mastery of the state-defined reading knowledge and skills necessary for advancing to the next grade level. The percentage correct required for a scaled score of 2100 varies from year to year, as based on the difficulty of the test, and so ranges from 58% to 69% correct (Texas Education Agency, 2006).

Procedures

Oral reading fluency measures were individually administered to participants by trained graduate students and research associates in the winter and spring of first grade and in the fall, winter, and spring of second and third grades. The number of examiners ranged from 12 to 19 across the eight rounds of testing. After examiners were trained, they participated in administration practice sessions before each round of testing. Examiners were required to record reading errors on a protocol while listening to someone read a passage with scripted errors. Scoring was checked, and if discrepancies were found, examiners completed a second round of practice to ensure accurate scoring. Interrater reliability was then calculated by taking the number of agreements and dividing by the sum of the number of agreements and disagreements. All examiners were required to complete practice administrations to a

criterion of 90% reliability or higher before each round of testing. In addition, to ensure the maintenance of this level of reliability (90% or higher), as well as the adherence to scoring procedures, fidelity-of-implementation and reliability-of-scoring checks (with forms provided in the DIBELS administration and scoring guide) were conducted for the examiners during each round of testing, during unannounced visits. All testers met criterion of 90% or higher during the test administration.

The TAKS was administered as part of the state requirements for testing, whereas the SAT-10 was administered as part of the research study. Classroom teachers at each school administered the TAKS at the end of February and the SAT-10 at the beginning of May. Research team members proctored each administration of the SAT-10 to ensure reliable administration. No discrepancies were noted. School procedures were followed for TAKS administration, and no outside personnel were allowed in the school during administration. Although the state of Texas has implemented statewide assessment since 1980, the current TAKS measure had been administered by all schools in the state for 4 years at the time of this study. It was the second year that the third-grade teachers were administering the SAT-10 measure. The SAT-10 measure was scored by Harcourt via a Scantron machine, and scores were sent to the research team. The state of Texas provided the district with the students' TAKS scores, and the district provided these scores to the research team.

Results

Results are organized according to the two research questions: First, what parameters describe the best fitting curves across first, second, and third grades on a nationally normed measure and a state-normed measure? Second, what are the probabilities of later success on the nationally normed measure and the state-normed measure, given oral reading fluency performance at earlier points in time? Distributional assumptions were evaluated during preliminary analyses, and all relevant indicators of normality were within established limits.

A total of 461 students met the selection criteria, from which 87.0% achieved a passing score on the TAKS (scaled score of 2100 or higher) and 41.5% scored at a proficient level or higher on the SAT-10 (scaled score of 634 or higher).

Table 2 summarizes the oral reading fluency results and the number of cases at each measurement point.

Curve-Fitting Analyses

To address the first research question, we estimated latent factors representing students' growth in oral reading fluency and used them as indicators of later proficiency on two third-grade measures of reading achievement (i.e., TAKS and SAT-10). Growth was modeled as linear across the three school years, using measures from spring of first grade and from fall, winter, and spring of second and third grades. Latent growth factors of the best-fitting line were used as predictors of *passing* status on the TAKS and the SAT-10. Missing data were treated as random, and full information maximum likelihood was used for estimation, allowing the inclusion of all cases with a criterion score and at least one oral reading fluency data point (see Singer & Willet, 2003). Full information maximum likelihood utilizes all available data, when a missing-at-random assumption is tenable. Full information maximum likelihood assumes that unobserved data points have an expectation equal to a model-derived value estimated from the remaining data points (Muthén & Muthén, 2006). Maximum likelihood estimation outperforms listwise deletion for parameters involving many recouped cases and provides more reliable standard error estimates (Newman, 2003). The percentage of missing cases in the sample of 461 ranged from 53% at Time 1 (winter of first grade) to 5% at Time 8 (spring of third grade), with a generally linear decrease in

percentage missing over time; all 461 cases were present at Time 9. Table 3 presents the results of the curve-fitting analysis for the overall models. The average winter of first-grade oral reading fluency score in the linear model was 16.31 words correct per minute with variance of approximately 750 and a standard deviation of 27.4. The average slope was 11.33 words correct per minute per time point, with variance of 19.2 and a standard deviation of 4.38.

Probabilities of Grade-Level Proficiency on Outcome Measures

To address the second research question, we conducted a logistic regression of SAT-10 and TAKS status on the latent growth factors, yielding estimates of the increased likelihood associated with changes in slope and/or intercepts, expressed as odds ratios. Student growth on oral reading fluency and its relationship to SAT-10 and TAKS (i.e., the logistic relationship of growth factors to outcome) were specified within a single model.

Tables 4 and 5 present the results for the logistic regression. Each term in the different equations represents contributions to the estimated log odds such that for each unit change in X_j, there is a predicted change of b_j units in the log odds in favor of $Y = 1$ (in this case, grade-level proficiency on the TAKS or SAT-10). Odds ratio is the exponentiation of log-odds estimates, and it represents the relative frequency with which outcomes occur—specifically, the number of cases with the event in a group to the number of cases without the event. The probability of an event's occurring is derived from odds using the logistic function. Of primary interest in this study were the probabilities of achieving grade-level proficiency scores on the TAKS and SAT-10 given different estimates of slope and intercept.

For the TAKS model, coefficient values were .055 and .343 for the winter first-grade oral reading fluency and slope, respectively (Table 4). These estimates were statistically significant. The winter first-grade oral reading fluency and slope estimates for the SAT-10 linear model were .044 and .290, respectively (Table 5). The probability of passing for the given values of intercept and slope was determined by solving Equation 1 (for the TAKS) and Equation 2 (for the SAT-10) for intercept and slope values of interest.

$$\pi_{\text{linearTAKS}} = e^{-2.00+.055(\text{intercept})+.343(\text{slope})} / 1 + e^{-2.00+.055(\text{intercept})+.343(\text{slope})} \tag{1}$$

$$\pi_{\text{linearSAT-10}} = e^{-4.26+.044(\text{intercept})+.290(\text{slope})} / 1 + e^{-4.26+.044(\text{intercept})+.290(\text{slope})} \tag{2}$$

Figure 1 presents estimated aim lines for the two measures, based on results of the logistic regression. Initial status was held constant at 20 words correct per minute (the DIBELS benchmark for winter of first grade; Good & Kaminski, 2002), and the probability of reaching grade-level proficiency was set at .8 and .9 for the two third-grade measures. The equations were solved for slope, and the plotted lines reflect the resulting estimates. For TAKS, the approximate slopes were 6.7 for a probability of .8 of proficiency and 9.0 for a probability of .9 (i.e., gains of about 6.7 and 9.0 words correct per minute from one oral reading fluency administration to the next, with no correction for summer learning loss). Reaching grade-level proficiency on the SAT-10 proved considerably more difficult. For this nationally normed assessment, a probability of .8 of proficiency required a gain of about 16.4 words correct per minute from one oral reading fluency measurement point to the next (again, without accounting for summer decrements), whereas an increase of approximately 19.8 words correct per minute from one measurement point to the next was required to have a probability of .9 of proficiency.

The data presented in Table 6 represent the probabilities of reaching proficiency given different intercept and slope scenarios. Predicted values for intercept and slope for this sample were converted to percentile ranks, and the probability of third-grade proficiency was calculated for values corresponding to the 10th, 25th, 50th, and 75th percentiles by solving Equations 1 and 2 for these percentile ranks. The results illustrate differences between the SAT-10 and TAKS, both in terms of relative difficulty and in terms of the oral reading fluency performance in Grades 1, 2, and 3 necessary to obtain proficiency on the tests. For example, students at the 10th percentile for intercept and slope had a probability of .48 for proficiency on the TAKS but only a probability of .06 for proficiency on the SAT-10. At the 25th percentile for intercept and slope, the probabilities increased to .65 for TAKS and .1 for SAT-10. At the 50th and 75th percentiles, the probabilities for proficiency on the TAKS were .92 and .99, respectively, whereas for the SAT-10, the probabilities for proficiency at the 50th and 75th percentiles were .29 and .68, respectively.

Discussion

The purpose of this study was to examine differences in the predictive validity of oral reading fluency for two reading comprehension outcome measures used for high-stakes decisions: a nationally normed measure (SAT-10) and a state-developed measure (TAKS). Reliable models for the oral reading fluency–SAT-10 and oral reading fluency–TAKS predictive relationships were fit, indicating a positive relationship between oral reading fluency and the comprehension outcome measures. Previous research has consistently demonstrated oral reading fluency's predictive validity within grade level (e.g., Stage & Jacobsen, 2001; Wiley & Deno, 2005). This study confirms oral reading fluency as a reliable predictor of performance on both a nationally normed assessment and a state-normed assessment across grade levels.

In addition to analyzing the predictive validity of oral reading fluency for two types of measures, we sought to examine oral reading fluency performance throughout the grades, as well as proficiency on each type of measure at the end of third grade. The findings suggest that different student achievement is needed to reach proficiency levels for the two measures. Using the linear model, we considered the student growth necessary to have a high probability (.8–.9) of obtaining the identified third-grade proficiency levels on each outcome measure, given measure-specific oral reading fluency status at winter of first grade. The DIBELS benchmarks indicate that a score of 20 words per minute on oral reading fluency in the winter of first grade presents a high probability that the student will make the next benchmark (end of first grade).

Using the winter benchmark, we found that students would need an estimated slope of 6.7 words per minute between each measurement period in first through third grades to remain on track and have a probability of .8 of proficiency on the TAKS. Students would need more than double that word gain, 16.4 words per minute between each measurement period, to have a probability of .8 of proficiency on the SAT-10. Estimated gains of 9.0 words per minute between each measurement period for TAKS and 19.8 words per minute between each measurement period for SAT-10 are needed for a student to remain on track and have a probability of .9 of proficiency on each test. As a comparison, to meet the DIBELS benchmark goal of 110 words per minute or more by the end of third grade, students would need an estimated slope of 12.86 words correct per minute between each measurement period. Hasbrouck and Tindal (2006) also report a gain of about 12 words correct per minute for each measurement period for students to remain at the 50th and 75th percentiles. Thus, our data suggest that students were able to reach proficiency on the TAKS measure with relatively low growth in words correct per minute, whereas a steeper level of growth beyond

recommended benchmarks was needed to ensure a high probability of reaching proficiency on the SAT-10.

Note that the estimated slope values were calculated for use as one of several predictors of subsequent status on high-stakes assessments. The purpose of this study is not to model within-grade-level growth nor establish time-specific indicators of performance (e.g., fall of first grade). Thus, the point values are not proposed as individual benchmarks for evaluating student performance, because the linear analysis does not account for summer learning loss and other sources of nonsystematic variation (e.g., passage difficulty at different grade levels). Rather, these values provide interpretation of the key differences in the overall slope values, or the overall student growth required to reach proficiency on the TAKS and SAT-10 measures.

Examining student percentile ranks in oral reading fluency intercept and slope provides a similar picture of the differences in the oral reading fluency levels needed for later proficiency on the two outcome measures. For example, students at the 25th percentile in oral reading fluency intercept and slope in our sample have a probability of .65 of proficiency on the TAKS but only a probability of .1 of demonstrating proficiency on the SAT-10. These data suggest a relatively low level of reading fluency needed to predict success on the TAKS.

Practitioners engaging in data-based decision making about instruction need to be aware of the differences in growth trajectories that predict success on various outcome measures. School psychologists would benefit from carefully examining the proficiency levels required to reach positive outcomes on each of these measures, to guide teachers and other key personnel in providing effective instruction to students. Our data demonstrate that students are more likely to obtain proficiency on the Texas state assessment than to show proficiency on SAT-10. If student response to instruction is evaluated with progress toward passing the state assessment, students may appear to be responding when they are actually still at risk in terms of their overall reading achievement. The rate of growth needed through second and third grades to achieve proficiency on SAT-10 is more than double that needed to achieve success on the state test, suggesting a higher level of reading achievement for those students proficient on the SAT-10 measure. In fact, all students in our sample who were proficient on SAT-10 passed the TAKS.

Our results are in line with reports of discrepancies in passing rates on state tests around the country and in passing rates on the National Assessment of Educational Progress (Lee, 2006). Passing rates on state tests for most states are significantly higher than those for the National Assessment of Educational Progress, indicating that either the standard for many state tests is low or the standard for the National Assessment of Educational Progress is high. We argue that in many cases, the level of passing for state tests should be considered a minimum level of achievement. As mentioned, passing standards for state tests are often determined on the basis of high-stakes decisions. For example, proficiency on the state test may equate to the minimum level of achievement needed to continue to the next grade level. These test standards have not necessarily been set up to indicate that students have a high probability of achieving future reading proficiency levels in later grades. In fact, a comparison of the percentage of students who passed the TAKS at third grade and fifth grade shows that in the last 4 years, passing rates have been between 89% and 91% at third grade but only 75% to 81% at fifth grade. For the 2005–2006 school year, 90% of the state's third graders passed the state test, whereas 81% of fifth graders and 80% of seventh graders had the same successful outcome.

When practitioners and policy makers focus on ensuring that all students meet these minimum levels, they may set expectations too low for students and so may falsely assume that many low-performing students are on track for future success after achieving these minimum levels. As a result, students may miss out on additional assistance and the interventions that they need to reach true proficiency in reading. Practically speaking, these data suggest that practitioners may need to consider an outcome measure not attached to high-stakes decisions for measuring student levels of proficiency, to ensure that students meet more than just minimum competency levels.

Limitations

We examined the relationship of oral reading fluency with two reading outcome measures: TAKS and SAT-10. TAKS is administered by the school each year and is already connected to high-stakes decisions in the state of Texas, including student retention, school quality status, and No Child Left Behind standards. The schools monitor student progress toward passing the test throughout the third-grade year using practice TAKS tests. Third-grade students identified as being at risk for failing the TAKS are provided additional instruction toward passing it. In addition, teachers incorporate TAKS practice into their instruction for all students throughout the school months leading up to the TAKS administration. The SAT-10 was also administered as part of this study. Classroom teachers were asked to administer the SAT-10 assessment to their students, and proctors from the research team were available to ensure administration fidelity. The schools agreed to administer the test after the TAKS administration was complete. The SAT-10 results were not connected to any high-stakes decisions in the schools. The results of the SAT-10 measure, as well as other student data from the project, were shared with the school district and each school principal. Therefore, the "urgency" under which the TAKS and SAT-10 measures were administered may have differed, as well as the importance of the measures, as seen by teachers and students.

In addition, as is common with state-developed tests, the validity of the TAKS assessment has not been studied. We were able to calculate the correlation between TAKS and SAT-10 for this sample. The correlation between the TAKS reading comprehension test and the SAT-10 reading comprehension subtest (.67) provides some evidence that TAKS may be a valid assessment of reading comprehension. However, the relationship is not known between TAKS and other reading assessments or a broader sample of students. Finally, family-wise error and the potential for inflated Type I errors should be acknowledged, although the nature of the model-building analysis diminishes the degree of threat (Singer & Willet, 2003).

Future Research

In this study, we identified student growth trajectories for predicting successful performance on two third-grade reading measures used as high-stakes assessments. Practitioners can use data such as these to identify students for early intervention and provide them with necessary instruction to accelerate their progress. Future research is needed to examine how this process of identifying students (i.e., who are at risk for failing outcome assessments) and providing intervention affects student trajectories across grade levels as well as their ultimate performance on these measures.

Examining student trends in oral reading fluency across grade levels is complicated by the increasing difficulty of passages that are typically calibrated by grade level (e.g., first-grade students read passages at a first-grade level, second graders read second-grade passages). The variability in passage difficulty can make the interpretation of trends across more than a single year difficult (e.g., 40 words read correctly per minute on a first-grade passage may

not be equivalent to 40 words per minute on a third-grade passage)—hence, an issue when establishing aim lines that indicate levels of performance necessary to be on track for later success on high-stakes measures. Equating passage difficulty by considering the data in pieces that correspond to grade levels may be one way to address variability in text difficulty, and it may provide a more reliable fit. This approach has the added advantage of controlling for time off from school in the summer months. Our data were too limited to fully apply this analysis; however, we do suggest it as an area for continued research. Data with four or more collection points per grade level will allow each grade level to be considered a piece in the model, fully taking into account summer breaks and changes in passage difficulty. Information regarding the fit of the piecewise model could assist educators in making decisions about response to instruction and intervention services. For example, given the interruption of instruction over the summer, as well as the change in difficulty of oral reading fluency passages from grade to grade, more specific expectations for scores and slope within each grade level provided by a piecewise model could help teachers better identify students who are on track or in need of intervention.

Additional research information could also be gained by vertically scaling oral reading fluency measures across grade levels and examining student growth on these measures in relation to reading outcome measures. Furthermore, the scaling would make it possible to better examine student growth across grade levels. For example, in the absence of scaling, when a student achieves a score of 50 words per minute at the end of first grade and a score of 40 words per minute at the beginning of second grade, we do not know whether the drop is due to the increased difficulty of the second-grade text or to an actual drop in reading fluency, possibly from lack of instruction or practice over the summer. Similarly, the use of a fixed text level (e.g., second-grade-level text for all students) could provide more consistency in measurement over time and thus assist researchers in further examining student growth and the effects of text difficulty on student oral reading fluency (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Further research in these areas is necessary.

Acknowledgments

References

Baker SK, Smolkowski K, Katz R, Hank F, Seeley JR, Kame'enui EJ, et al. Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. School Psychology Review. 2008; 37:18–37.

Crawford L, Tindal G, Stieber S. Using oral reading rate to predict student performance on statewide achievement tests. Educational Assessment. 2001; 7:303–323.

CTB/McGraw-Hill. TerraNova Second Edition: California Achievement Tests technical report. Monterey, CA: Author; 2003.

Deno SL. Curriculum-based measurement: The emerging alternative. Exceptional Children. 1985; 18:19–32.

Deno SL. Developments in curriculum-based measurement. Journal of Special Education. 2003; 37:184–192.

Florida Statute, Section 1008.25. 2002.

Fuchs LS, Fuchs D, Hosp MK, Jenkins JR. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. Scientific Studies of Reading. 2001; 5:239–256.

Fuchs LS, Fuchs D, Maxwell L. The validity of informal reading comprehension measures. Remedial and Special Education. 1988; 9(2):20–29.

Good, RH., III; Kaminski, RA., editors. Dynamic Indicators of Basic Early Literacy Skills. 6th ed.. Eugene: University of Oregon, Institute for the Development of Educational Achievement; 2002.

Good, RH., III; Kaminski, RA.; Shinn, M.; Bratten, J.; Shinn, M.; Laimon, L., et al. Technical adequacy and decision making utility of DIBELS (Technical Report No. 7). Eugene: University of Oregon; 2001.

Good RH III, Simmons DC, Kame'enui EJ. The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. Scientific Studies of Reading. 2001; 5:257–288.

Harcourt Educational Measurement. Stanford Achievement Test. 10th ed.. San Antonio, TX: Harcourt Assessment; 2003.

Hasbrouck J, Tindal GA. Oral reading fluency norms: A valuable assessment tool for reading teachers. The Reading Teacher. 2006; 59:636–644.

Hintze JM, Silberglitt B. A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. School Psychology Review. 2005; 34:372–386.

Hoover, H.; Dunbar, S.; Frisbie, D. Iowa Test of Basic Skills. Rolling Meadows, IL: Riverside; 2005.

Hosp MK, Fuchs LS. Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? School Psychology Review. 2005; 34:9–26.

Jenkins JR, Fuchs LS, van den Broek P, Espin CL, Deno SL. Sources of individual differences in reading comprehension and reading fluency. Journal of Educational Psychology. 2003; 95:719–729.

Klein JR, Jimerson SR. Examining ethnic, gender, language, and socioeconomic bias in oral reading fluency scores among Caucasian and Hispanic students. School Psychology Quarterly. 2005; 20:23–50.

Lee, J. Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into the national and state reading and math outcome trends. Cambridge, MA: Civil Rights Project at Harvard University; 2006.

McGlinchey MT, Hixson MD. Using curriculum based measurement to predict performance on state assessments in reading. School Psychology Review. 2004; 33:193–203.

Muthén, LK.; Muthén, BO. Mplus user's guide. 4th ed.. Los Angeles: Author; 2006.

Newman DA. Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. Organizational Research Methods. 2003; 6:328–362.

No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425. 2002.

Riedel BW. The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. Reading Research Quarterly. 2007; 42:546–567.

Roehrig AD, Petscher Y, Nettles SM, Hudson RF, Torgesen JK. Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. Journal of School Psychology. 2008; 46:343–366. [PubMed: 19083363]

Schilling SG, Carlisle JF, Scott SE, Zeng J. Are fluency measures accurate predictors of reading achievement? The Elementary School Journal. 2007; 107:429–448.

Shinn MR, Good RH III, Knutson N, Tilly WD III, Collins VL. Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. School Psychology Review. 1992; 21:459–479.

Silberglitt B, Burns MK, Madyun NH, Lail KE. Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. Psychology in the Schools. 2006; 43:527–536.

Silberglitt B, Hintze J. Formative assessment using CMB-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. Journal of Psychoeducational Assessment. 2005; 23:304–325.

Singer, JD.; Willett, JB. Applied longitudinal data analysis: Modeling change and event occurrence. Oxford, UK: Oxford University Press; 2003.

Snow, CE.; Burns, MS.; Griffin, P. Preventing reading difficulties in young children. Washington, DC: National Academic Press; 1998.

Spear-Swerling L. Children's reading comprehension and oral reading fluency in easy text. Reading and Writing. 2006; 19:199–220.

Stage SA, Jacobsen MD. Predicting student success on state-mandated performance-based assessment using oral reading fluency. School Psychology Review. 2001; 30:407–419.

Texas Education Agency. Texas Assessment of Knowledge and Skills. Austin, TX: Author; 2004.

Texas Education Agency. Testing and accountability. 2006. from http://www.tea.state.tx.us/index.aspx?id=3426&menu_id=660&menu_id2=795

Vaughn, S.; Linan-Thompson, S.; Woodruff, AL.; Murray, CS.; Wanzek, J.; Scammaca, N., et al. Effects of professional development on improving at-risk students' performance in reading. In: Greenwood, CR.; Kratochwill, TR.; Clements, M., editors. Schoolwide prevention models: Lessons learned in elementary schools. New York: Guilford Press; 2008. p. 115-142.

Wiley HI, Deno SL. Oral reading and maze measures as predictors of success for English learners on a state standards assessment. Remedial and Special Education. 2005; 26:207–214.

Wood DE. Modeling the relationship between oral reading fluency and performance on a statewide reading test. Educational Assessment. 2006; 11:85–104.

Zimmerman BI, Dibenedetto MK. Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. Psychology in the Schools. 2008; 45:206–216.

Biographies

Jeanne Wanzek, PhD, is an assistant professor in special education at Florida State University, and she is on the research faculty at the Florida Center for Reading Research. Her research interests include learning disabilities, reading, effective instruction, and response to intervention.

Greg Roberts, PhD, is an associate director of the Meadows Center for Preventing Educational Risk and the director of the Vaughn Gross Center, both at the University of Texas at Austin. His interests include program evaluation, measurement, and response to intervention.

Sylvia Linan-Thompson, PhD, is an associate professor in the Department of Special Education and a fellow in the Cissy McDaniel Parker Fellow Fund at the University of Texas at Austin, as well as an associate director of the National Research and Development Center on English Language Learners.

Sharon Vaughn, PhD, holds the H. E. Hartfelder/Southland Corp. Regents Chair in Human Development. She is the executive director of the Meadows Center for Preventing Educational Risk. She is the author of numerous books and research articles that address the reading and social outcomes of students with learning difficulties. She is currently the principal investigator or co–principal investigator on several research grants (from the Institute for Education Science, the National Institute for Child Health and Human Development, and the Office of Special Education Programs) investigating effective interventions for students with reading difficulties and students who are English-language learners.

Althea L. Woodruff, PhD, is the primary reading specialist in Del Valle Independent School District outside of Austin, Texas, and a lecturer at the University of Texas at Austin.

Christy S. Murray, MA, currently serves as deputy director of the National Center on Instruction's Special Education Strand.
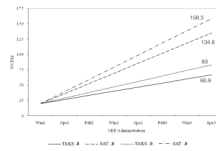
Figure 1.
Estimated Aim Lines for a Probability of .8 and .9 of Passing the *Texas Assessment of Knowledge and Skills* and the *Stanford Achievement Test*
Note: First-grade winter oral reading fluency (ORF) was fixed at 20 words correct per minute (WCPM).

Table 1

Demographics of Sample

Variable	n (%)
Gender	
Male	231 (50.1)
Female	230 (49.9)
Ethnicity	
Hispanic	303 (65.7)
White	61 (13.2)
African American	88 (19.1)
Other	9 (2.0)
Free or reduced-price lunch	343 (74.4)
Disability	
Learning disability	2 (0.4)
Speech/language	19 (4.1)
Other health impaired	3 (0.7)
Limited-English proficient	64 (13.9)

Table 2

Oral Reading Fluency Estimates by Time

| Assessment Period | *n* | Words Correct per Minute | | | | *r* | |
		MIN	MAX	*M*	*SD*	TAKS	SAT-10
First grade							
Winter	268	0	143	30.04	25.55	.44	.54
Spring	270	8	143	52.19	28.75	.51	.64
Second grade							
Fall	307	2	165	46.02	25.68	.53	.61
Winter	333	8	198	77.38	32.45	.58	.66
Spring	347	16	188	88.92	32.02	.57	.68
Third grade							
Fall	404	16	183	73.85	28.37	.58	.68
Winter	437	14	185	88.25	32.43	.60	.70
Spring	461	14	190	102.22	32.01	.60	.69

Note: TAKS = *Texas Assessment of Knowledge and Skills*; SAT-10 = *Stanford Achievement Test*, 10th edition.

Table 3

Growth Factors on Oral Reading Fluency

Model	Intercept	Slope
Linear	16.31 (749.9)	11.33 (19.2)

Note: Variance estimates in parentheses.

Table 4

Logistic Results for *Texas Assessment of Knowledge and Skills* Analysis

Model	Threshold	β	
		Intercept 1	**Slope 1**
Linear	2.00[**] (.61)	.055[**] (.01)	.343[**] (.06)

Note: Standard errors in parentheses.

[**]*p* < .01.

Table 5

Logistic Results for *Stanford Achievement Test* Analysis

Model	Threshold	β	
		Intercept 1	**Intercept 2**
Linear	4.26[**] (.53)	.044[**] (.01)	.290[**] (.04)

Note: Standard errors in parentheses.

[**]$p < .01$.

Table 6

Probability of Passing Conditioned on Several Parameter Combinations

Intercept and Slope	Probability of Passing (%)	
	TAKS	SAT-10
10th percentile	.48	.06
25th percentile	.65	.10
50th percentile	.92	.29
75th percentile	.99	.68

Note: TAKS = *Texas Assessment of Knowledge and Skills*; SAT-10 = *Stanford Achievement Test*, 10th edition.