Published in final edited form as:

Learn Disabil Q. 2012 May ; 35(2): 76–89. doi:10.1177/0731948711432509.

The Validity of a Holistically Scored Retell Protocol for Determining the Reading Comprehension of Middle School Students

Deborah K. Reed¹, Sharon Vaughn², and Yaacov Petscher³

¹The University of Texas at El Paso, USA

²The University of Texas at Austin, USA

³The Florida State University, Tallahassee, USA

Abstract

In this study, the authors examined the validity of a holistically scored retell within a confirmatory factor analysis framework by comparing the fit of a three-factor model of reading with the data from a diverse sample of seventh and eighth graders. The final model demonstrated adequate fit, $\chi^2(32) = 97.316$; comparative fit index = .96; Tucker–Lewis index = .94; and root mean square error of approximation = .08. Retell’s chi-square difference, $\Delta\chi^2(1) = 16.652$, $p < .001$, and factor loading (.250, $p < .001$) were higher for the comprehension construct. Similarly, retell’s correlation to comprehension measures ($r = .155-.257$, $p < .01$) was stronger than its relationship to measures of fluency ($r = .158-.183$, $p < .01$) or word identification ($r = .132$, $p < .05$). However, retell had a large residual variance (.938) and low interrater reliability ($\kappa = .37$), suggesting that improvements to the instrument are needed. Despite overall latent differences, retell did not demonstrate differential item functioning.

Keywords

retell; reading comprehension; confirmatory factor analysis; middle school

Current approaches to reading intervention are dependent on easily and frequently administered assessments to inform instruction (Coyne, Kame’enui, Simmons, & Harn, 2004; Stecker & Fuchs, 2000). In Grades 1 through 5, oral reading fluency (ORF) measures are commonly used to detect problems and monitor reading progress by calculating a student’s rate and accuracy when reading aloud, often with a 1-min limit. The scores of elementary students have demonstrated high correlations to standardized measures of reading (Burke & Hagan-Burke, 2007; Spear-Swerling, 2006) as well as state criterion-referenced reading assessments (Wiley & Deno, 2005). Researchers therefore have theorized that ORF is indicative of students’ general reading ability, and those who do not read

© Hammill Institute on Disabilities 2012

Corresponding Author: Deborah K. Reed, The University of Texas at El Paso, 500 W. University Ave., Ed 702, El Paso, TX 79968, USA, dkreed@utep.edu.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health or Texas Education Agency

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>

quickly and accurately are the children who would profit from instructional intervention (Burke, Hagan-Burke, Kwok, & Parker, 2009; Madelaine & Wheldall, 2005).

Research with older students has indicated a lower correlation between ORF and reading comprehension (Schatschneider et al., 2004; Wiley & Deno, 2005). Moreover, rate and accuracy scores have shown a tendency to asymptote in the middle grades as the slope of students' growth in ORF decelerates so that the line mapping their progress flattens (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Stage & Jacobsen, 2001). This is also when the contribution of decoding to reading comprehension diminishes somewhat (Keenan, Betjemann, & Olson, 2008) and students have more highly developed compensatory strategies that lessen their reliance on word identification skills (Savage, 2006).

Research has suggested that supplementing ORF with retell might assist in identifying adolescents who read disfluently but with adequate understanding or, conversely, those who read fluently but with poor comprehension (Marcotte & Hintze, 2009). Word-level deficits are considered distinct from comprehension deficits in students with learning disabilities (LD; Lyon et al., 2001). If retell improves the ability to discriminate among adolescents with particular reading skill deficits, it would assist teachers in providing targeted interventions to students with LD.

Background and Significance

Retell, or free recall, is a frequent component of reading comprehension measures (Nilsson, 2008; Talbott, Lloyd, & Tankersley, 1994). After reading a passage either orally or silently, students are prompted to convey the gist of what they read, without referring back to the text. In comprehension research, *retelling*, *recalling*, *summarizing*, and *paraphrasing* are considered distinct skills that require different degrees of telling or transforming knowledge (Kintsch & van Dijk, 1978; Scardamalia & Bereiter, 1987). Table 1 provides a brief explanation of each skill and a sample instrument that incorporates "retell" in this way. Within studies examining retell as a measurement tool, however, these skills are treated interchangeably (Duffelmeyer & Duffelmeyer, 1987). Depending on the instrument or study, *retell* and *recall* could be used to elicit main ideas, summaries, or a thorough restatement of the passage.

Retell, in its many forms, is an appealing compliment to ORF because it does not add considerable time or expense to the assessment, can present a consistent probe of comprehension across passages that is reflective of typical classroom instruction, and meets criteria for social validity (Roberts, Good, & Corcoran, 2005). In a study with students aged 8 to 18 years, the retell task of an informal assessment was much less sensitive to decoding ability (as measured by word recognition of isolated real words and word attack of nonsense words) than other standardized measures of comprehension (Keenan et al., 2008). In addition, errors consistent with the meaning of the sentence or passage were more strongly related to fourth-graders' recall of important ideas than their reading accuracy scores (Kucer, 2009). Unfortunately, commonly available reading assessments with a retell task have not sufficiently established the technical adequacy of the retell component (Klesius & Homan, 1985; Nilsson, 2008). Therefore, this study sought to examine the contribution of retell to a theoretical model of reading for middle school students.

As a measurement study, rather than an intervention study, the framework derives from theories of how the construct(s) of reading are defined. It examines how performance on measured reading skills contributes to latent variables or theoretically defined components of reading. Extant research suggests various conceptions of *reading competence* as a single construct through Grade 3 (Burke & Hagan-Burke, 2007; Shinn et al., 1992) or as a composite of distinct constructs (i.e., decoding, fluency, comprehension) among older

students (Buly & Valencia, 2003; Jackson, 2005; Vukovic, Wilson, & Nash, 2004). Given the existing empirical support, the present study modeled reading competence of adolescents utilizing three *a priori* constructs (word identification, fluency, and comprehension) within a confirmatory factor analysis (CFA) framework to determine whether retell measures observable skills that predict reading comprehension. The specific instrument under examination was the retell component of the Texas Middle School Fluency Assessment (TMSFA; Texas Education Agency, University of Houston, & The University of Texas System, 2008).

Research Questions

Similar to the procedures used in previous research on retell as an assessment for the construct of comprehension among adolescents (e.g., Burke & Hagan-Burke, 2007; Shinn et al., 1992), the TMSFA retell protocol is administered after a student reads a passage under timed conditions. Based on the premise that immediate recall is one common element of reading comprehension measures, retell is intended to provide unique information on how well the student understood the passage at a literal level (Jackson, 2005).

The primary research questions addressed about the TMSFA retell were as follows:

Research Question 1: What is the factor structure of reading competence expressed in the data obtained from a large, heterogeneous sample of students in Grades 7 and 8?

Research Question 2: Is model fit improved by including only narrative retell, only expository retell, narrative and expository retells entered individually, or the average retell performance on narrative and expository passages combined?

Research Question 3: How and to what extent does retell contribute to comprehension? Does retell contribute to fluency and word identification?

Research Question 4: What are the patterns of associations (correlation, regression) between the TMSFA retell instrument and other standardized measures of reading?

Research Question 5: Is retell influenced by differences in primary language, ability level, or socioeconomic status over and above the effects of reading comprehension?

Method

Research Setting and Participants

This study relies on an extant database compiled by researchers at The University of Texas at Austin and the University of Houston. Participants were from seven middle schools in Texas. In all, 394 students were tested: 260 enrolled in Grade 7 and 134 in Grade 8. Among the sample, approximately 47% were females and 53% were males; 37% were African American students, 1% Asian, 47% Hispanic, 14% Caucasian; and 63% were classified as economically disadvantaged (based on free/reduced lunch status). Students represented a range of ability levels, with 13% receiving special education services, 16% classified as English language learners (ELLs), and 23% classified as having reading difficulties (based on scale scores on the state criterion-referenced reading assessment, the *Texas Assessment of Knowledge and Skills* [TAKS]).

After removing outliers (described in the “Results” section), the final sample consisted of 311 students, evenly divided between males and females. The racial/ethnic makeup did not change. The percentage of students in special education (12%) and the percentage of students classified as having reading difficulties (22%) were only slightly less than that in the original sample. There were, however, a greater number of ELLs (24%) and students

classified as economically disadvantaged (71%). The handling of missing data is addressed in the procedures.

Measures

Students were administered 10 reading assessments (all with established reliability and validity) conceptualized as measuring word identification, fluency, or comprehension. With the addition of an intelligence test (used as an indicator of ability level) and the TMSFA retell (the instrument under study), the total number of measures included in this study was 12. Descriptive statistics are provided in Tables 2 and 3, and all measures except the *Kaufman Brief Intelligence Test* (KBIT) are depicted in the baseline factor model (see Figure 1).

Procedures

Test administration—The assessments were administered at the school sites by research assistants who attended at least 6 hr of training. All assistants had to achieve 100% accuracy in the administration and scoring procedures, which could take 2 to 4 hr longer than the standard training time. For group-administered assessments, research assistants would bring together 10 to 100 students in a room (e.g., library, cafeteria). Students were seated in rows, facing forward, and provided with individual stimuli and pencils. One research assistant would read the directions from the assessment manual to the full group while 2 to 12 additional assistants (depending on group size) would monitor students throughout the room and ensure adherence to the procedures outlined in the assessment manuals.

For individually administered assessments, including the retell, research assistants would pull students from class one-at-a-time and take them to the testing room. The assistant would sit directly across from the student and follow the administration procedures in the assessment manual and/or preprinted on the examiner document. Stimuli were placed in front of students, typically inside plastic sheet protectors and held in binders. After testing a student, the research assistant tallied and recorded data on the examiner document(s).

The retell component of the TMSFA was administered after the 1-min reading of each passage in the *Passage Reading Fluency subtest* (an ORF instrument). The examiner concealed the text and delivered the prompt: “Tell me in your own words what this passage is mostly about.” If the student provided only the title or a single word, the examiner prompted again with “Tell me more.” This additional prompt was offered only one time. The examiner transcribed the student response as accurately as possible on the record sheet and scored the response using a rubric. Because each student read to a different location in the passage within the minute time limit for the preceding ORF portion, the holistic approach to scoring allowed the tester to evaluate a student’s grasp of the passage’s “gist” against the particular amount of information read. Each student, therefore, had an equal chance of achieving a high score, regardless of how many words were read. Scores from 0 = *no response* to 3 = *strong comprehension* were awarded based on accuracy, completeness, and coherence (see appendix for rubric with exemplar responses).

Handling of data—After each testing session, research assistants checked for completion of the student answer documents from group-administered assessments and the examiner documents from individually administered assessments. Packets with missing data were flagged for makeup testing. Due to absences and schedule restrictions, some students did not take all assessments. Approximately 8% of data from the final sample were missing, all from the KBIT. Missing data were handled in the analysis with direct maximum likelihood (ML) estimator, considered more efficient and unbiased than ad hoc methods (Allison, 2003; Schafer & Graham, 2002).

Students recorded their answers from group-administered assessments on teleforms, computer-readable documents that allowed for electronic scoring. Research assistants only checked these to ensure students had completely filled-in the bubbles. Individually administered assessments that required calculations of words/items missed, rate, or accuracy were double-checked for correctness by an assistant other than the one who administered the assessment(s).

Interrater reliability of retell scoring—Transcribed student retells were scored once by the original examiner and a second time by the researcher. Both scorers were trained in the use of the rubric. Observed interrater agreement was 66%, consistent with previous findings of holistic approaches (Nilsson, 2008). The estimate of interrater reliability for the TMSFA retell scores was then adjusted for the possibility of chance (Cohen, 1960). The resulting kappa ($\kappa = .47$) was interpreted as a moderate agreement (Landis & Koch, 1977). It is important to note that interrater reliability for the individual passages was the same. That is, the raw percentage agreement (66%) and the kappa statistic (.47) were found for Passage 1 scores, Passage 2 scores, and Passage 3 scores.

When computing the interrater agreement for the average of the three retells, it was necessary to first compute the intraindividual average for each rater across the three passages and, subsequently, estimate the agreement as a function of the intraindividual mean. After averaging the three retells, interrater agreement decreased to 63%. The resulting kappa ($\kappa = .37$) was interpreted as a fair agreement. Although the interrater reliability of the TMSFA rubric score was low to moderate, it was higher than the 59% agreement reported in a previous study that used quantitative scores from oral retells to determine the incremental and concurrent validity of a formative measure of reading comprehension (Marcotte & Hintze, 2009). Most commercially available instruments with a retell protocol utilize holistic/rubric scores, and few report the agreement among scorers (e.g., Beaver, 2006; Johns, 2008; Roe & Burns, 2007; Silvaroli & Wheelock, 2004; Woods & Moe, 2007). Therefore, proceeding with the analysis was deemed valuable, albeit cautionary, to contribute additional information about the adequacy of these approaches that might be used to plan instruction for students with LD.

Design and Analysis

CFA provided the overall framework for analysis. An integrated model building approach was used to address the research aims, with each analysis providing a foundation for subsequent models. CFA belongs to the class of structural equation models. Accordingly, it provides error-adjusted measures of latent constructs based on the covariance structures of observed variables, yielding more precise estimates of relevant factors than the observed values on which the analysis is based. Structural equation models, including CFA, also provide indices of model fit as a means of evaluating the degree to which the available data conform to the specified model. The comparative fit index (CFI) is frequently reported in CFA research, but sample sizes of more than 100 can inflate CFI (Brown, 2006). For this reason, other indices of comparative fit and parsimony correction were included in the evaluation of model fit. These indices included the Tucker–Lewis index (TLI; Tucker & Lewis, 1973) and root mean square error of approximation (RMSEA; Steiger & Lind, 1980). For this study, SPSS version 17.0 (SPSS Inc., 2009) was used to manage the data and calculate descriptive statistics. *Mplus* version 5.21 (L. Muthen & Muthen, 2009) was used to estimate confirmatory models.

Results

An examination of the variables' skewness and kurtosis revealed that TAKS scores had a leptokurtic distribution. Therefore, Mahalanobis distances, $\chi^2(14, N = 394) = 36.123, p < .001$, were evaluated for the variables of interest, and 83 cases exceeding the critical value were removed from the data set. This reduced the sample size to 311, which was still sufficient for the CFA because it met or exceeded the three cases:parameter ratio.

The initial analysis concerned the fit of the baseline model depicted in Figure 1. This did not converge in 50,000 iterations, so the dependent variables were redefined conceptually to identify the most theoretically supported measures for each of the three constructs. Based on the correlations among the measures (see Table 3), the *TMSFA Word Reading Fluency subtest* was removed from the model because it did not contribute unique information above what was contributed by the *Test of Word Reading Efficiency* (TOWRE) summary, and the *Test of Sentence Reading Efficiency* (TOSRE) was removed because it could not be clearly distinguished as a dependent variable for any one latent construct. In addition, the AIMSweb reading maze was moved within the model to be a measure of fluency.

This conceptually redefined model (see Figure 2) converged, $\chi^2 = 97.316(32)$; CFI = .958; TLI = .941; RMSEA = .081. CFIs and TLIs greater than .95 and RMSEAs less than .10 are considered indicative of good fitting models (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996). In addition, the standardized root mean square residual (SRMR) of the model was .039, which is close to the desired .0 value (Brown, 2006). Taking all indices of fit into consideration, the conceptually redefined three-factor model of reading competence was retained.

Model Fit by Passage Type

The next phase of analysis examined whether the fit could be improved by using one or more retell scores from individual passages rather than the average score. The three retell scores were derived from two passages labeled *expository* and one labeled *narrative*. When entered individually into the model, the retell score from the narrative passage was the best predictor with a moderate but significant factor loading on comprehension of .352 ($p < .001$) compared with the weak but significant factor loadings of the expository passages (.264 and .221; $p < .001$). The Akaike information criterion (AIC) and Bayesian information criterion (BIC) values were lower for the model with the narrative retell score alone than for the model that included all three retell scores. Therefore, this nested model comparison was evaluated with the chi-square difference test, which was significant, $\Delta\chi^2 = 134.261(19)$; $p < .001$.

However, entering the average across the three retell scores produced a more parsimonious model than entering the retell score from the narrative passage alone. Not only were the AIC and BIC values lower for the average retell score ($\Delta\text{AIC} = 58.275$; $\Delta\text{BIC} = 58.275$), but the RMSEA and SRMR values were also slightly lower (see Table 4). The relative fit indices, CFI and TLI, further confirm that entering the narrative retell score alone decreased model fit. Therefore, the average of the three retell scores was used for all subsequent analysis.

Contribution of Retell to Latent Constructs

With the retell score that produced the "best fitting" model identified, the contribution of retell to each of the three latent factors was evaluated through chi-square difference testing of nested models. The difference between the constrained versus the fully specified path coefficients to retell was significant for comprehension, $\Delta\chi^2 = 16.652(1)$, $p < .001$; fluency, $\Delta\chi^2 = 10.882(1)$, $p < .001$; and word identification, $\Delta\chi^2 = 7.84(1)$, $p < .005$. However, the chi-square difference and the factor loading (see Table 5) were greater for comprehension,

so the model depicted in Figure 2 was not revised. The average retell score remained as an indicator of comprehension only, suggesting that it is less indicative of students' word identification and fluency ability.

Patterns of Associations Among Measures and Constructs

The measures were, in general, moderately correlated with each other (see Table 3) with no observed multicollinearity. The strongest correlation was between Woodcock-Johnson III (WJ-III) Word Attack and WJ-III Letter-Word Identification ($r = .742$) with the weakest significant relationship at .132 between WJ-III Letter-Word Identification and average retell. Retell had the strongest relationship to WJ-III Passage Comprehension ($r = .208$) and TAKS ($r = .257$), the measures with the highest factor loadings on the comprehension construct (see Table 6). Average retell demonstrated the weakest correlations to the measures of word identification where only the correlation coefficient for the WJ-III Letter-Word Identification was significant. This is contrary to the pattern demonstrated by the TMSFA *Passage Reading Fluency subtest* and AIMSweb reading maze, which had the strongest relationships to assessments of word identification ($r = .550-.595$) and the weakest relationships to assessments of comprehension ($r = .430-.498$).

The relationships of the factors to the measures are provided in Table 6 and Figure 2. Retell was a weak but significant predictor of comprehension with a very high residual variance (depicted in the model in a small circle with an arrow pointing toward the measure), which indicates the variable does not function well as a measure of the construct. It is possible the weak interrater reliability is contributing to the poor functioning of the average retell scores in the model (Marcotte & Hintze, 2009). Nevertheless, retell was less indicative of comprehension ability than standardized measures (Group Reading Assessment and Diagnostic Evaluation [GRADE], WJ-III Passage Comprehension, and TAKS).

As expected, the constructs were all significantly correlated. The strongest relationship was between fluency and word identification ($r = .799$) with the weakest correlation at .640 between comprehension and fluency. Given that all measures except TAKS were timed, the correlations might be due to the pressure timed tests place on the speed with which words can be read or processed (Catts, Gillispie, Leonard, Kail, & Miller, 2002; Cutting & Scarborough, 2006).

Influence of Socioeconomic Status, Primary Language, and Ability Level

In the final phase of the analysis, students' socioeconomic status (SES), language proficiency, and ability level were entered into the model as covariates. Each covariate was treated as a dichotomous variable. For example, students who were receiving free/reduced-priced lunch were coded as "1," and all others were coded as "0" on the variable "SES." The KBIT composite scores were converted into a categorical indicator for ability level. Students whose standard score was ≥ 100 were coded as "1" for *above average*, and those whose score < 100 were coded as "0" for *below average*.

Given the smaller sample size, multigroup modeling with nested comparisons based on group (Bovaird, 2007; Mehta & Neale, 2005) was not used. Rather, a multiple indicators, multiple causes (MIMIC) model (Joreskog & Goldberger, 1975) was conducted by adding the aforementioned groups as covariates. MIMIC models with categorical indicators have demonstrated equivalence to differential item functioning (DIF) analysis and have the advantage of modeling a direct effect of the covariate on the latent factor (MacIntosh & Hashim, 2003; B. O. Muthen, Kao, & Burstein, 1991). The MIMIC testing followed a two-step approach. The first testing was for overall latent differences on the covariates. As Table 7 indicates, there were significant, small to moderate group differences on the latent factors

defining reading competence. This was particularly true with respect to comprehension performance where almost all groups demonstrated DIF.

This initial step of the MIMIC testing also included an examination of the modification indices to identify any specific observed indicator differences that should be tested in the second step. Although the model did not suggest average retell demonstrated DIF, retell was tested to be consistent with the research questions for this study. As shown in Table 8, there were no significant differences for any group on average retell performance. The lack of differences in factor means for retell indicate that the groups do not differ on intercepts.

Because there were no significant differences among groups on average retell scores, the differences observed on the comprehension construct cannot be attributed to students' retell performance. Students who were classified as economically disadvantaged, bilingual, ELLs, in special education, or below average on the KBIT performed significantly worse on the standardized measures of comprehension, whereas on the retell measure, students in and out of these categorical groups all had comparable scores.

Discussion

The present study addressed the need for more data on the technical adequacy of retell, administered within an ORF approach, as a significant and efficient measure of adolescents' reading competence. Although ORF measures have been shown to be reflective of the overall reading ability of students in Grades 1 through 5 (Burke et al., 2009; Fuchs et al., 2001; Spear-Swerling, 2006), research suggests that ORF may not be a comprehensive index of reading comprehension for adolescents. Specifically, the correlation between ORF and reading comprehension is less robust for students above Grade 5 (Schatschneider et al., 2004; Wiley & Deno, 2005), an age at which rate and accuracy scores begin to asymptote (Fuchs et al., 2001; Stage & Jacobsen, 2001). The addition of a retell task to ORF might provide a solution to efficient progress monitoring of reading comprehension for middle school students; however, existing measures have not been validated for this purpose. By including the retell component of the TMSFA in a CFA, the study sought to determine whether and to what extent the TMSFA retell scores were indicative of seventh- and eighth-grade students' reading comprehension ability.

Overall, findings suggest that the TMSFA retell protocol currently lacks the technical adequacy to be a valid and reliable measure of reading comprehension for this age group. If an assessment is no more valid than it is reliable, the primary concern is that the retell data used in the models tested here were based on holistic scores with a low interrater reliability. To make valid interpretations of students' scores, a more reliable method of scoring retell responses is needed. The next sections will summarize the findings and their implications with respect to each research question.

Factor Structure of Reading Competence

With a normally distributed sample of 311 seventh and eighth graders, a three-factor model of reading competence converged and had adequate fit. Unlike previous studies (Buly & Valencia, 2003; Hock et al., 2009; Valencia & Buly, 2004), scores from adolescents at a range of ability levels, including typically achieving readers, were included in the analysis. This increases confidence in retaining the model; however, it is possible that the fit might have been improved if students were grouped by ability level or if the model included vocabulary as a fourth latent construct.

The decision to test only a three-factor model of reading competence was theoretically based on the fact that vocabulary and comprehension have demonstrated such a strong,

bidirectional relationship as to make distinctions between the two abilities difficult (Leach, Scarborough, & Rescorla, 2003; Snow, 2002; Wagner, Muse, & Tannenbaum, 2007). However, a principal components analysis conducted with ninth graders suggests that vocabulary may be a distinct domain of reading that discriminates a small percentage of students (4%) who have adequate comprehension but low vocabulary knowledge (Hock et al., 2009). Because vocabulary was not tested as a separate latent construct, the results from the present study do not allow for conclusions as to whether a four-factor model of reading competence would be more parsimonious for this age group.

Retell by Passage Type

Previous research found students recall less information from expository than from narrative passages (Best, Floyd, & McNamara, 2008). Narrative TMSFA retell scores in the present model were better predictors of comprehension than expository retell scores or the combination of all three scores, but average retell scores produced the most parsimonious model. Unlike the study by Best and colleagues (2008), the TMSFA retell scores are based on holistic evaluations of accuracy, completeness, and quality rather than on quantitative counts of predetermined idea units. In addition, the TMSFA passages are not clearly distinguishable as expository but might be more accurately described as informational narratives with a descriptive or biographical structure. Passages with more challenging implicit causal relations (Richgels, McGee, Lomax, & Sheard, 1987; Zinar, 1990) might have resulted in student retells with different effects on existing model parameters.

Although not reflective of more traditional or complex expository text, the TMSFA passages labeled “expository” include more facts and require more understanding of history, geography, and culture than the “narrative” passages. Therefore, it is possible that retell performance on the expository passages was influenced by students’ prior knowledge. Background knowledge was the second largest contributor to ninth graders’ comprehension (Cromley & Azevedo, 2007) and the best predictor of third graders’ expository retell performance (Best et al., 2008).

Extent to Which Retell Contributes to Latent Constructs

The path coefficients to TMSFA retell were significant for all three constructs, but the data from this study suggest that retell was most indicative of students’ comprehension ability and least indicative of word identification ability. Although average retell was only weakly correlated to the comprehension construct, the model demonstrated adequate fit, $\chi^2 = 97.316(32)$; CFI = .958; TLI = .941; RMSEA = .081, with retell included. In contrast, the *TMSFA Word Reading Fluency subtest* and TOSRE had to be removed before the model would converge. TMSFA retell did, however, have a large residual variance (.938), suggesting that it did not function well as a measure of comprehension as currently administered. Given the weak interrater reliability of the holistic scoring mechanism used in the TMSFA retell component and its administration following a timed fluency test, it is possible that the data on the contribution of retell presented here are more a function of these particular scores than of the actual validity of “retell” as a dependent variable.

Patterns of Associations Among Measures and Constructs

Similar to its relationship with the latent constructs, TMSFA retell also demonstrated weak correlations to the other measures of comprehension. These findings were in contrast to previous studies that demonstrated moderate to high correlations between retell and measures of comprehension (e.g., Fuchs, Fuchs, & Maxwell, 1988; Marcotte & Hintze, 2009). However, retell in this study bore the strongest relationship to the best functioning predictors of comprehension (i.e., WJ-III Passage Comprehension and TAKS) and almost no relationship to measures of word identification.

This makes the findings noteworthy for two reasons. First, the *TMSFA Passage Reading Fluency subtest* and AIMSweb reading maze, demonstrated a pattern contrary to that of TMSFA retell: The two fluency measures were more associated with word identification than comprehension. Therefore, the TMSFA retell has the potential to provide a different portrait of students' reading competence than that depicted by the ORF component of the same assessment. It should be stressed, however, that the ORF and maze measures correlated more highly with comprehension than TMSFA retell and that the reliability of the TMSFA retell measure is low to moderate.

Despite these cautions, the results provide a second important finding about the measures used in the model. The best predictors of comprehension were moderately correlated to measures of word identification, whereas there was only one weakly significant correlation between TMSFA retell and the WJ-III Letter–Word Identification. This seems to confirm the finding of Keenan et al. (2008) that retell was less sensitive to decoding ability than other standardized measures of reading comprehension.

Word-level deficits are the primary focus of early identification and prevention of reading disabilities but are considered distinct from comprehension deficits in students with LD (Lyon et al., 2001). About 25% of elementary children provided with intensive, explicit, and systematic instruction in auditory discrimination, phonics, and word identification still demonstrated persistent difficulties beyond the elementary years (Torgesen et al., 2001; Velluntino et al., 1996). Word-level reading disabilities, or dyslexia (Fletcher, Lyon, Fuchs, & Barnes, 2007), can make it difficult to accurately assess reading comprehension, particularly when the instruments are timed (Catts et al., 2002). If retell can provide unique information on the reading skills of adolescents in special education, teachers and reading interventionists would be better able to plan targeted instruction in the appropriate domains.

Covariates

A number of studies have reported that students with LD recall less information than students without identified disabilities (e.g., Carlisle, 1999; Curran, Kintsch, & Hedberg, 1996; Williams, 1991). Others have cautioned that socioeconomic and cultural–linguistic differences might influence performance on comprehension tasks, such as retell, that require oral language processing (Snyder, Caccamise, & Wise, 2005). Despite overall latent differences, average TMSFA retell scores did not exhibit any significant differences by group. Students who were economically disadvantaged, ELLs, enrolled in special education, or had lower academic ability had significantly poorer performance on standardized comprehension measures but not on retell.

Ability differences have been found in other research when utilizing quantitative counts of predetermined idea units because lower ability students and those with LD did not offer as much information unless specifically prompted to do so (Gambrell, Koskinen, & Kapinus, 1991; Gambrell & Jawitz, 1993; Zinar, 1990). In the current study, the subjective nature of the holistic approach to evaluating TMSFA retells may have facilitated taking certain student characteristics into account. Rather than basing the score on a straightforward count of idea units, raters could draw on other impressions of quality and completeness that might have accommodated for what would otherwise be considered an insufficient response. However, this cannot be determined from available data.

Limitations

In addition to the overarching limitation imposed by the weak interrater reliability of the scoring mechanism, there are several other characteristics of the study to consider when interpreting the results. First, nearly all the data were derived from assessments administered

under timed conditions, which place added pressure on speed of reading (Cutting & Scarborough, 2006; Jackson, 2005). This created a conceptual difficulty in specifying the baseline model in that it was difficult to separate measures of reading accuracy (the word identification construct) from measures of reading rate (the fluency construct). Hence, there were dependent variables that had to be removed or moved within the model. What is not known is how much the timed nature of the assessments might be influencing the correlations among the measures and constructs.

Although the retell itself was not timed, it was assessed after the ORF component. This is consistent with the procedures in previous studies (e.g., Burke & Hagan-Burke, 2007; Shinn et al., 1992). However, other studies provided students an unlimited amount of time to read and/or allowed students to read the passages silently (e.g., Best et al., 2008; van den Broek, Tzeng, Risdien, Trabasso, & Basche, 2001; Walczyk & Raska, 1992). The outcomes under those conditions could be different from what is reported for the retell protocol used in this study.

A second limitation concerns the issue of social validity, which is one of the rationales for adding a retell component to an ORF assessment (Roberts et al., 2005). This is related to Messick's (1989) validity model that includes considerations of relevance and utility with construct validity. In other words, an instrument may demonstrate technically that it assesses the construct it purports to measure (construct validity), but if the individuals who use it still do not believe it measures that construct, the instrument lacks validity on its face. Unfortunately, no social validity data were available for analysis in this study to determine teachers' confidence in the retell or ORF scores as indicators of students' reading competence.

Finally, it should be noted that the lack of DIF in the TMSFA retell scores might be due to the limited variability and the use of dichotomous scores for the covariates. Although true DIF might not be as strongly detected because of the small range (0–3) in the TMSFA retell scores, there was no clear pattern of some groups' scores being closer to 0 or to 3.

Recommendations for Future Research

Questions still remain about the reliability and validity of retell as a measure of adolescents' reading comprehension that would compliment traditional ORF instruments. In its current state, the retell component of the TMSFA was not technically adequate, but it might reflect the holistic approach to scoring responses or the context of the TMSFA rather than the utility of retell in general. Although quantitative procedures for scoring responses are more common in studies of retell (Reed & Vaughn, in press), rubric and other subjective means of scoring are far more prevalent in instruments commercially available (e.g., Beaver, 2006; Johns, 2008; Roe & Burns, 2007; Silvaroli & Wheelock, 2004; Woods & Moe, 2007).

Future research could replicate the CFA with retells scored using a more reliable mechanism. A quantitative mechanism might improve the functioning of retell in the model and reveal stronger correlations to the other measures. In addition, future research might compare retell within ORF with retell administered after reading completed silently and/or for an unlimited amount of time. Replicating the analysis by including retell condition as a potential covariate would yield important information about the optimal administration procedures.

To address questions of social validity (Kazdin, 1977; Wolf, 1978), future research should gather information on teachers' perceptions and use of standardized comprehension test results, ORF scores, and retell ratings. It might not be worthwhile to pursue improvements to retell tasks if teachers already had confidence in and were relying on data from more

psychometrically sound measures. Conversely, if teachers disregard data from instruments with high technical adequacy in favor of a retell, it would be critical to advance this line of research and ensure valid information was available.

As with any reading assessment, the value of a retell task lies in what it can reveal about a students' abilities that would be useful for planning and evaluating instruction. Even if subsequent studies can substantiate that a reliably scored retell is a strong predictor of comprehension that is trusted by teachers, more guidance is needed in how to use the retell scores to make instructional decisions. It is not clear how teachers would interpret retell responses to group students, plan targeted skills instruction beyond retelling information, or connect to inferential comprehension. In summary, there is a great deal yet to learn about the utility of retell assessments.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Grant P50 HD052117 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and contract No 2356-27235 from the Texas Education Agency.

References

- Allison PD. Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*. 2003; 112:545–557. [PubMed: 14674868]
- Beaver, JM. Developmental reading assessment, Grades K-3. 2. Parsippany, NJ: Pearson Education; 2006.
- Best RM, Floyd RG, McNamara DS. Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*. 2008; 29:137–164.
- Bovaird, JA. Multilevel structural equation models for contextual factors. In: Little, TA.; Bovaird, JA., editors. *Modeling contextual effects in longitudinal studies*. Mahwah, NJ: Lawrence Erlbaum; 2007. p. 149-182.
- Brown, TA. *Confirmatory factor analysis for applied research: Methodology in the social sciences*. New York, NY: Guilford; 2006.
- Buly, MR.; Valencia, S. Meeting the needs of ailing readers: Cautions and considerations for state policy. Seattle: University of Washington, Center for the Study of Teaching and Policy; 2003. Retrieved from <http://depts.washington.edu/ctpmail/PDFs/Reading-MRBSV-04-2003.pdf>
- Burke MD, Hagan-Burke S. Concurrent criterion-related validity of early literacy indicators for middle of first grade. *Assessment for effective Intervention*. 2007; 32:66–77.
- Burke MD, Hagan-Burke S, Kwok O, Parker R. Predictive validity of early literacy indicators from the middle of kindergarten to second grade. *Journal of Special Education*. 2009; 42:209–226.
- Carlisle JF. Free recall as a test of reading comprehension for students with learning disabilities. *Learning Disability Quarterly*. 1999; 22:11–22.
- Catts HW, Gillispie M, Leonard LB, Kail RV, Miller CA. The role of speed processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities*. 2002; 35:510–525.
- Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
- Coyne MD, Kame'enui EJ, Simmons DC, Harn B. Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities*. 2004; 37:90–105. [PubMed: 15493232]
- Cromley JG, Azevedo R. Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*. 2007; 99:311–325.

- Curran CE, Kintsch E, Hedberg N. Learning-disabled adolescents' comprehension of naturalistic narratives. *Journal of Educational Psychology*. 1996; 88:494–507.
- Cutting LE, Scarborough HS. Prediction of reading comprehension: Relative contribution of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*. 2006; 10:277–299.
- Duffelmeyer FA, Duffelmeyer BB. Main idea questions on informal reading inventories. *Reading Teacher*. 1987; 41:162–166.
- Fletcher, JM.; Lyon, GR.; Fuchs, LS.; Barnes, MA. *Learning disabilities: From identification to intervention*. New York, NY: Guilford; 2007.
- 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; 9:20–29.
- Gambrell LB, Jawitz PB. Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly*. 1993; 28:265–276.
- Gambrell LB, Koskinen PS, Kapinus BA. Retelling and the reading comprehension of proficient and less-proficient readers. *Journal of Educational Research*. 1991; 84:356–362.
- Good, RH.; Kaminski, RA. *Vital indicators of progress (VIP)*. Dallas, TX: Voyager Expanded Learning; 2002.
- Hock MF, Brasseur IF, Deshler DD, Catts HW, Marquis JG, Mark CA, Stribling JW. What is the reading component skill profile of adolescent struggling readers in urban schools? *Learning Disability Quarterly*. 2009; 32:21–38.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structural analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Jackson NE. Are university students' component reading skills related to their text comprehension and academic achievement? *Learning and Individual Differences*. 2005; 15:113–139.
- Johns, JL. *Basic reading inventory: Pre-primer through grade twelve and early literacy assessments*. 10. Dubuque, IA: Kendall/Hunt; 2008.
- Joreskog KG, Goldberger AS. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*. 1975; 70:631–639.
- Kazdin AE. Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*. 1977; 1:427–451.
- Keenan JM, Betjemann RS, Olson RK. Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*. 2008; 12:281–300.
- Kintsch W, van Dijk TA. Towards a model of text comprehension and production. *Psychological Review*. 1978; 85:363–349.
- Klesius JP, Homan SP. A validity and reliability update on the informal reading inventory with suggestions for improvement. *Journal of Learning Disabilities*. 1985; 18:71–76. [PubMed: 3882869]
- Kucer SB. Examining the relationship between text processing and text comprehension in fourth-grade readers. *Reading Psychology*. 2009; 30:340–358.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
- Leach JM, Scarborough HS, Rescorla L. Late-emerging reading disabilities. *Journal of Educational Psychology*. 2003; 95(2):211–224.
- Leslie, L.; Caldwell, J. *Qualitative reading inventory*. 5. Boston, MA: Pearson Education; 2011.
- Lyon, GR.; Fletcher, JM.; Shaywitz, SE.; Shaywitz, BA.; Torgesen, JK.; Wood, FB.; Olson, R. Rethinking learning disabilities. In: Finn, CE., Jr; Rotherham, RAJ.; Hokanson, CR., Jr, editors. *Rethinking special education for a new century*. Washington, DC: Thomas B. Fordham Foundation; 2001. p. 259-287.

- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Method*. 1996; 1:130–149.
- MacIntosh R, Hashim S. Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. *Applied Psychological Measurement*. 2003; 27:372–379.
- Madelaine A, Wheldall K. Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development, and Education*. 2005; 52:33–42.
- Marcotte AM, Hintze JM. Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology*. 2009; 47:315–335. [PubMed: 19712779]
- Mehta P, Neale M. People are variables too: Multilevel structural equation models. *Psychological Methods*. 2005; 10:259–284. [PubMed: 16221028]
- Messick, S. Validity. In: Linn, RL., editor. *Educational measurement*. New York, NY: Macmillan; 1989. p. 13–104.
- Muthen BO, Kao C, Burstein L. Instructional sensitivity in mathematics achievement test items: Applications of a new IRT-based detection technique. *Journal of Educational Measurement*. 1991; 28:1–22.
- Muthen, L.; Muthen, B. *Mplus (version 5.21)* [Computer Software]. Los Angeles, CA: Author; 2009.
- Nilsson NL. A critical analysis of eight informal reading inventories. *Reading Teacher*. 2008; 61:526–536.
- Reed DK, Vaughn S. Retell as an indicator of reading comprehension. *Scientific Studies of Reading*. in press. 10.1080/10888438.2010.538780
- Richgels DJ, McGee LM, Lomax RG, Sheard C. Awareness of text structures: Effects on recall of expository text. *Reading Research Quarterly*. 1987; 22:177–196.
- Roberts G, Good R, Corcoran S. Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly*. 2005; 20:304–317.
- Roe, BD.; Burns, PC. *Informal reading inventory*. Boston, MA: Houghton Mifflin; 2007.
- Savage R. Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading*. 2006; 10:143–164.
- Scardamalia, M.; Bereiter, C. Knowledge telling and knowledge transforming in written composition. In: Rosenberg, S., editor. *Advances in applied psycholinguistics: Disorders of first-language development*. Vol. 1. New York, NY: Cambridge University Press; 1987. p. 142–175.
- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2002; 7:147–177. [PubMed: 12090408]
- Schatschneider, C.; Buck, J.; Wagner, R.; Hassler, L.; Hecht, S.; Powell-Smith, K. Technical Report No 5. Tallahassee: Florida Center for Reading Research; 2004. A multivariate study of individual differences in performance on the reading portion of the Florida Comprehensive Assessment Test: A brief report.
- Shinn MR, Good R, Knutson N, Tilly WD, Collins VL. Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*. 1992; 21:459–479.
- Silvaroli, NJ.; Wheelock, WH. *Classroom reading inventory*. 10. Boston, MA: McGraw-Hill; 2004.
- Snow, C. *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Education; 2002.
- Snyder L, Caccamise D, Wise B. The assessment of reading comprehension: Considerations and cautions. *Topics in Language Disorders*. 2005; 25:33–50.
- Spear-Swerling L. Children's reading comprehension and oral reading fluency in easy text. *Reading and Writing: An Interdisciplinary Journal*. 2006; 19:199–220.
- SPSS Inc.. *SPSS PASW Statistics (version 17.0)* [Computer Software]. Chicago, IL: Author; 2009.
- Stage SA, Jacobsen MD. Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*. 2001; 30:407–419.

- Stecker PM, Fuchs LS. Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disability Research & Practice*. 2000; 15:128–134.
- Steiger, JH.; Lind, JM. Statistically based tests for the number of common factors; Paper presented at the meeting of the Psychometric Society; Iowa City, IA. 1980 May.
- Talbott E, Lloyd JW, Tankersley M. Effects of reading comprehension interventions for students with learning disabilities. *Learning Disability Quarterly*. 1994; 17:223–232.
- Texas Education Agency, University of Houston, & The University of Texas System.. Texas middle school fluency assessment, version 1.0. Austin, TX: Author; 2008.
- Torgesen JK, Alexander AW, Wagner RK, Rashotte CA, Voeller KKS, Conway T. Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*. 2001; 34:33–58. [PubMed: 15497271]
- Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. 1973; 38:1–10.
- Valencia SW, Buly MR. Behind test scores: What struggling readers really need. *Reading Teacher*. 2004; 57:520–531.
- van den Broek P, Tzeng Y, Risden K, Trabasso T, Basche P. Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology*. 2001; 93:521–592.
- Velluntino F, Scanlon D, Sipay E, Small S, Pratt A, Chen R, Denckla MYY. Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*. 1996; 88:601–638.
- Vukovic RK, Wilson AM, Nash KK. Naming speed deficits in adults with reading disabilities: A test of the double-deficit hypothesis. *Journal of Learning Disabilities*. 2004; 37:440–450. [PubMed: 15460350]
- Wagner, RK.; Muse, AE.; Tannenbaum, KR. Promising avenues for better understanding implications of vocabulary development for reading comprehension. In: Wagner, RK.; Muse, AE.; Tannenbaum, KR., editors. *Vocabulary acquisition: Implications for reading comprehension*. New York, NY: Guilford; 2007. p. 276-291.
- Walczyk JJ, Raska LJ. The relation between low- and high-level reading skills in children. *Contemporary Educational Psychology*. 1992; 17:38–46.
- 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.
- Williams JP. Comprehension by learning-disabled and nondisabled adolescents of personal/social problems presented in text. *American Journal of Psychology*. 1991; 104:563–586. [PubMed: 1793127]
- Wolf MM. Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*. 1978; 11:203–214. [PubMed: 16795590]
- Woods, ML.; Moe, AJ. Analytical reading inventory: Comprehensive standards-based assessment for all students, including gifted and remedial. 8. Upper Saddle River, NJ: Pearson/Prentice Hall; 2007.
- Zinar S. Fifth-graders' recall of propositional content and causal relationships from expository prose. *Journal of Reading Behavior*. 1990; 22:181–199.

Appendix. Texas Middle School Fluency Assessment, Version 1.0, Retell Component

Retell Rubric

Score

0 : No response.

1 : Weak comprehension

- Contains information that was not in the passage or that misinterprets information in the passage.
- Consists of an isolated fact or name.
- Is rambling or incoherent.
- No apparent understanding of the main idea.

Example response: “People living underground in the desert.”

Explanation: This student provides only an isolated fact from the passage. There is no indication that she understood the significance of people living underground or how the conditions necessitated certain building features.

2 : Partial comprehension

- Contains some minor inaccuracies.
- Contains basic information from the passage, but not the most important point.
- Is not concisely stated and/or does not reflect the relationship among the ideas.
- Partial understanding of the main idea.

Example response: “The Australians. The way they made their homes and why they made their homes.”

Explanation: This student’s response is accurate and contains more than an isolated fact. However, she has not made an attempt to provide the important information about the homes being underground or to relate this to the desert conditions. It is not clear in the last statement whether the student is referring to the temperatures and dust or to the opal mine. The response is, perhaps, too concise to determine whether the student had adequate understanding.

3 : Strong comprehension

- Accurately reports information in the passage.
- Contains the most important point from the passage.
- Is coherently stated and reflects the proper relationship among the ideas.
- Response reflects a clear understanding of the main idea.

Example response: “It’s about how underground homes are carved in the Earth because it’s in a desert in Southern Australia. There are few windows in the homes. The temperatures get as high as 120 degrees in the summer. But it’s only 32 degrees in winter.”

Explanation: The student’s response accurately reports the most important information about the town in Australia and why the homes are built underground. He knows details about the geographic location, the homes, and the temperatures during the different seasons. Although the response is not as concisely stated as might be preferred, it does reflect the proper relationship among the ideas. It seems this student had a strong understanding of what he read.

Figure 1.

Baseline model

Note: WJ-III = Woodcock-Johnson III; L-W Identification = Letter-Word Identification; TMSFA Texas Middle School Fluency Assessment; TOSRE *Test of Sentence Reading Efficiency*; TOWRE = *Test of Word Reading Efficiency*; GRADE = Group Reading Assessment and Diagnostic Evaluation; TAKS = Texas Assessment of Academic Skills.

Figure 2.

Final model

Note: WJ-III = Woodcock-Johnson III; L-W Identification = Letter-Word Identification; TOWRE = *Test of Word Reading Efficiency*; TMSFA = Texas Middle School Fluency Assessment; GRADE = Group Reading Assessment and Diagnostic Evaluation; TAKS = Texas Assessment of Academic Skills.

Table 1

Potential Skills Targeted by a Retell Protocol

Skill	Definition	Sample instrument
Retell	Telling back the passage completely and in sequence	QRI
Recall	Stating particular information from memory that matches the text in wording but not in sequence	VIP
Summarize	Providing the main ideas and significant details of the passage in a shortened, organized form	DRA
Paraphrase	Restating particular content in different words, may involve an implied main idea statement	TMSFA

Note: QRI = Qualitative Reading Inventory-5 (Leslie & Caldwell, 2011); VIP = Vital Indicators of Progress (Good & Kaminski, 2002); DRA = Developmental Reading Assessment (Beaver, 2006); TMSFA = Texas Middle School Fluency Assessment (Texas Education Agency, University of Houston, & The University of Texas System, 2008).

Table 2

Descriptive Statistics: Means and Standard Deviations

Measure	<i>n</i>	<i>M</i>	<i>SD</i>
KBIT composite	287	97.66	13.675
TAKS	311	2,189.79	168.036
TMFSA			
Retell component	311	1.85	.616
Word reading fluency	311	74.78	18.541
Passage reading fluency	311	145.75	31.804
WJ-III			
Letter–Word ID	311	99.12	11.645
Word Attack	311	99.14	10.990
Passage Comprehension	311	93.43	10.866
Spelling	311	96.18	14.860
TOWRE summary scores	311	98.27	14.363
AIMSweb reading maze	311	190.84	58.356
GRADE composite	311	90.62	11.212
TOSRE summary score	311	91.13	14.123
Valid <i>n</i> (listwise)	287		

Note: KBIT = *Kaufman Brief Intelligence Test*; TAKS = Texas Assessment of Academic Skills; TMSFA = Texas Middle School Fluency Assessment; WJ-III = Woodcock-Johnson III; TOWRE = *Test of Word Reading Efficiency*; AIMSweb; GRADE = Group Reading Assessment and Diagnostic Evaluation; TOSRE = *Test of Sentence Reading Efficiency*.

Table 3

Descriptive Statistics: Correlations Among the Measures

Variable	1	2	3	4	5	6	7	8	9
1. WJ Word Attack									
2. WJ Letter-Word	.742 ^a								
3. WJ Spelling	.555 ^a	.609 ^a							
4. TOWRE Summary	.630 ^a	.644 ^a	.579 ^a						
5. TMSFA AvePassES	.550 ^a	.595 ^a	.572 ^a	.813 ^a					
6. AIMS maze	.328 ^a	.358 ^a	.324 ^a	.462 ^a	.517 ^a				
7. GRADE composite	.264 ^a	.408 ^a	.339 ^a	.336 ^a	.430 ^a	.195 ^a			
8. WJ Passage Comprehension	.479 ^a	.578 ^a	.508 ^a	.453 ^a	.498 ^a	.329 ^a	.564 ^a		
9. TAKS	.322 ^a	.509 ^a	.387 ^a	.383 ^a	.471 ^a	.281 ^a	.555 ^a	.581 ^a	
10. Average retell	.037	.132 ^b	.075	.158 ^a	.180 ^a	.183 ^a	.155 ^a	.206 ^a	.257 ^a

Note: WJ = Woodcock-Johnson III; TOWRE = *Test of Word Reading Efficiency*; TMSFA = Texas Middle School Fluency Assessment; AvePassES = TMSFA Passage Reading Fluency average equated score; GRADE = Group Reading Assessment and Diagnostic Evaluation; TAKS = Texas Assessment of Academic Skills.

^aCorrelation is significant at the .01 level (two-tailed).

^bCorrelation is significant at the .05 level (two-tailed).

Table 4

Tests of Model Fit

Retell	CFI	TLI	RMSEA	SRMR	AIC	BIC
Narrative	.956	.938	.084	.042	24,412.788	24,536.201
Average	.958	.941	.081	.039	24,354.513	24,477.926

Note: CFI = comparative fit index; TLI = Tucker–Lewis Index (Tucker & Lewis, 1973); RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

Table 5

Factor Loadings of Retell on the Latent Constructs

Construct	Average retell	
	Estimate	SE
Word identification, by average retell	.167 *	.058
Fluency, by average retell	.194 *	.057
Comprehension, by average retell	.250 *	.059

*
 $p < .001$.

Table 6

Relationships of the Factors to the Measures

Factor/measure	Estimate	SE	Residual variance
Word Identification by			
WJ Letter–Word	.891*	.019	.205
WJ Word Attack	.808*	.024	.347
WJ Spelling	.711*	.033	.495
Fluency by			
TOWRE Summary	.903*	.018	.185
TMSFA AvePassES	.901*	.019	.187
AIMS maze	.541*	.043	.707
Comprehension by			
GRADE composite	.685*	.037	.531
WJ Passage Comprehension	.825*	.030	.320
TAKS	.737*	.035	.457
Average retell	.250*	.059	.938

Note: WJ = Woodcock-Johnson III; TOWRE = *Test of Word Reading Efficiency*; TMSFA = Texas Middle School Fluency Assessment; AvePassES = TMSFA Passage Reading Fluency average equated score; GRADE = Group Reading Assessment and Diagnostic Evaluation; TAKS = Texas Assessment of Academic Skills.

$p < .001$.

Table 7

Influence of Student Characteristics on Factors

Factor/student characteristic	Estimate	SE
Word identification on		
SES	-.109	.057
Bilingual	-.221 **	.070
ELL	-.098	.072
LEP	.174 ***	.078
Special education	-.330 *	.055
KBIT	.312 *	.055
Fluency on		
SES	.002	.059
Bilingual	-.204 **	.072
ELL	-.080	.074
LEP	.017	.081
Special education	-.307 ***	.054
KBIT	.221 ***	.058
Comprehension on		
SES	-.154 **	.053
Bilingual	-.315 ***	.065
ELL	-.180 **	.067
LEP	.066	.073
Special education	-.219 ***	.050
KBIT	.469 ***	.049

Note: SES = socioeconomic status; ELL = English language learner; LEP = limited English proficient; KBIT = *Kaufman Brief Intelligence Test*.

*
 $p < .001$.

**
 $p < .01$.

 $p < .05$.

Table 8

Influence of Student Characteristics on Retell Measure

Average retell on	Estimate	SE
SES	.010	.062
Bilingual	-.016	.079
ELL	.041	.077
LEP	-.048	.082
Special education	.066	.060
KBIT	-.077	.074

Note: SES = socioeconomic status; ELL = English language learner; LEP = limited English proficient; KBIT = *Kaufman Brief Intelligence Test*.