Published in final edited form as:

Read Psychol. 2012 ; 33(6): 562–585. doi:10.1080/02702711.2011.557333.

The Influence of Testing Prompt and Condition on Middle School Students' Retell Performance

Deborah K. Reed and Yaacov Petscher

Florida Center for Reading Research, Florida State University

Abstract

This study examined whether the type of prompt or the method of passage reading had an effect on the retell performance of 6th–8th graders randomly assigned to one of four retell testing conditions. Both the type of prompt and the use of follow-up prompting were significantly related to the percentage of predetermined idea units retold. Effect sizes were approximately moderate ($d = .44-.62$) when one change was made to the prompt but were strong ($d = .96-1.05$) with a combination of changes. The addition of silent reading did not significantly improve performance.

Retell and free recall are among the most commonly used approaches for assessing students' reading comprehension, particularly in classroom assessments (Cohen, Krustedt, & May, 2009) and informal reading inventories (IRIs; Nillsson, 2008). After reading a passage, students are asked to tell what they remember in their own words and, usually, without referring back to the passage. Recently, retell has been added to oral reading fluency (ORF) measures, such as the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS; Good & Kaminski, 2011), to improve the face validity of the instruments used to frequently monitor students' reading progress.

Retell is believed to have value beyond simply improving teachers' acceptance of ORF assessment results. Research findings suggest that retells capture information about students' understanding of text that may be inconsistent with the rate and accuracy with which they read the words (Keenan, Betjemann, & Olson, 2008; Kucer, 2009). Moreover, reliability and validity studies indicate that retell tasks used in conjunction with ORF provide information about students' reading comprehension abilities that might be useful in planning instruction (Marcotte & Hintze, 2009; Reed, Vaughn, & Petscher, 2012). The utility of retell measures may be constrained, however, by the lack of consistency with which they are administered and the lack of reliability with which responses are scored (Nillsson, 2008).

Background and Significance

As used in assessment protocols, *retell* is not easily distinguished from paraphrasing, summarizing, or identifying the main idea of a passage (Duffelmeyer & Duffelmeyer, 1987).

Address correspondence to Deborah K. Reed, Florida Center for Reading Research, Florida State University, 2010 Levy Avenue, Ste 100, Tallahassee, FL 32310. dkreed@fcrr.org.

The content is solely the responsibility of the author and does not necessarily represent the official views of the Texas Education Agency.

This is due, in part, to the wide variability with which the retell is elicited from students, as well as whether and what kind of a follow-up prompt is administered. Students may have been asked to: tell about the passage, retell the passage, retell as many details as possible, retell/recall everything, provide a summary, paraphrase the passage, retell as much as could be remembered, retell the passage so someone who had not read it could understand it, retell in your own words, give the first 10 facts remembered, tell everything the passage said word-for-word, retell the passage and what it was about, tell what the passage was mostly about, or freely recall the passage using the exact words from the passage (e.g., Cohen et al., 2009; Cote, Goldman, & Saul, 1998; Curran, Kintsch, & Hedberg, 1996; Freebody & Anderson, 1983; Fuchs & Fuchs, 1992; Gagne, Bing, & Bing, 1977; Hagtvet, 2003; Kucer, 2009; Marcotte & Hintze, 2009; Vieiro & Garcia-Madruga, 1997). After these initial prompts, students may or may not be provided follow-up prompting when they hesitate or pause. Follow-up prompts might be general encouragement to continue providing information, cued elicitation of particular information, a series of comprehension questions, a directive to summarize the retell/passage, or some combination thereof.

Rarely, if ever, are explanations provided for why a particular prompt was chosen for the retell task. It might be assumed that the wording is based on how the responses are to be scored. However, the same prompt has been used with different scoring mechanisms (e.g., Marcotte & Hintze, 2009; Tenenbaum, 1977) and, conversely, different prompts were administered when responses were scored with a similar mechanism (e.g., Cohen et al., 2009; Kucer, 2009). Some instruments or studies have offered testers a choice among prompts, even though the same evaluation criteria are applied to all responses regardless of the prompt chosen (e.g., Bader & Pearce, 2009; Pearman, 2008).

The implication of these variations is that the way in which the retell is elicited is of no consequence to the resulting response, but some researchers provide evidence to the contrary. A study of 10th- through 12th-graders indicated that how students are oriented to the task can influence their responses in qualitatively different ways (Gagne et al., 1977). Others have argued that the mere addition of the phrase “in your own words” alters the nature of the task from one of formal summarizing and synthesizing of information to the more ubiquitous paraphrasing without regard to the relative importance among ideas (Browne, Campione, & Day, 1981; Kintsch & van Dijk, 1978). However, the prompting conditions associated with the best performance on retell measures have not yet been established in the extant research.

Further variability among instruments is apparent in the method of passage reading conducted prior to the administration of the retell prompt. Students may be required to read orally (e.g., Good & Kaminski, 2011), silently (e.g., Cooter, Flynt, & Cooter, 2007), or both (e.g., Beaver, 2003; Johns, 2008). Still other protocols allow the tester or student to choose the method of reading (e.g., Applegate, Quinn, & Applegate, 2008; Leslie & Caldwell, 2006; Roe & Burns, 2007; Woods & Moe, 2007). It could be hypothesized that retell would be better after silent reading because of the added pressure placed on reading “performance” and the speed with which words are processed when the retell follows ORF (Gold, 1983). Most standardized comprehension assessments require students to read silently prior to answering questions (Paris, 2007), but research examining whether oral or silent reading is

more efficacious for comprehension has been inconclusive (e.g., Hale et al., 2007; McCallum, Sharp, Bell, & George, 2004; Miller & Smith, 1990; Prior & Welling, 2001). The only identified study to use text recall as an outcome measure following oral versus silent reading had a limited sample of boys identified with learning disabilities (Fuchs & Maxwell, 1988). Although results indicated oral reading was associated with better recall, it is not known whether this finding would generalize to a more heterogeneous population of students.

The purpose of this study, therefore, is to improve the utility of retell protocols by more clearly identifying the testing conditions under which students demonstrate better performance. Given their reportedly widespread use, teachers and test developers would benefit from having more information on the optimal procedures for administering retell measures. The research question used to develop the study was: Is the wording of the initial prompt, the inclusion of a follow-up prompt, or the opportunity to silently reread the passage related to retell performance among students in grades 6 through 8?

Because retells are more reliably scored quantitatively by the percent of predetermined idea units included (e.g., Best, Floyd, & McNamara, 2008; McGee, 1982; van den Broek, Tzeng, Ridsen, Trabasso, & Basche, 2001; Wright & Newhoff, 2001), it was hypothesized that a prompt asking students to tell everything they remember would be associated with better scores than a prompt asking students to tell what a passage is “mostly about.” Furthermore, it was hypothesized that students of various ability levels would have improved retells if provided an opportunity to reread a passage silently prior to retelling it, rather than delivering their responses immediately following a 1-minute timed ORF component.

METHOD

Research Setting and Participants

A total of 589 students from two middle schools located in different Texas cities returned parent permission forms to participate in the study. Seventeen of those students were absent or otherwise unavailable on the days of testing, and 45 students experienced testing abnormalities as explained in the section on fidelity of test administration procedures. This left a final sample of 527 students: 150 in grade 6, 168 in grade 7, and 209 in grade 8. There were slightly more females ($n = 54\%$) than males ($n = 46\%$), and only a small percentage of students were identified as English language learners ($n = 8\%$) or as having learning disabilities ($n = 5\%$). However, half the sample received free or reduced-price lunch. The majority of students were Hispanic ($n = 65\%$) and an additional 28% were white. Among the remaining students, 4% were African American, 2% were Asian, and 1% were Native American.

All participants were randomly assigned to one of four retell testing conditions: (1) onetime oral reading and prompted to tell in their own words what the passage was mostly about; (2) onetime oral reading and prompted to tell everything they remembered; (3) one-time oral reading, prompted to tell everything they remembered, and provided follow-up prompting; (4) oral reading followed by silent rereading, prompted to tell everything they remembered, and provided follow-up prompting. The first condition was intended to capture an alternative

way of asking for a retelling that would not specifically preference summarizing or giving a main idea (“what the passage was mostly about”) and that would leave open the possibility of paraphrasing (“in your own words”). The latter was done to create more equivalency to the prompt used in the other three conditions, “tell me everything you remember reading in the passage.” Similarly, the “tell everything” prompt avoided using the word “retell,” so that the comparison between groups could more confidently be based on nuanced variations in prompt wording rather than on students’ level of familiarity with a named skilled (i.e., summarizing, identifying a main idea, retelling).

The fourth condition introduced silent reading after the oral reading required for the ORF component of the assessment. This was done for two reasons: (1) having a silent reading only condition would require that three additional passages be administered, thus potentially confounding results with measurement artifacts; and (2) combining oral and silent reading would allow for a comparison between conditions 3 and 4 that was not based on an advantage of one method of reading over another, but on the efficacy of allowing more time to reread for meaning.

Measures

All students were administered the Passage Reading Fluency and Word Reading Fluency subtests of the Texas Middle School Fluency Assessment (TMSFA; Texas Education Agency, University of Houston, & The University of Texas System, 2008a), an individually administered assessment primarily based on ORF. However, both subtests of the TMSFA have demonstrated predictive validity to the former state criterion-referenced assessment of reading comprehension, the Texas Assessment of Knowledge and Skills (TAKS; Texas Education Agency, 2004), as will be described in the sections below.

Passage Reading Fluency subtest (Texas Education Agency et al., 2008a). In Passage Reading Fluency, students read a series of three passages of increasing difficulty as defined by the text’s Lexile level (The Lexile Framework, 2007). Passages are predetermined by grade (i.e., 6th, 7th, or 8th grade) and testing point (e.g., beginning of year, middle of year, end of year). Within each set, the passages may be a combination of narrative and informational or all may be informational.

All passages were equated, and the mean intercorrelation of the performances on five passages across testing points ranged from .86 to .98 with a sample of students in grades 6 through 8 (Texas Education Agency, University of Houston, & The University of Texas System, 2008b). The equating was designed to remove the effects of practice, text difficulty, text type, or other measurement artifacts associated with using different passages (Francis et al., 2008). Hence, raw scores of words correct per minute (WCPM) were converted to an equated score using tables normed by grade level and testing point.

Students read each of the three passages aloud under a 1 minute time limit while a tester marks substitutions, mispronunciations, alterations, reversals, hesitations lasting 3 seconds, and skips as errors. The criterion validity of the Passage Reading Fluency subtest ($r = .50$) was established with TAKS reading test (Texas Education Agency, 2004).

After students read as far as possible in each passage within the one minute, the retell portion of the subtest is administered. In the three conditions that do not allow silent reading, the examiner immediately conceals the text and delivers the retell prompt. In the fourth condition, students have an opportunity to reread the passage silently prior to the administration of the retell prompt. In all conditions, the examiner transcribes the student response as accurately as possible on the record sheet and scores the response by calculating the percent of the predetermined idea units recalled.

The retell portion of the TMSFA was originally validated through confirmatory factor analysis by comparing the fit of a three-factor model of reading to the data from a diverse sample of 7th- and 8th-graders (Reed et al., 2012). The final model demonstrated adequate fit ($\chi^2 = 97.316$ {32}; CFI = 0.958; TLI = 0.941; RMSEA = .081) with retell having a greater χ^2 difference ($\chi^2 = 16.652$ {1}, $p < .001$) and factor loading (.250, $p < .001$) on the comprehension construct as compared to the fluency or word identification constructs.

Word Reading Fluency subtest (Texas Education Agency et al., 2008a). Students read aloud a series of three lists of real words for 1 minute each while the tester marks substitutions, mispronunciations, alterations, reversals, skips, and 3-second hesitations as errors. The lists are of increasing difficulty as defined by the length and frequency of the words (Zeno, 1995), but the lists only vary by testing point and not by grade level. The mean intercorrelation of performances ranged from .89 to .98 with a sample of students in grades 6 through 8 (Texas Education Agency et al., 2008b). The criterion validity of the Word Reading Fluency subtest ($r = .36$) was established with the TAKS reading test (Texas Education Agency, 2004). Although not considered a measure of students' comprehension ability, the Word Reading Fluency subtest was administered in this study because it has demonstrated predictive validity to the TAKS and is a part of the routine administration of the TMSFA. Similarly to the Passage Reading Fluency subtest, raw WCPM scores on the Word Reading Fluency subtest are converted to an equated score using tables normed by grade level and testing point to remove the effects of measurement artifacts.

Procedures

TEST ADMINISTRATION—Students participated in the regularly scheduled spring Benchmark and TAKS testing according to the procedures and timelines specified by the school district or state department of education, respectively. Parents authorized the school to release those scores to the researchers for the purposes of this study.

The subtests of the TMSFA were administered at the school sites by research assistants who participated in a half-day training. Procedures for the Word Reading Fluency subtest and for the ORF portion of the Passage Reading Fluency subtest were identical across conditions. Only the procedures for the retell portion of the Passage Reading Fluency subtest varied by group assignment. Research assistants had to demonstrate WCPM scoring accuracy through practice administrations of the assessment that utilized audiotapes of a student reading the passages and word lists. Those whose marking of errors was off by more than ± 1 on a single passage or word list continued practicing with a partner until they were within ± 1 error. In addition, the research assistants practiced each of the retell testing conditions and had to demonstrate 100% accuracy in all four protocols.

Approximately 94% of the students were tested during their English language arts class period. Those who were absent or involved in other school activities made up the testing on a subsequent day. In those situations, students were pulled from another class and brought to the testing room. Research assistants sat across from each student and followed the scripted testing procedures preprinted on the cover sheet of the examiner documents. Individual testing began with the Passage Reading Fluency subtest. Students in conditions 1 through 3 received the appropriate retell prompt(s) immediately upon completing the 1-minute reading of each passage. Condition 1 students were prompted with: "Tell me in your own words what this passage was mostly about." They did not receive any follow-up prompting, and research assistants were trained to impassively transcribe what each student said without suggesting in body language or speech that the student provide any more information than what was freely recalled.

Students in conditions 2 and 3 were prompted with: "Can you tell me everything you remember reading in the passage?" Similarly to condition 1, research assistants were trained to impassively transcribe what condition 2 student said without encouraging any elaboration. However, each time a student in condition 3 paused in retelling, the research assistants were trained to deliver the follow-up prompt, "Do you remember anything else?" until the student indicated he/she could recall no more information from the passage.

In condition 4, students were provided additional time after the ORF component to silently reread the passage. Research assistants reset the timer for 60 seconds and instructed the student to read the passage again, silently this time, from the beginning and stopping at the word where he/she ended during the 1 minute of oral reading. Students could spend the entire additional minute rereading, or they could indicate when they had finished rereading and were ready to provide their retell. When students indicated they were ready or when the 1-minute timer sounded, the research assistant delivered the initial prompt, "Can you tell me everything you remember reading in the passage?" and the follow-up prompt(s), "Do you remember anything else?" until the student indicated he/she could recall no more information from the passage. Hence, the only difference between conditions 3 and 4 was the inclusion of the silent rereading time.

After following the appropriate procedures for administering all three passages of the Passage Reading Fluency subtest, research assistants then administered the Word Reading Fluency subtest to each student. The procedures for this subtest were identical across the four groups.

In between testing students, research assistants tabulated errors made in oral reading, calculated the raw WCPMs, converted the WCPMs to equated scores, averaged the equated scores from the three passages, averaged the equated scores from the three word lists, made sure the transcribed retells were legible, and counted the number of words in student retells.

FIDELITY OF TEST ADMINISTRATION PROCEDURES—Throughout the test administration, the principal investigator observed the research assistants. Each had a colored folder designating the condition to which the student being tested had been assigned. If any violation of the procedures was observed, the principal investigator would flag the

testing packet for review. In all, 45 testing packets (8% of the full sample) were marked for administration abnormalities that precluded the use of the data. Approximately 3% of excluded cases were due to the tester forgetting to administer the retell on one or more passages, forgetting to administer one or more passage(s) or word list(s), delivering the wrong prompt for the condition, forgetting to conceal the passage during student retelling, and forgetting to set the timer for one or more ORF components. Most of these types of abnormalities occurred in condition 4, which had the greatest number of procedural steps and took the longest amount of time for testing.

In addition, one research assistant was observed using hand motions to encourage students in condition 1 to continue retelling, and one research assistant was determined to have a hearing loss that prevented him from accurately recording students' reading or retelling in condition 3. The data from 11 students in condition 1 and 7 students in condition 3 were not included in the analysis.

HANDLING OF DATA—After each testing session, research assistants checked each other's documents for completion and accuracy. The ORF components of the TMSFA subtests were initially scored by the individual who administered the test and were scored a second time by another research assistant who verified the number of errors recorded, the calculation of the WCPM, and the conversion of the raw score to the equated score.

The transcribed retell responses were scored by the principal investigator and two research assistants who had participated in the development of the scoring guides. Each response was compared to a set of predetermined idea units for the passage. The scoring guides include allowable alternatives for wording the idea units, and scores are based on the percent of idea units accurately recalled out of the number possible for the word count attained. That is, students read to different places in the passage during the ORF portion of the assessment and, therefore, were eligible to recall a different number of idea units based on the total number of words read. The scoring guides provide the minimum word count needed to have read the information contained in each idea unit (see Appendix A for a sample scoring guide).

All data were manually entered into a database by the two research assistants involved in the development of the retell scoring guides. This provided another opportunity to check the accuracy of the scores recorded on the examiner documents. The research assistants also typed all student retells into the database before scoring them a second time to calculate inter-rater reliability. When entering the retell, assistants verified the number of words in the retell as recorded by the original examiner.

INTER-RATER RELIABILITY OF RETELL SCORING—Twenty-five percent of the retell responses were double scored for the purposes of determining inter-rater reliability among the three individuals applying the scoring guides. Using the database with the typed retells, an individual other than the original scorer applied the scoring guide to the student response a second time. Observed agreement on scores for individual passage retells ranged from a low of 82% (6th-grade passage 1) to a high of 95% (7th-grade passage 2). However,

the average of the scores obtained on the three passages is the main unit of analysis for the TMSFA.

Therefore, agreement was also calculated by determining the correlation of the average retell scores obtained by the two raters. The relationship, as measured by the intraclass correlation (i.e., .978), was strong and suggested that agreement would not have occurred by chance. This was a substantial improvement over the previous approach to scoring the TMSFA retell component, which relied upon a holistic approach to evaluating the quality, coherence, and completeness of responses. Agreement on qualitative scores was reportedly low, 63%, and adjusted for chance with the kappa statistic, .37 (Reed et al., in press).

RESULTS

Preliminary analysis of the data indicated that only 1% of the data across the 11 variables were missing (i.e., on average retell). Little's test of data missing completely at random (Little, 1988) indicated that the data were not missing completely at random ($\chi^2(7) = 17.47$, $p = .015$). Because attrition in randomized experimental designs may create differences among the groups, it was important to assess the extent to which attrition contributed to a bias of estimated treatment effects. This was examined through a comparison of the differential proportion of missing data between the treatment and control groups with the overall proportion of missing data. Using guidelines outlined by What Works Clearinghouse (2008), we determined that the relationship between total missing and differential missing data for each outcome was at an acceptable level of bias under conservative assumptions.

Descriptive statistics for the full sample are provided in Table 1. As can be seen, students retold, on average, only about 31% of the idea units they read in the passage. Their responses were about 33 to 38 words in length. Correlations among the components of the TMSFA are provided in Table 2. The ORF portion of the Passage Reading Fluency subtest was strongly correlated ($r = .86$) to the Word Reading Fluency subtest, an ORF assessment of words in isolation. However, retell scores bore little relationship to ORF scores. With the exception of the weak correlation between retell scores on passage 2 with the Passage Reading Fluency equated score, only the lengths of utterances in retelling were significant ($r = .19$ to $.30$) and mostly weakly related to the ORF components. This suggests that retell is measuring different reading skills.

Means and standard deviations by testing condition are provided in Table 3. All data were analyzed using one-way analysis of variance (ANOVA). For all treatment effect models, a Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) was applied to statistically significant effects. The use of multilevel modeling serves as a method to control the Type I error rate to account for clustering; however, when simultaneously testing for treatment effects across multiple outcomes, it is critical that a correction is applied to statistically significant findings (What Works Clearinghouse, 2008). Though a Bonferroni has traditionally served as popular correction to multiple hypothesis testing, it has been found to be stringent in practical settings. As such, the Benjamini-Hochberg correction was utilized to control for the false-discovery rate, rather than the family-wise error rate.

In the first ANOVA model, testing condition served as the independent variable, and average retell score served as the dependent variable. The mean scores indicated a main effect (see Table 4), so post-hoc analysis was conducted using Scheffe and Tukey tests. These revealed several significant differences among groups.

Students prompted to tell “what this passage was mostly about” without follow-up prompting (condition 1) had lower retell scores than students in all other conditions who were prompted to tell “everything you remember reading in the passage.” This slight difference in the wording of the prompt delivered to students had a significant effect, $p < .001$, on their subsequent retell performance as evaluated by a quantitative scoring guide. Because the change in wording was the only distinction between conditions 1 and 2, an effect size was calculated using Cohen’s d (Cohen, 1988), or the quotient of the difference in the group means divided by the pooled standard deviation, to estimate the magnitude of the difference in student performance. The resulting $d = .62$ was interpreted as a moderate effect.

The students provided follow-up prompting, “Do you remember anything else?” had significantly better retell scores than students in conditions 1, $p < .001$, and 2, $p < .01$, who were not provided any encouragement to continue retelling. With the same initial prompt as that delivered in condition 2, the addition of the follow-up prompting in condition 3 produced a small-to-moderate effect size, $d = .40$. Whereas, using a different initial prompt as well as adding the follow-up prompting produced a large effect size, $d = .96$, between student performance in conditions 1 and 3. This suggests that the combination of changes in wording plus encouragement to continue retelling had a greater magnitude of effect on scores than either one of the differences introduced individually.

The final comparison was between the retell scores of students in condition 4, provided an opportunity to silently reread the passage, and the students in condition 3, who received the same prompt and follow-up prompting, but no opportunity to silently reread. Scheffe and Tukey tests revealed no significant difference between the two groups. This was confirmed by the almost nonexistent effect size, $d = .03$. Whereas the addition of silent rereading did not impact student retell performance, the change in wording and the inclusion of follow-up prompting remained a significant influence ($p < .001$, $d = 1.05$ when comparing conditions 1 and 4; and $p < .01$, $d = .44$ when comparing conditions 2 and 4).

Because the method of scoring was based on the percent of predetermined idea units recalled, the relationship between length of utterance and retell score was explored. The correlations between the number of words retold and the retell score on individual passages were all moderate ($r = .52$ to $.65$, $p < .01$). The second ANOVA used testing condition as the independent variable and length of utterance as the dependent variable and revealed a main effect, $p < .001$ (see Table 5). Post-hoc Scheffe and Tukey tests produced results similar to those when comparing the average retell scores across groups. That is, students said significantly more words when prompted to “tell everything” ($p < .001$) and when encouraged to continue with follow-up prompting ($p < .01$), but length of utterance was not significantly different when provided an opportunity to silently reread the passage before

retelling on passages 2 and 3. On passage 1, the easiest of the three passages read, student retells were modestly though significantly longer when allowed to silently reread, $p < .05$.

DISCUSSION

This study sought to improve the utility of retell protocols by more clearly identifying the prompt(s) and method of reading (i.e., oral only or oral + silent rereading) associated with better performance. Students in grades 6–8 were randomized into one of four retell testing conditions: (1) one-time oral reading and prompted to tell in their own words what the passage was mostly about; (2) one-time oral reading and prompted to tell everything they remembered; (3) one-time oral reading, prompted to tell everything they remembered, and provided follow-up prompting; (4) oral reading followed by silent rereading, prompted to tell everything they remembered, and provided follow-up prompting.

A reported limitation of retell instruments is their low inter-rater reliability, particularly when scored with holistic mechanisms or rubrics evaluating overall quality and/or coherence (Klesius & Homan, 1985; Nilsson, 2008). To improve the reliability of the TMSFA retell scoring (Reed et al., 2012), a quantitative approach was developed. Inter-rater agreement on the percent of predetermined idea units recalled as measured by the intraclass correlation was strong (i.e., .98).

Using the average retell score from three passages, differences in means among groups were significant, favoring an initial prompt that asks students to tell everything they remember as compared to telling what the passage was “mostly about.” Moreover, encouraging students to continue retelling by asking if they remembered anything else also significantly increased students’ scores. Effect sizes for making one of these changes, either to the wording of the initial prompt or by adding follow-up prompting, produced approximately moderate effect sizes ($d = .44-.62$). However, combining both the “tell everything” prompt with follow-up prompting produced strong effect sizes over only asking students what the passage was “mostly about” ($d = .96-1.05$).

The results were similar when considering the number of words students included in their retells. The length of utterance was significantly longer when asked to tell everything and was longer still when encouraged to keep retelling. It seems students considered the initial prompt a literal indicator of what they were to do: either give a succinct “gist” of what the passage was “mostly about” or give as much information as possible. Yet even with the “tell everything” prompt, students did not provide all the information they actually recalled unless they were repeatedly prompted to continue retelling.

These are important results given the great amount of variance in the prompts used across retell instruments. Consistent with Gagne and colleagues’ (1977) finding that changes to the way students are oriented to the task can produce qualitative differences in their responses, results in the present study suggest that even small alterations of the prompt also produce quantitative differences in responses. Combinations of changes result in an even greater magnitude of effect. This potentially precludes attempts to synthesize findings from studies employing different retell protocols. At the least, it makes interpreting retell scores

problematic if there is not a movement toward greater consistency in the wording and delivery of prompts.

An equally compelling finding was the lack of significant differences in the scores of students provided identically worded prompts and follow-up prompting, but a difference in the method of reading prior to retelling. Intuitively, the addition of silent reading to the oral component would seem to provide an advantage in retelling because it reduces the pressure placed on reading “performance” and the speed with which words are identified when reading orally (Gold, 1983) and gives students an opportunity to reread for meaning. In fact, silent reading has been employed in a great number of studies of retell (e.g., Best et al., 2008; Curran et al., 1996; Fuchs & Fuchs, 1992; Gagne et al., 1977; Irwin, 1979; Tindal & Parker, 1989; van den Broek et al., 2001; Walczyk & Raska, 1992). Here, however, students provided an opportunity to silently reread the passage prior to retelling did not recall significantly more idea units. On only the easiest passage did students say a modestly but significantly greater number of words in their retell. Remember, however, that the greater length of utterance still did not equate to a statistical improvement in their scores.

Because retell is being included with ORF progress monitoring instruments (e.g., Good & Kaminski, 2011), adding just 1 minute onto the testing time for each passage can impact the feasibility of using frequently administered measures. In terms of the TMSFA Passage Reading Fluency subtest, this equates to an additional 3 minutes of testing per student per testing point, or 9 minutes per student per year. If only 20 students are tested, a teacher potentially loses 180 minutes of instructional time by allowing the silent rereading. On a middle school campus with 45-minute periods, that is about four class periods of instruction that could have been devoted to improving comprehension rather than measuring it. The lack of strong correlations between retell and the ORF components of the TMSFA suggest that retell is measuring some other skill than what is being captured by ORF (Marcotte & Hintze, 2009; Reed et al., 2012). Therefore, retell has the potential to better inform reading instruction if administered under conditions that preserve instructional time and optimize student performance.

Limitations

As discussed, student retells were scored quantitatively. Therefore, the quality of a student response can only be judged as one in which more predetermined information from the passage was accurately stated. No inferences can be made about the organization of those ideas or the level of critical thinking exhibited by the student. Most idea units were based on explicitly stated information, although suggested means of restating the ideas were included on the scoring guides. Comprehension researchers typically prefer to distinguish the degree of telling versus transforming knowledge required by the skill the reader is performing (Kintsch & van Dijk, 1978; Scardinalia & Bereiter, 1987). Because the scoring guide allows for verbatim as well as paraphrased responses, scores cannot be considered indicative of a student’s level of complex thought.

Another limitation of this study is that no information is available on how reliably teachers are able to administer the instrument or apply the scoring guides. Research assistants who received thorough training still experienced difficulty adhering to the proper procedures, and

nearly 8% of the data were lost due to testing abnormalities. In addition, only two research assistants who helped develop the retell scoring guides were involved in scoring the students' responses. Hence, there are questions about the amount of training, oversight, and ongoing guidance it would take for teachers to make practical use of the TMSFA or other similar progress monitoring instruments with ORF and retell components.

In addition, the passages included in the TMSFA are not necessarily typical of texts used in middle school content area classrooms. The expository passages were biographical or descriptive in nature and, potentially, placed greater demands on students' background (e.g., history, geography, culture) than the narrative passages. However, they were written as informational narratives and therefore were not representative of the complexity or conceptual density that characterizes subject matter textbooks (Kosanovich, Reed, & Miller, 2010). The passages were all from released state reading assessments, which are considered appropriate measures of students' reading comprehension (Torgesen & Miller, 2009), but students' ability to retell information from the passages may not provide much instructional guidance to content area teachers.

Recommendations for Future Research

As shown on Table 3, the mean retell scores indicate that students only recalled between 26 and 37% of the idea units they read. The extant literature reports similar findings of students recalling low percentages of predetermined idea units (e.g., Hansen, 1978; McGee, 1982; Tindal & Parker, 1989). Future research might explore the interpretation of these results for planning reading instruction. Is recalling 30% of the idea units considered "good comprehension"? Is there a threshold proportion below which students could be considered "struggling readers"?

The present study did not attempt to manipulate students' retell ability with instruction, but many intervention studies use retell as an outcome measure. Therefore, it might be worthwhile to examine whether there is a point at which further improvements to the percentage of idea units recalled are no longer worth the investment of instructional resources. Retell is one skill associated with comprehension, but it may not capture the level of skill that would produce stronger effect sizes.

Finally, frequently administered assessments are most often used to monitor students with reading difficulties. The participants in this study were a representative sample of students at different ability levels in the schools but were less representative of the population of struggling readers. Future research might address whether different prompts and/or the opportunity to reread the passage silently would produce similar results with participants who were identified for inclusion on the basis of low standardized reading scores.

Acknowledgments

This research was supported by contract No 2356–27235 from the Texas Education Agency.

References

- Applegate, MD.; Quinn, KB.; Applegate, AJ. The Critical Reading Inventory: Assessing students' reading and thinking. 2. Upper Saddle River, NJ: Pearson Prentice Hall; 2008.
- Bader, LA.; Pearce, DL. BADER Reading and Language Inventory. Boston, MA: Allyn & Bacon; 2009.
- Beaver, JM. Developmental Reading Assessment, grades 4–8. 2. Parsippany, NJ: Pearson Education; 2003.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1995; 57:289–300.
- Best RM, Floyd RG, McNamara DS. Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*. 2008; 29(2):137–164.
- Brown AL, Campione JC, Day JD. Learning to learn: On training students to learn from texts. *Educational Researcher*. 1981; 10(2):14–21.
- Cohen, J. Statistical power analysis of the behavioral sciences. 2. Hillsdale, NJ: Erlbaum; 1988.
- Cohen L, Krustedt RL, May M. Fluency, text structure, and retelling: A complex relationship. *Reading Horizons*. 2009; 49(2):101–124.
- Cooter, RB.; Flynt, ES.; Cooter, KS. Comprehensive Reading Inventory: Measuring reading development in regular and special education classrooms. Upper Saddle River, NJ: Pearson Prentice Hall; 2007.
- Cote N, Goldman SR, Saul EU. Students making sense of informational text: Relations between processing and representation. *Discourse Processes*. 1998; 25(1):1–53.
- Curran CE, Kintsch E, Hedberg N. Learning-disabled adolescents' comprehension of naturalistic narratives. *Journal of Educational Psychology*. 1996; 88(3):494–507.
- Duffelmeyer FA, Duffelmeyer BB. Main idea questions on informal reading inventories. *The Reading Teacher*. 1987; 41(2):162–166.
- Francis DJ, Santi KL, Barr C, Fletcher JM, Varisco AI, Foorman BR. Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*. 2008; 46:315–342. [PubMed: 19083362]
- Freebody P, Anderson RC. Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*. 1983; 15(3):19–39.
- Fuchs LS, Fuchs D. Identifying a measure for monitoring student reading progress. *School Psychology Review*. 1992; 21(1):45–58.
- Fuchs LS, Maxwell L. Interactive effects of reading mode, production format, and structural importance of text among LD pupils. *Learning Disability Quarterly*. 1988; 11:97–105.
- Gagne ED, Bing SB, Bing JR. Combined effect of goal organization and test expectations on organization in free recall following learning from text. *Journal of Educational Psychology*. 1977; 69(4):428–431.
- Gold R. Recall of story schema categories by reading disabled adults: Effect of mode of presentation. *Perceptual and Motor Skills*. 1983; 56:387–396. [PubMed: 6866649]
- Good, RH.; Kaminski, RA. DIBELS^R Next assessment manual. Eugene, OR: Dynamic Measurement Group, Inc; 2011. Retrieved from <http://resourcecenter.adams14.org/media/27044/dibelsnextassessmentmanual.pdf>
- Hagtvet BE. Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing: An Interdisciplinary Journal*. 2003; 16:505–539.
- Hale AD, Skinner CH, Williams J, Hawkins R, Neddenriep CE, Dizer J. Comparing comprehension following silent and aloud reading across elementary and secondary students: Implication for curriculum-based measurement. *The Behavior Analyst Today*. 2007; 8(1):9–23.
- Hansen CL. Story retelling used with average and learning disabled readers as a measure of reading comprehension. *Learning Disability Quarterly*. 1978; 1(3):62–69.

- Irwin JW. Fifth grade readers' comprehension of explicit and implicit connective propositions. *Journal of Reading Behavior*. 1979; 11(3):261–271.
- Johns, JL. *Basic Reading Inventory: Pre-primer through grade twelve and early literacy assessments*. 10. Dubuque, IA: Kendall/Hunt Publishing Co; 2008.
- 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(3):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(2):71–76. [PubMed: 3882869]
- Kosanovich, ML.; Reed, DK.; Miller, DH. *Bringing literacy strategies into content instruction: Professional learning for secondary-level teachers*. Portsmouth, NH: RMC Research Corporation, Center on Instruction; 2010.
- Kucer SB. Examining the relationship between text processing and text comprehension in fourth-grade readers. *Reading Psychology*. 2009; 30(4):340–358.
- Leslie, L.; Caldwell, J. *Qualitative Reading Inventory*. 4. Boston, MA: Pearson Education; 2006.
- The Lexile Framework. The Lexile calculator. 2007. Retrieved from <http://www.lexile.com>
- Little RJA. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*. 1988; 83:1198–1202.
- 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]
- McCallum RS, Sharp S, Bell SM, George T. Silent versus oral reading comprehension and efficiency. *Psychology in the Schools*. 2004; 41(2):241–246.
- McGee LM. Awareness of text structure: Effects on children's recall of expository text. *Reading Research Quarterly*. 1982; 17:581–590.
- Miller SD, Smith DEP. Relations among oral reading, silent reading and listening comprehension of students at differing competency levels. *Reading Research and Instruction*. 1990; 29:73–84.
- Nillsson NL. A critical analysis of eight informal reading inventories. *The Reading Teacher*. 2008; 61(7):526–536.
- Paris, SG. *Encyclopedia of language and literacy development*. London, ON: Canadian Language and Literacy Research Network; 2007. Assessment of reading comprehension; p. 1-8. Retrieved from <http://www.literacyencyclopedia.ca/pdfs/topic.php?topId=226>
- Pearman CJ. Independent reading of CD-ROM storybooks: Measuring comprehension with oral retellings. *The Reading Teacher*. 2008; 61(8):594–602.
- Prior SM, Welling KA. Read in your head: A Vygotskian analysis of the transition from oral to silent reading. *Reading Psychology*. 2001; 22:1–15.
- Reed DK, Vaughn S, Petscher Y. The validity of a holistically-scored retell protocol for determining the reading comprehension of middle school students. *Learning Disability Quarterly*. 2012; 35:76–89.10.1177/0731948711432509 [PubMed: 23097594]
- Roe, BD.; Burns, PC. *Informal Reading Inventory*. Boston, MA: Houghton Mifflin; 2007.
- 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.
- Tenenbaum AB. Task-dependent effects of organization and context upon comprehension of prose. *Journal of Educational Psychology*. 1977; 69(5):528–536.
- Texas Education Agency. *Texas Assessment of Knowledge and Skills (TAKS) information booklet: Reading, grade 6, revised*. Austin, TX: Author; 2004. Retrieved from <http://www.tea.state.tx.us/index3.aspx?id=3693&menuid3=793>
- Texas Education Agency, University of Houston, & The University of Texas System. *Texas middle school fluency assessment*. Austin, TX: Author; 2008a.

- Texas Education Agency, University of Houston, & The University of Texas System. Texas middle school fluency assessment teacher's guide. Austin, TX: Author; 2008b.
- Tindal G, Parker R. Development of written retell as a curriculum-based measure in secondary programs. *School Psychology Review*. 1989; 18(3):328–343.
- Torgesen, JK.; Miller, DH. Assessments to guide adolescent literacy instruction. Portsmouth, NH: RMC Research Corporation, Center on Instruction; 2009.
- 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(3):521–592.
- Vieiro P, Garcia-Madruga JA. An analysis of story comprehension through spoken and written summaries in school-age children. *Reading and Writing: An Interdisciplinary Journal*. 1997; 9:41–53.
- Walczyk JJ, Raska LJ. The relation between low- and high-level reading skills in children. *Contemporary Educational Psychology*. 1992; 17:38–46.
- What Works Clearinghouse. Procedures and standards handbook: Version 2.0. Washington, DC: Author; 2008.
- 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.
- Wright HH, Newhoff M. Narration abilities of children with language-learning disabilities in response to oral and written stimuli. *American Journal of Speech-Language Pathology*. 2001; 10(3):308–319.
- Zeno, SM.; Ivens, SH.; Millard, RT.; Duvvuri, R. The educator's word frequency guide. New York, NY: Touchstone Applied Science Associates, Inc; 1995.

Appendix A: Retell Scoring Guide

Underground Town

Each row of the table represents an idea unit and is worth 1 point.

# of Words Read	Recalled Idea Unit	Maximum Possible Idea Units
19	People Families in Australia in Coober Pedy live underground in dugouts	1
35	They The dugouts The underground homes are similar to regular homes houses have kitchens, bedrooms, and living rooms* or	2
47	They The dugouts The underground homes have few windows mostly artificial light little sunlight or The people use lamps for light overhead lights	3
80	The roofs are made of layers of dirt	4
91	They The people Australians build dugouts live underground have this kind of home because of where they live because they live in a desert	5
111	It's hot 120 degrees in the summer or cold 32 degrees in the winter	6
132	There are † dust storms lots of flies (bugs) make it hard to live there or Dust storms Flies (bugs) make life miserable hard intolerable	7
139	Underground in the dugout it is comfortable 75 degrees the same temperature	8
163	They The people live stay in the town to mine opals because the opals are there	9
181	The town Coober Pedy produces is responsible for most of the opals in the world or Most of the world's opals come from the town Coober Pedy	10
199	They The people Early settlers have had homes underground dugouts to avoid harsh temperatures	11
212	Half the people live underground in dugouts	12
220	Businesses Schools Restaurants Other buildings are also underground or People Families in Coober Pedy enjoy living "down under" underground	13

* To earn credit for this idea unit option, the student must supply at least two of the three details.

† The student may also earn credit for this idea unit by referring to the lack of dust, flies, or bugs underground.

Table 1

Descriptive Statistics of Full Sample

	<i>N</i>	Mean	SD
Retell 1 score	527	30%	20%
Retell 2 score	527	29%	22%
Retell 3 score	527	35%	26%
Ave. retell	526	31%	16%
Retell 1 word count	527	38.17	17.70
Retell 2 word count	527	32.62	16.23
Retell 3 word count	527	37.33	18.44
PRF	527	160.66	34.12
WRF	527	76.82	20.91
TAKS	509	802.02	97.55
Benchmark	497	87.49	11.50

Note. PRF = Passage Reading Fluency subtest of the TMSFA; WRF = Word Reading Fluency subtest of the TMSFA

Table 2

Correlations Among Measures

	1	2	3	4	5	6	7	8
1. Retell 1 score								
2. Retell 2 score	.22**							
3. Retell 3 score	.32**	.21**						
4. Ave. retell	.70**	.66**	.76**					
5. Retell 1 word count	.65**	.27**	.29**	.55**				
6. Retell 2 word count	.41**	.52**	.26**	.54**	.66**			
7. Retell 3 word count	.38**	.30**	.54**	.58**	.55**	.55**		
8. PRF	-.07	-.10*	.01	-.07	.30**	.25**	.21**	
9. WRF	-.07	.002	.000	-.03	.30**	.28**	.19**	.86**

**
 $p < .01$

*
 $p < .05$

Table 3

Descriptive Statistics by Condition

Measure	Condition 1			Condition 2			Condition 3			Condition 4		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Retell 1	129	20%	18%	151	29%	18%	134	30%	20%	113	37%	22%
Retell 2	129	19%	19%	151	28%	20%	134	30%	20%	113	35%	22%
Retell 3	129	26%	25%	151	35%	24%	134	40%	30%	113	39%	26%
Ave. retell	129	22%	14%	151	30%	14%	134	40%	20%	113	37%	15%
Retell 1 WC	129	27.86	13.35	151	36.11	15.18	134	42.4	15.8	113	47.71	20.51
Retell 2 WC	129	24.86	14.21	151	30.39	12.71	134	35.8	16.1	113	40.78	18.15
Retell 3 WC	129	29.06	15.50	151	36.47	15.46	134	41.8	20.7	113	42.62	18.98
PRF	129	159.82	39.58	151	162.11	34.34	134	158.6	29.3	113	162.13	32.70
WRF	129	75.46	22.59	151	76.56	20.93	134	76.0	19.6	113	79.70	20.40

Note. PRF = Passage Reading Fluency subtest of the TMSFA; WRF = Word Reading Fluency subtest of the TMSFA.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

One-Way Analysis of Variance: Average Retell

	Sum of Squares	df	Mean Square	F	Sig.
Between groups	1.882	3	.627	28.156	.000
Within groups	11.633	522	.022		
Total	13.515	525			

Table 5

One-Way Analysis of Variance: Length of Utterance

		Sum of Squares	df	Mean Square	F	Sig.
Retail 1 word count	Between groups	26992.037	3	8997.346	34.141	.000
	Within groups	137828.247	523	263.534		
	Total	164820.284	526			
Retail 2 word count	Between groups	17349.081	3	5783.027	24.939	.000
	Within groups	121278.278	523	231.890		
	Total	138627.359	526			
Retail 3 word count	Between groups	14758.303	3	4919.434	15.678	.000
	Within groups	164103.906	523	313.774		
	Total	178862.209	526			